

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده مهندسی معدن، نفت و ژئوفیزیک

گروه اکتشاف معدن

رساله دکتری اکتشاف مواد معدنی

استفاده از الگوریتم TSVMSC به منظور بهبود

تلفیق نتایج حاصل از وارون سازی داده های

توموگرافی لرزه انکساری و مقاومت ویژه

نگارنده: بهمن محمدی

استاد راهنما

دکتر ابوالقاسم کامکار روحانی

شهریور ۱۳۹۷

سپاس خدای را که سخنوران، در ستودن او بمانند و شمارندگان، شمردن نعمت های او ندانند و کوشندگان، حق او را گزاردن نتوانند. و سلام و دورد بر محمد و خاندان پاک او، طاهران معصوم، هم آنان که وجودمان وامدار وجودشان است؛ و نفرین پیوسته بر دشمنان ایشان تا روز رستاخیز...

با درود فراوان به روح بزرگ پدر عزیزم و سپاس بیکران بر همدلی و همراهی و همگامی مادر و همسر دلسوز و مهربانم که سجده‌ی ایثارشان گل محبت را در وجودم پروراند، از صمیم قلب تشکر و قدردانی می‌کنم.

بدون شک جایگاه و منزلت معلم، اجل از آن است که در مقام قدردانی از زحمات بی شائبه‌ی او، با زبان قاصر و دست ناتوان، چیزی بنگاریم. اما از آنجایی که تجلیل از معلم، سپاس از انسانی است که هدف و غایت آفرینش را تامین می‌کند و سلامت امانت‌هایی را که به دستش سپرده‌اند، تضمین؛ بر حسب وظیفه و از باب "من لم یشکر المنعم من المخلوقین لم یشکر الله عزّ و جلّ" :

از استاد با کمالات و شایسته؛ جناب آقای دکتر ابوالقاسم کامکار روحانی که در کمال سعه صدر، با حسن خلق و فروتنی، از هیچ کمکی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این رساله را بر عهده گرفتند؛ و از اساتید فرزانه و دلسوز؛ جناب آقای دکتر علی مرادزاده، دکتر علیرضا عرب امیری و دکتر امین روشندل که زحمت داوری این رساله را متقبل شدند؛ کمال تشکر و قدردانی را دارم. باشد که این خردترین، بخشی از زحمات آنان را سپاس گوید.

در پایان از تمامی دوستان و همکلاسی‌هایم کمال تشکر و قدر دانی را دارم.

بهمن محمدی

شهریور ماه ۹۷

تقدیم بہ:
روح پاک پدر بزرگوارم

کہ عالمانہ بہ من آموخت تا چگونه در عرصہ

زندگی، ایستادگی را تجربہ نمایم
و بہ مادر و ہمسر عزیزم،

دریای بی کران فداکاری و عشق کہ وجودم برایشان

ہمہ رنج بود و وجودشان برایم ہمہ مہر

و ہمہ استفادہ کنندگان از این رسالہ...

تعمدنامه

اینجانب بهمن محمدی دانشجوی دوره دکتری رشته مهندسی معدن - اکتشاف دانشکده مهندسی معدن، نفت و ژئوفیزیک دانشگاه صنعتی شاهرود نویسنده پایان نامه استفاده از الگوریتم TSVMSC به منظور بهبود تلفیق نتایج حاصل از وارون سازی داده های توموگرافی لرزه انکساری و مقاومت ویژه تحت راهنمایی جناب آقای دکتر ابوالقاسم کامکار روحانی متعهد می شوم.

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود . استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده

الگوریتم استنتاج تبدیلی ماشین بردار پشتیبان بر اساس خوشه‌بندی طیفی (TSVMSC)، ناهمگنی فضایی در همه مدل‌ها را در بر می‌گیرد و نیز انعطاف‌پذیر بوده و می‌تواند با هر نوع داده‌ای برای شناسایی الگوها و ساختارهای غالب موجود در داده‌های چند متغیره سازگاری داشته باشد. این الگوریتم عملکرد پایدار و سطح دقت بالایی دارد. هدف از این رساله، بهبود تلفیق نتایج حاصل از وارون‌سازی داده‌های لرزه‌نگاری انکساری و مقاومت‌ویژه با استفاده از روش TSVMSC که یک روش دانش‌محور است، می‌باشد. داده‌های توموگرافی لرزه انکساری و مقاومت‌ویژه الکتریکی از محدوده یکی از سدهای در حال ساخت ایران برداشت و مورد تحلیل قرار گرفت. برای بررسی تلفیق مدل در یک وضعیت کنترل شده از داده‌های مصنوعی استفاده شد. بعد از وارون‌سازی داده‌های مصنوعی و صحرایی، برای تلفیق نتایج حاصله، از خوشه‌بندی‌های میانگین k ، میانگین فازی، گوستافسون کسل فازی و الگوریتم TSVMSC استفاده شد. با اعتبارسنجی خوشه‌بندی‌های مختلف از لحاظ تفکیک مناطق آبرفت، سنگ بستر با کیفیت متوسط و سنگ بستر با کیفیت خوب، الگوریتم TSVMSC به دلیل در بر گرفتن ناهمگنی‌ها و پیچیدگی‌های دو مدل دقیق‌تر و موثرتر بوده است. خوشه‌بندی گوستافسون کسل فازی نیز نسبت به خوشه‌بندی‌های دیگر موفق‌تر عمل نموده است. با اعتبارسنجی نتایج وارون‌سازی توموگرافی لرزه‌ای انکساری، خطای تخمین سرعت معادل $30/95$ درصد محاسبه گردید. بر اساس نتایج حاصل از آزمایش‌ها برای سرعت تراکمی، سرعت برشی و تخلخل، روابط تجربی برای محاسبه سرعت‌های تراکمی و برشی، در این محدوده، به ترتیب با درصد همبستگی $76/5$ و $77/5$ به دست آمد.

کلمات کلیدی: TSVMSC، توموگرافی لرزه‌ای انکساری، مقاومت‌ویژه الکتریکی، خوشه‌بندی

لیست مقالات مستخرج از پایان نامه

۱- ارزیابی کیفیت توده سنگ زیر سطحی و تفکیک مناطق عمقی مختلف با استفاده از روش توموگرافی لرزه

انکساری (بهمن محمدی، ابوالقاسم کامکار روحانی)، هفدهمین کنفرانس ژئوفیزیک ایران، تاریخ ارسال:

۹۴/۱۲/۲۵، پذیرش: ۹۵/۲/۲۳

۲- تعیین سرعت موج لرزه‌ای در حوضه های رسوبی به روش شبیه سازی چند متغیره گوسی **FFTMA**،

(بهمن محمدی، ابوالقاسم کامکار روحانی)، بیستمین همایش انجمن زمین شناسی ایران، تاریخ ارسال:

۹۵/۰۵/۲۰، پذیرش: ۹۵/۰۶/۱۶

۳- به کارگیری روش های خوشه بندی میانگین **K**، میانگین فازی و گوستافسون کسل در تلفیق نتایج وارون-

سازی داده های توموگرافی لرزه ای انکساری و مقاومت ویژه الکتریکی برای ارزیابی آبرفت و سنگ بستر،

(بهمن محمدی، ابوالقاسم کامکار روحانی)، مجله علوم زمین خوارزمی، تاریخ ارسال: ۹۵/۰۸/۱۰

۴- تلفیق نتایج وارون سازی داده های توموگرافی لرزه ای انکساری و مقاومت ویژه الکتریکی با استفاده از

روش خوشه بندی طیفی به منظور ارزیابی کیفیت توده سنگ زیر سطحی و تفکیک مناطق عمقی مختلف،

(بهمن محمدی، ابوالقاسم کامکار روحانی)، مجله مهندسی معدن دانشگاه تهران، تاریخ ارسال: ۹۵/۰۸/۱۶

فهرست مطالب

صفحه

عنوان

د.....فهرست جداول

ه.....فهرست اشکال

۱ فصل اول: کلیات

۲.....۱-۱ مقدمه

۶.....۲-۱ بحث عدم قطعیت

۷.....۳-۱ تلفیق مدل های توموگرافی سطحی

۹.....۴-۱ تحلیل خوشه ای طیفی

۱۱.....۵-۱ مطالعه مدل مصنوعی

۱۱.....۱-۵-۱ مدل آبخوان

۱۴.....۲-۵-۱ آزمایشها و وارون سازی داده های مصنوعی

۱۵.....۳-۵-۱ تحلیل خوشه ای طیفی

۱۷.....۶-۱ الگوریتم TSVM

۱۹.....۷-۱ نتایج پردازش و تحلیل مثال موردی

۲۱.....۸-۱ ضرورت و اهداف تحقیق

۲۲.....۹-۱ روش انجام تحقیق

۲۴.....۱۰-۱ سازماندهی مطالب

۲۵ فصل دوم: مبانی توموگرافی مقاومت ویژه و لرزه انکساری

۲۶.....۱-۲ روش توموگرافی الکتریکی

۲۷.....۱-۱-۲ مدل سازی داده های توموگرافی

۲۸.....۲-۱-۲ مدل سازی پیشرو

۲۸.....۳-۱-۲ مدل سازی وارون

۲۹.....۲-۲ روش توموگرافی لرزه ای

۳۰.....۱-۲-۲ مدل سازی مستقیم

۳۱.....۲-۲-۲ روش های بازسازی تکراری

۳۳.....۳-۲-۲ معادلات ریاضی مربوط به روش SIRT

۳۹.....۴-۲-۲ وارون سازی توموگرافی

۳۹.....	۵-۲-۲ توموگرافی اولین زمان رسید
۳۹.....	۳-۲ جمع‌بندی و نتیجه‌گیری
۴۱	فصل سوم: روش‌های طبقه‌بندی و خوشه‌بندی
۴۲.....	۱-۳ روش‌های طبقه‌بندی
۴۲.....	۱-۱-۳ ماشین بردار پشتیبان
۴۳.....	۲-۱-۳ جداسازی داده‌ها به صورت خطی
۴۷.....	۳-۱-۳ شرایط کاراش - کان - تاکر
۴۷.....	۴-۱-۳ طبقه‌بندی
۴۸.....	۵-۱-۳ داده‌های جدانشدنی به طور خطی
۵۳.....	۶-۱-۳ ماشین‌های بردار پشتیبان چند کلاسی
۵۴.....	۷-۱-۳ ماشین‌های بردار پشتیبان غیرخطی
۵۶.....	۸-۱-۳ انواع کرنل
۵۸.....	۹-۱-۳ انتخاب مدل
۵۹.....	۱۰-۱-۳ مروری بر مبانی روش TSVM
۶۴.....	۱۱-۱-۳ الگوریتم TSVM بر اساس خوشه‌بندی طیفی (TSVMSC)
۶۵.....	۲-۳ روش‌های خوشه‌بندی
۶۶.....	۱-۲-۳ قواعد کلی خوشه‌بندی
۶۸.....	۲-۲-۳ روش خوشه‌بندی k-means
۷۰.....	۳-۲-۳ خوشه‌بندی Fuzzy c-means یا FCM
۷۲.....	۴-۲-۳ الگوریتم گوستافسون کسل
۷۳.....	۵-۲-۳ مقایسه طبقه‌بندی و خوشه‌بندی
۷۵	فصل چهارم: فرمولاسیون مسئله
۷۶.....	۱-۴ ماتریس تشابه در تحلیل خوشه‌ای طیفی
۷۹.....	۲-۴ معادلات خوشه‌بندی طیفی
۸۱.....	۳-۴ روش شاخص دان
۸۲.....	۴-۴ رابطه تجربی هان و همکاران برای محاسبه سرعت لرزه‌ای
۸۲.....	۱-۴-۴ توصیف نمونه
۸۴.....	۲-۴-۴ اندازه‌گیری صوتی
۸۵.....	۳-۴-۴ اثرات تخلخل و محتوای رس بر روی سرعت‌های V_s و V_p
۹۱.....	۵-۴ رابطه تجربی نول برای محاسبه مقاومت ویژه
۹۱.....	۱-۵-۴ توصیف نمونه‌ها و فرآیند آزمایش
۹۳.....	۲-۵-۴ نتایج تجربی

۶-۴	معادلات روش میانگین متحرک تبدیل فوریه سریع	۹۴
۷-۴	جمع بندی و نتیجه گیری	۹۶
۹۷	فصل پنجم: وارون سازی داده های مصنوعی و صحرایی داده های لرزه ای و مقاومت ویژه	
۱-۵	مقدمه	۹۸
۲-۵	داده های توموگرافی لرزه انکساری	۱۰۱
۳-۵	داده های مقاومت ویژه الکتریکی	۱۰۵
۴-۵	مطالعه مدل مصنوعی	۱۰۸
۱-۴-۵	تولید و وارون سازی داده های مصنوعی با استفاده از روابط تجربی	۱۰۸
۵-۵	جمع بندی و نتیجه گیری	۱۱۰
	فصل ششم: تلفیق نتایج حاصل از وارون سازی داده های توموگرافی لرزه ای و مقاومت ویژه و اعتبار سنجی آن	
۱۱۱		
۱-۶	مقدمه	۱۱۲
۲-۶	تلفیق داده های توموگرافی لرزه انکساری و مقاومت ویژه الکتریکی حاصل از روابط تجربی	
۱۱۲	هان (۱۹۸۶) و نول (۱۹۹۶) و استفاده از الگوریتم TSVMSC در مدل مصنوعی	۱۱۲
۳-۶	تحلیل داده های صحرایی	۱۱۴
۱-۳-۶	داده های توموگرافی لرزه ای	۱۱۵
۲-۳-۶	داده های مقاومت ویژه الکتریکی	۱۱۶
۳-۳-۶	اعتبار سنجی نتایج حاصل از وارون سازی داده ها	۱۱۷
۴-۶	تلفیق نتایج حاصل از وارون سازی داده های واقعی توموگرافی لرزه ای و مقاومت ویژه الکتریکی	۱۱۹
۵-۶	اعمال الگوریتم TSVM بر اساس خوشه بندی طیفی (TSVMSC) بر روی نتایج حاصل از وارون سازی داده های ژئوفیزیکی	۱۲۳
۶-۶	اعتبار سنجی نتایج وارون سازی توموگرافی لرزه ای انکساری	۱۲۷
۷-۶	جمع بندی و نتیجه گیری	۱۳۰
۱۳۱	فصل هفتم: نتیجه گیری و پیشنهادات	
۱-۷	مقدمه	۱۳۲
۲-۷	نتیجه گیری	۱۳۲
۳-۷	پیشنهادهای	۱۳۴
۱۳۶	مراجع	

فهرست جداول

فصل اول: کلیات

جدول ۱-۱- مقایسه بین نتایج یادگیری بر روی مجموعه داده‌های رویترز با استفاده از روش‌های TSVM و TSVMSC..... ۲۰

جدول ۲-۱- مقایسه نتایج یادگیری TSVMSC با در نظر گرفتن تعداد خوشه مختلف..... ۲۱

فصل سوم: روش‌های طبقه‌بندی و خوشه‌بندی

جدول ۱-۳- کرنل‌های ماشین بردار پشتیبان ۵۷

فصل چهارم: فرمولاسیون مسئله

جدول ۱-۴- مدل‌های برازش شده در داده‌های سرعت تجربی ۸۷

فصل ششم: تلفیق نتایج حاصل از وارون‌سازی داده‌های توموگرافی لرزه‌ای و مقاومت‌ویژه و اعتبار سنجی آن

جدول ۱-۶- نگار گمانه حفاری شده در کف سد (راهنمای نگار گمانه در جدول ۲ ۶ آمده است)..... ۱۱۸

جدول ۲-۶- راهنمای نگار گمانه حفاری شده..... ۱۱۹

جدول ۳-۶- اندازه‌گیری سرعت‌های تراکمی (V_p) و برشی (V_s) به همراه دیگر پارامترها..... ۱۲۷

جدول ۴-۶- سرعت موج تراکمی (V_p) و برشی (V_s) و تخلخل متناظر نمونه‌ها..... ۱۲۸

فهرست اشکال

فصل اول: کلیات

- شکل ۱-۱- فلوچارت رهیافت تلفیق مدل یکپارچه و تخمین پارامتر هدف، هسته روش، الگوریتم خوشه‌ای FCM است که در پارامتر فضایی چندبعدی به دست آمده از مدل‌های ورودی موجود به کار رفته است..... ۸
- شکل ۱-۲- هسته اصلاح شده رهیافت تلفیق مدل با استفاده از خوشه‌بندی طیفی (شکل ۱-۱). توجه این که خوشه‌بندی در فضای ویژه از مدل‌های ورودی انجام شده است..... ۱۰
- شکل ۱-۳- الف) توزیع تخلخل و ب) محتوای رس یک مدل آبخوان غیر پیوسته مصنوعی..... ۱۲
- شکل ۱-۴- الف) سرعت موج P و ب) مدل مقاومت ویژه الکتریکی به دست آمده از تغییر پتروفیزیکی خواص آبخوان نشان داده شده در شکل ۱-۳..... ۱۳
- شکل ۱-۵- داده‌های مصنوعی وارون‌سازی شده الف) سرعت موج P حاصل از تست لرزه نگاری انکساری و ب) مدل مقاومت ویژه الکتریکی به دست آمده از تست مقاومت ویژه و ژئوالکتریک جریان مستقیم مصنوعی بر روی مدل آبخوان نشان داده شده در شکل ۱-۴. توجه این که به طور سیستماتیکی کیفیت مدل با عمق کاهش می‌یابد (مقایسه با شکل ۱-۴)..... ۱۵
- شکل ۱-۶- مدل منطقه‌بندی به دست آمده از تحلیل خوشه‌ای طیفی از مدل‌های وارون‌سازی شده شکل ۱-۵. زون‌ها (طبقه‌ها) با رنگ مشخص شده‌اند..... ۱۶
- شکل ۱-۷- توزیع عضویت خوشه به دست آمده برای هر کلاس از مدل منطقه‌بندی شده (شکل ۱-۶). درجه بالای عضویت، قابلیت اعتماد بالای یک پارامتر مدل (سلول مدل) را در کلاس مشخص، نشان می‌دهد..... ۱۶

فصل دوم: مبانی توموگرافی مقاومت ویژه و لرزه انکساری

- شکل ۱-۲- ترتیب بلوک‌های مستطیلی مورد استفاده در مدل‌سازی سه..... ۲۹
- شکل ۲-۲- توصیف شماتیک دو الگوریتم ART و SIRT..... ۳۳
- شکل ۲-۳- فلوچارت مربوط به روش SIRT پارامتر S_0 تخمین اولیه کندی، S_q کندی در جریان که تصحیحات روی آن انجام می‌شود، T_{calc} زمان رسید محاسبه شده و T_{meas} زمان رسید مشاهده شده است. پارامتر S_q به طور مرتب اصلاح می‌شود تا جایی که اختلاف T_{meas} و T_{calc} به حد مجاز برسد..... ۳۸

فصل سوم: روش‌های طبقه‌بندی و خوشه‌بندی

- شکل ۱-۳- دو مجموعه مجزا به صورت خطی با ابرصفحه‌های جداکننده. ابرصفحه جداکننده در سمت راست (خط ضخیم) در فاصله بیشینه از نزدیک‌ترین نقاط قرار می‌گیرد. خطوط نازک مرز یا حاشیه را مشخص می‌کنند..... ۴۳

شکل ۳-۲- خطوط $H1$ و $H2$ ابرصفحه‌های کانونی هستند. حاشیه، فاصله عمودی بین ابرصفحه جداکننده $(g(x)=0)$ و یک ابرصفحه گذرنده از نزدیک‌ترین نقاط (با دایره مشخص شده‌اند) است. این‌ها بردارهای پشتیبان نامیده می‌شوند..... ۴۵.

شکل ۳-۳- ابرصفحه جدا کننده خطی برای داده غیر قابل تفکیک..... ۴۹.

شکل ۳-۴- داده تفکیک پذیر به طور خطی (چپ) و داده تفکیک ناپذیر (راست، $C=20$ که C پارامتر منظم سازی است)..... ۵۲.

شکل ۳-۵- بیشینه مرز ابر صفحه. مثال‌های مثبت و منفی با $+/-$ و مثال‌های تست با نقطه علامت گذاری شده‌اند. خط تیره حل‌های SVM است. خطوط پر، طبقه بندی ترانس داکتیو را نشان می‌دهد..... ۶۳.

شکل ۳-۶- بخش‌های معمول یک سیستم خوشه بندی..... ۶۷.

فصل چهارم: فرمولاسیون مسئله

شکل ۴-۱- مثالی از تحلیل خوشه‌ای طیفی با استفاده از مثال مصنوعی. الف) دو مدل ورودی جدا از هم در مجاورت یکدیگر که توزیع‌های پارامتر فیزیکی مختلف و جدا از هم را نشان می‌دهد. ب) پارامتر فضایی دو بعدی به دست آمده از مدل‌های ورودی. ج) ماتریس تشابه، فواصل اقلیدوسی وزن دار شده را بین نقاط در فضای پارامتری دو بعدی نرمال شده، نشان می‌دهد. د) اولین هشت مقدار ویژه و چهار بردار ویژه از ماتریس به دست آمده در قسمت (ج). ه) توزیع عضویت فازی برای یک جواب چهار خوشه بر اساس بردار ویژه به دست آمده در قسمت (د)..... ۷۸.

شکل ۴-۲- تغییرات محتوای رس و تخلخل برای ۷۵ نمونه ماسه سنگ. تغییرات تخلخل از ۲ تا ۳۰ درصد و تغییرات محتوای رس از ۰ تا ۵۰ درصد می‌باشد. داده‌ها نشان می‌دهند که ماسه سنگ‌ها با محتوای رس بالا، تخلخل پایینی دارند..... ۸۳.

شکل ۴-۳- الف) موج تراکمی اندازه گیری شده و ب) سرعت موج برشی اندازه گیری شده در برابر تخلخل برای ۷۵ نمونه ماسه سنگ در فشار ۴۰ مگا پاسکال و ۱ مگا پاسکال..... ۸۶.

شکل ۴-۴- سرعت موج تراکمی V_p در فشار ۴۰ و ۱ مگا پاسکال که از معادله زمان- میانگین به دست آمده است. الف) انحراف نسبی در برابر تخلخل که با پراکندگی بالا نشان داده شده است. ب) انحراف نسبی یک همبستگی واضح را با محتوای رس نشان می‌دهد..... ۸۸.

شکل ۴-۵- سرعت‌های موج تراکمی V_p و برشی V_s که با معادله برازش شده است. الف) انحراف نسبی V_p در برابر تخلخل، با پراکندگی بالا. ب) انحراف نسبی V_p در برابر محتوای رس. ج) انحراف نسبی V_s در برابر تخلخل، با پراکندگی بالا. ب) انحراف نسبی V_s در برابر محتوای رس..... ۸۹.

شکل ۴-۶- شکل مفهومی هندسه تراکم برای ماسه خالص، رس خالص و ترکیب رس - ماسه در شرایط پارامترهای تجربی: تخلخل (یا درجه فشردگی) و محتوای رس. پارامتر ϕ محتوای رس و ψ تخلخل ماسه خالص که در متن آمده است..... ۹۲

شکل ۴-۷- روابط پیش‌بینی شده و مشاهده شده بین تخلخل و درصد وزنی رس برای ترکیب ماسه - کائولینیت در فشار همه جانبه پایین. مقادیر پیش‌بینی شده بر اساس مدل میکرو هندسی برای ترکیب ماسه - رس توسط ماریون و همکاران (۱۹۹۲) بیان شده است..... ۹۳

شکل ۴-۸- الف) ثابت دی الکتریک بر حسب تخلخل، و ب) رسانندگی بر حسب تخلخل، برای نمونه‌های ماسه کائولینیت در سطح‌های مختلف از اشباع شدگی. علامت‌های تو خالی خواص الکتریکی اندازه‌گیری شده در شرایط اتاق خشک، علامت‌های تو پر خواص الکتریکی اندازه‌گیری شده بعد از این که نمونه‌ها با آب مقطر اشباع شدند را نشان می‌دهند ($\sigma_w \approx 4 \times 10^{-4} \text{ S/m}$). فرکانس‌های اندازه‌گیری شده ۱۰۰ کیلوهرتز، ۱ مگاهرتز و ۱۰ مگاهرتز هستند. فلش‌ها جهت افزایش محتوای رس را نشان می‌دهند..... ۹۴

فصل پنجم: وارون‌سازی داده‌های مصنوعی و صحرایی داده‌های لرزه‌ای و مقاومت‌ویژه

شکل ۵-۱- نقشه زمین‌شناسی و موقعیت گمانه‌های توموگرافی لرزه‌ای انکساری و سونداژهای برداشت مقاومت‌ویژه در محدوده مورد مطالعه..... ۱۰۰

شکل ۵-۲- تعیین اولین زمان رسید رکوردهای توموگرافی لرزه‌ای انکساری..... ۱۰۱

شکل ۵-۳- نتایج وارون‌سازی توموگرافی لرزه انکساری بین دو گمانه k224 و kch-19..... ۱۰۳

شکل ۵-۴- نتایج وارون‌سازی توموگرافی لرزه انکساری بین دو گمانه k224 و k210..... ۱۰۵

شکل ۵-۵- تحلیل یک بعدی سونداژهای مقاومت ویژه H16, H17, H18, H19..... ۱۰۷

شکل ۵-۶- نتایج حاصل از وارون‌سازی داده‌های مقاومت‌ویژه الکتریکی برداشت شده در طول پروفیل ۲۰۰ متر با استفاده از نرم‌افزار RES2DINV با خطای RMS برابر ۴/۷ درصد..... ۱۰۷

شکل ۵-۷- مقطع مقاومت‌ویژه الکتریکی در عمق پایین‌تر از ۱۴ متر (این عمق در زیر سطح ایستابی قرار دارد)..... ۱۰۸

شکل ۵-۸- محاسبه سرعت لرزه‌ای و مقاومت‌ویژه الکتریکی با استفاده از تخلخل و محتوای رس (روابط تجربی هان و همکاران (۱۹۸۶) و نول و همکاران (۱۹۹۶))..... ۱۰۹

فصل ششم: تلفیق نتایج حاصل از وارون‌سازی داده‌های توموگرافی لرزه‌ای و مقاومت‌ویژه و اعتبار سنجی آن

شکل ۶-۱- فلوچارت الگوریتم TSVMSC..... ۱۱۳

- شکل ۲-۶- نتایج حاصل از وارون سازی داده های مصنوعی توموگرافی لرزه انکساری (سمت راست) و مقاومت ویژه الکتریکی (سمت چپ) (استفاده از روابط تجربی هان (۱۹۸۶) و نول (۱۹۹۶))..... ۱۱۳
- شکل ۳-۶- تلفیق نتایج حاصل از وارون سازی داده های مصنوعی توموگرافی لرزه انکساری و مقاومت ویژه الکتریکی با استفاده از الگوریتم TSVMSC..... ۱۱۴
- شکل ۴-۶- نتایج وارون سازی توموگرافی لرزه های انکساری برای اعماق پایین تر از ۱۴ متری سطح زمین..... ۱۱۶
- شکل ۵-۶- نتایج وارون سازی داده های مقاومت ویژه الکتریکی برای اعماق پایین تر از ۱۴ متری سطح زمین..... ۱۱۷
- شکل ۶-۶- شاخص دان که برای محاسبه تعداد خوشه های بهینه مورد استفاده قرار گرفته است. در این شاخص حداکثر مقدار به دست آمده برای خوشه ها، به عنوان تعداد خوشه بهینه در نظر گرفته می شود. محور افقی تعداد خوشه ها و محور عمودی مقدار شاخص را نشان می دهد..... ۱۲۱
- شکل ۷-۶- مدل منطقه بندی به دست آمده از اعمال الگوریتم میانگین k بر روی مدل های وارون سازی شده. زون ها (طبقه ها) با رنگ مشخص شده اند..... ۱۲۱
- شکل ۸-۶- مدل منطقه بندی به دست آمده از اعمال الگوریتم میانگین فازی بر روی مدل های وارون سازی شده. زون ها (طبقه ها) با رنگ مشخص شده اند..... ۱۲۲
- شکل ۹-۶- مدل منطقه بندی به دست آمده از اعمال الگوریتم گوستافسون- کسل فازی بر روی مدل های وارون سازی شده. زون ها (طبقه ها) با رنگ مشخص شده اند..... ۱۲۲
- شکل ۱۰-۶- مدل منطقه بندی به دست آمده از اعمال خوشه بندی طیفی بر روی مدل های وارون. زون ها (طبقه ها) با رنگ های آبی تا قرمز مشخص شده اند..... ۱۲۴
- شکل ۱۱-۶- مدل منطقه بندی به دست آمده از الگوریتم TSVMSC بر روی مدل های وارون. زون ها (طبقه ها) با رنگ های آبی تا قرمز مشخص شده اند..... ۱۲۶
- شکل ۱۲-۶- نمونه آزمایشگاهی برای تعیین سرعت موج لرزه ای که از عمق ۴۸/۴۱ تا ۴۸/۸۵ متری گرفته شد..... ۱۲۸
- شکل ۱۳-۶- نمودار تعیین سرعت موج تراکمی و برشی با استفاده از تخلخل موجود..... ۱۲۹
- شکل ۱۴-۶- نمودار تعیین سرعت موج تراکمی و برشی با استفاده از تخلخل موجود..... ۱۲۹

فصل اول

کلیات

۱-۱ مقدمه

رهیافت‌های چند روشی^۱ به طور گسترده در کاهش عدم قطعیت تحلیل‌ها و تفسیرهای ژئوفیزیک سطحی، به کار برده می‌شوند [۱]. عدم قطعیت، یک پارامتر مرتبط با نتیجه اندازه‌گیری است که پراکندگی مقادیری که می‌توانند به شکل منطقی به اندازه کمیت اندازه‌گیری شونده نسبت داده شود را مشخص می‌کند. رویافت‌های رایج، تفسیر توأم مدل‌های خواص فیزیکی مختلف است که به طور مجزا از توموگرافی^۲‌های زیر سطحی مختلف، مشتق می‌شوند. تفسیر توأم مدل‌های چندگانه، به طور ذاتی مبتنی بر تجربه مفسر بوده و دانش محور است [۲].

در طی دهه اخیر چندین روش عملی برای تلفیق توموگرافی مدل‌های ژئوفیزیکی در یک مدل چند پارامتری بدست آمده است. انواع روش‌های تلفیق عبارتند از: مدل منطقی بولین، مدل‌های منطق فازی، وارون‌سازی توأم مدل‌های ژئوفیزیکی، روش تلفیق بصری ساده، از قبیل نمودارهای سه تایی^۳ [۳] یا همپوشانی مدل‌ها در فرم لایه‌های نیمه شفاف^۴ [۴]. به هر حال کاربرد روش تلفیق بصری محدود می‌باشد و این محدودیت به گونه‌ای است که تعداد مدل‌ها نباید بیشتر از چهار شود.

روش‌های پایدار و قدرتمند برای تلفیق مدل‌های بازسازی شده توموگرافی بر اساس تحلیل آماری چند متغیره^۵ وجود دارد. خصوصاً روش‌های تحلیل خوشه‌ای^۶، برای مثال الگوریتم‌های k-means یا روش خوشه‌بندی فازی c-means (FCM)^۷، [۵]، [۶]، برای تلفیق مدل‌های توموگرافی درون‌چاهی^۸ بسیار مفید می‌باشند. الگوریتم خوشه‌ای، نمونه‌ها را در یک فضای چندبعدی تحلیل کرده و به زیر مجموعه‌های از پیش تعریف شده تقسیم‌بندی می‌کند. سپس این الگوریتم، مدل‌های

¹ Multimethod approaches

² Tomographic

³ Ternary plots

⁴ Semi-transparent layers

⁵ Multivariate statistical analysis

⁶ Partitioning cluster analyses

⁷ Fuzzy c-means

⁸ Crosshole tomographic

زیرسطحی چند پارامتری را بر اساس مدل‌های ورودی چندگانه با روابط نامعلوم و غیرخطی متغیر از نظر فضایی، ایجاد می‌کند [۷]. روابط پیچیده بین پارامترهای فیزیکی مختلف، به طور نوعی وابسته به محیط‌های رسوبی سطحی تحکیم نیافته می‌باشد [۸]. به هر حال قبل از مشخص کردن یک مدل توموگرافی در تحلیل خوشه‌ای، کیفیت مدل باید به دقت ارزیابی شود. خطاهای ساختاری مدل، مثل خطاهای مصنوعی، ممکن است روی نتایج خوشه‌بندی تأثیر بگذارد [۲].

اپستین و دوقرتی^۱ (۱۹۹۶) [۹] از خوشه‌بندی k-means برای منطقه‌بندی سفره آبخوان و تخمین مقادیر قابلیت گذردهی^۲، استفاده کردند. فیچنر و دیتریچ^۳ (۱۹۹۷) [۱۰]، دیتریچ و همکاران^۴ (۱۹۹۸) [۱۱]، ترونیکه^۵ و همکاران (۲۰۰۲) [۱۲]، و ترونیکه و همکاران (۲۰۰۴) [۱۳]، از تحلیل خوشه‌ای k-means در تلفیق مدل‌های مختلف ژئوفیزیکی مقطعی در بهبود نمایش آبخوان‌های رسوبی و در نتیجه اکتشاف آن‌ها، استفاده کردند. بروچلر^۶ و همکاران (۲۰۱۲) [۱۴]، خوشه‌بندی k-means را برای تلفیق نتایج توموگرافی لرزه‌ای و هیدرولیکی برای بهبود توصیف هیدرولوژی آبخوان رسوبی، به کار گرفتند.

در سال‌های اخیر، الگوریتم‌های خوشه‌ای فازی^۷ محبوب‌تر از سایر روش‌ها برای خوشه‌بندی نتایج به کار برده شده است. به همین نحو در خوشه‌بندی^۸، الگوریتم‌های فازی می‌توانند در تلفیق مدل‌های چندگانه با گروه بندی نمونه‌ها در یک فضای چندبعدی، به کار روند. به هر حال مفهوم فازی به عضویت جزئی یک نمونه مشخص اجازه می‌دهد که در حالی که یک نمونه در یک خوشه معین گروه بندی شده باشد، اما در یک زمان می‌تواند عضو جزئی از دیگر خوشه‌ها باشد. مفهوم عضوهای

¹ Eppstein and Dougherty

² Transmissivity

³ Fechner and Dietrich

⁴ Dietrich

⁵ Troncke

⁶ Brauchler

⁷ Fuzzy partitioning cluster algorithms

⁸ Partitioning clustering

جزئی خوشه، ناهمگنی ساختاری همه مدل‌های ورودی را تلفیق کرده و آن را در یک مفهوم فازی توصیف می‌کند. سپس کمی کردن قابلیت اطمینان ساختارهای کشف شده در مدل ژئوفیزیکی چند پارامتری منطقه‌بندی شده^۱ را سبب می‌شود [۲].

پاشه^۲ و همکاران (۲۰۰۶) [۷] یک رهیافت را بر اساس تحلیل خوشه‌ای FCM برای تلفیق مدل‌های مختلف توموگرافی درون‌چاهی به منظور منطقه‌بندی آبخوان، توسعه دادند. اطلاعات عضویت در ساختارهای کشف شده در مدل‌های چند پارامتری منطقه‌بندی شده، به صورت رنگی نشان داده شده است. در عین حال، اطلاعات تلفیق شده عضویت فازی می‌تواند در پیش‌بینی توزیع فضایی پارامترهای هدف معلوم، استفاده شود. پاشه و همکاران (۲۰۰۶) از این رهیافت برای به دست آوردن توزیع فضایی پرتو گامای طبیعی و هدایت هیدرولیکی بر اساس سه مدل توموگرافی عرضی، چاه نگاری پرتو گاما^۳ و آزمایش‌های اسلاگ^۴ (ضربه)، استفاده کردند. این رهیافت بعداً برای تلفیق از کنار هم قرار دادن جزئی مدل‌های توموگرافی، توسعه پیدا کرد.

با این که این رهیافت پتانسیل خوبی در تلفیق مدل‌های توموگرافی ناشی از برداشت‌های ژئوفیزیکی می‌باشد، اما تنها مطالعات کمی هستند که تحلیل خوشه‌ای را برای تلفیق مدل‌های توموگرافی به دست آمده از برداشت‌های سطحی ژئوفیزیکی، به کار گرفته‌اند. برای مثال، گانترو و همکاران^۵ (۲۰۰۶) [۱۵] از آنالیز خوشه‌ای FCM در تلفیق مقاومت ویژه الکتریکی دو بعدی و مدل‌های سرعت موج P استفاده کردند. پاشه و همکاران (۲۰۰۷) [۱۶] تحلیل خوشه‌ای FCM را برای اکتشاف مسیر جریان آب بر اساس مدل سه بعدی مقاومت ویژه، توموگرافی‌هایی با چاه‌های متقابل و هدایت هیدرولیکی، به کار بردند. به هر حال در سال‌های اخیر برخی تلاش‌ها در جهت تلفیق مدل‌های

¹ Zoned

² Paasche

³ Gamma radiation logging

⁴ Slug tests

⁵ Günther

توموگرافی سطحی با استفاده از الگوریتم خوشه‌ای بوده است. یکی از مشکلات کاربرد الگوریتم‌های خوشه‌ای در تلفیق مدل‌های توموگرافی ژئوفیزیکی، تفاوت بین روش‌های برداشت می‌باشد [۲].

به منظور رسیدن به یک رهیافت مناسب برای تلفیق مدل‌های توموگرافی ناشی از برداشت‌های ژئوفیزیکی سطحی، رهیافت پاشه و همکاران (۲۰۰۶) [۷] توسط هاجمولر و پاشه^۱ (۲۰۱۳) [۲] بازنگری شده است. در عمل هسته روش با تغییراتی در الگوریتم خوشه‌ای FCM با یک الگوریتم خوشه‌بندی طیفی، بهبود یافت. یعنی یک روش خوشه‌بندی در تحلیل ساختار مدل محلی با تحلیل ساختار مدل کلی در همه الگوریتم‌های خوشه‌ای تقسیم‌بندی کننده، مقایسه شده است [۱۷]. بعد از بازنگری رهیافت پاشه و همکاران (۲۰۰۶)، خانواده الگوریتم‌های خوشه‌ای طیفی مشخص شد و با انتخاب یک الگوریتم مناسب، عاملیت^۲ و عملکرد آن در تلفیق مدل پاشه و همکاران (۲۰۰۶) ثابت می‌شود [۷]. برای این منظور، الگوریتم در یک مطالعه آبخوان رسوبی مصنوعی با داده‌های مقاومت‌ویژه دو بعدی، لرزه انکساری^۳، به کار برده شد. بعد از این که رهیافت به نتایج قابل اطمینانی رسید، داده‌های سرعت موج P و مدل‌های مقاومت‌ویژه با هم تلفیق شدند. داده‌های واقعی برای یک دشت سیلابی رسوبی در شرق آلمان بودند [۲].

در این فصل بعد از بررسی چگونگی بهبود روش FCM با استفاده از روش خوشه‌بندی طیفی، تلفیق داده‌های دو روش مقاومت‌ویژه و لرزه انکساری در مدل‌های مصنوعی و صحرایی مورد مطالعه قرار گرفته است. همچنین در این فصل الگوریتم بهبود یافته TSVMSC (ماشین بردار پشتیبان تبدیلی بر اساس خوشه‌بندی طیفی)^۴ با ذکر یک مثال موردی مطرح گردیده است. در این تحقیق تصمیم بر این است که به جای استفاده از تحلیل خوشه‌بندی طیفی^۵ (SC) در هسته تلفیق نتایج

¹ Hachmöller and Paasche

² Functionality

³ Refraction seismic

⁴ Transductive Support Vector Machine Algorithm Based on Spectral Clustering

⁵ Spectral clustering

حاصل از وارون سازی توموگرافی لرزه انکساری و مقاومت ویژه، از الگوریتم پایدار و با دقت TSVMSC استفاده شود.

۲-۱ بحث عدم قطعیت

عدم قطعیت، یک پارامتر مرتبط با نتیجه اندازه گیری است که پراکندگی مقادیری که می توانند به شکل منطقی به اندازه کمیت اندازه گیری شونده نسبت داده شود را مشخص می کند. خطا متفاوت از عدم قطعیت است. خطا به شکل اختلاف بین یک نتیجه تک و مقدار واقعی تعریف می شود. در عمل خطای اندازه گیری مشاهده شده، اختلاف بین مقدار مشاهده شده و مقدار مرجع می باشد. بدین ترتیب خطا چه تئوری، چه مشاهده شده، یک مقدار تک است. در حالی که عدم قطعیت شکل یک گستره یا بازه را به خود می گیرد. عموماً نمی توان از عدم قطعیت برای تصحیح یک نتیجه اندازه گیری استفاده کرد؛ در حالی که مقدار خطای معین می تواند به عنوان یک تصحیح برای نتیجه اندازه گیری به کار گرفته شود. پارامترهای عدم قطعیت در این مطالعه می تواند زمین شناسی و برداشت و همچنین وارون سازی داده ها باشد. در بحث برداشت ژئوفیزیکی و زمین شناسی می توان گفت اگر عملیات برداشت داده چندین بار تکرار شود، مؤلفه نوفه های تصادفی موجود در داده ها را می توان مستقیماً تعیین نمود. اما این حالت به ندرت در ژئوفیزیک اتفاق می افتد. اصولاً تخمین عدم قطعیت داده ها با تحلیل خطای عملیات شروع می شود، مانند ثبت نوفه زمینه [۱۸]، اما تحلیل خطا برای هر عملیات ژئوفیزیکی کاملاً پیچیده و شامل تعداد زیادی مجهولات می باشد. هدف از معکوس سازی داده های ژئوفیزیکی، استنباط کمی اطلاعات درباره زمین از تعداد محدودی مشاهده غیر مستقیم و حاوی نوفه می باشد. در عمل معمولاً با بهینه نمودن یک تابع اختلاف تعریف شده برای داده ها از عدم قطعیت آن ها صرف نظر می شود. اما این سؤال پیش خواهد آمد که تابع هدف چه موقع بهینه می شود.

برای تخمین عدم قطعیت داده‌ها و ساخت بازه‌های اطمینان برای پارامترهای مدل در مسائل معکوس خطی ابتدا با تنظیم تیخونف و استفاده از روش منحنی L برای تعیین پارامتر تنظیم [۱۹]، پارامترهای مدل حاصل می‌شوند، سپس بردار باقیمانده حاصل از این مدل‌ها برای تخمین عدم قطعیت داده‌ها استفاده می‌شود، این تخمین خطا برای ساخت مجموعه‌ای از مدل‌ها که داده‌ها را پیش بینی می‌کنند و همچنین مدل‌هایی که محدودیت‌هایی در بازه پارامترهای مدل دارند، به کار می‌رود. هنگامی که یک مدل معقول به دست آمد، از اطلاعات اولیه برای ساخت بازه‌های اطمینان مدل نهایی استفاده می‌شود. روش‌های ارائه شده برای تخمین عدم قطعیت داده‌ها، دقت بالایی دارند.

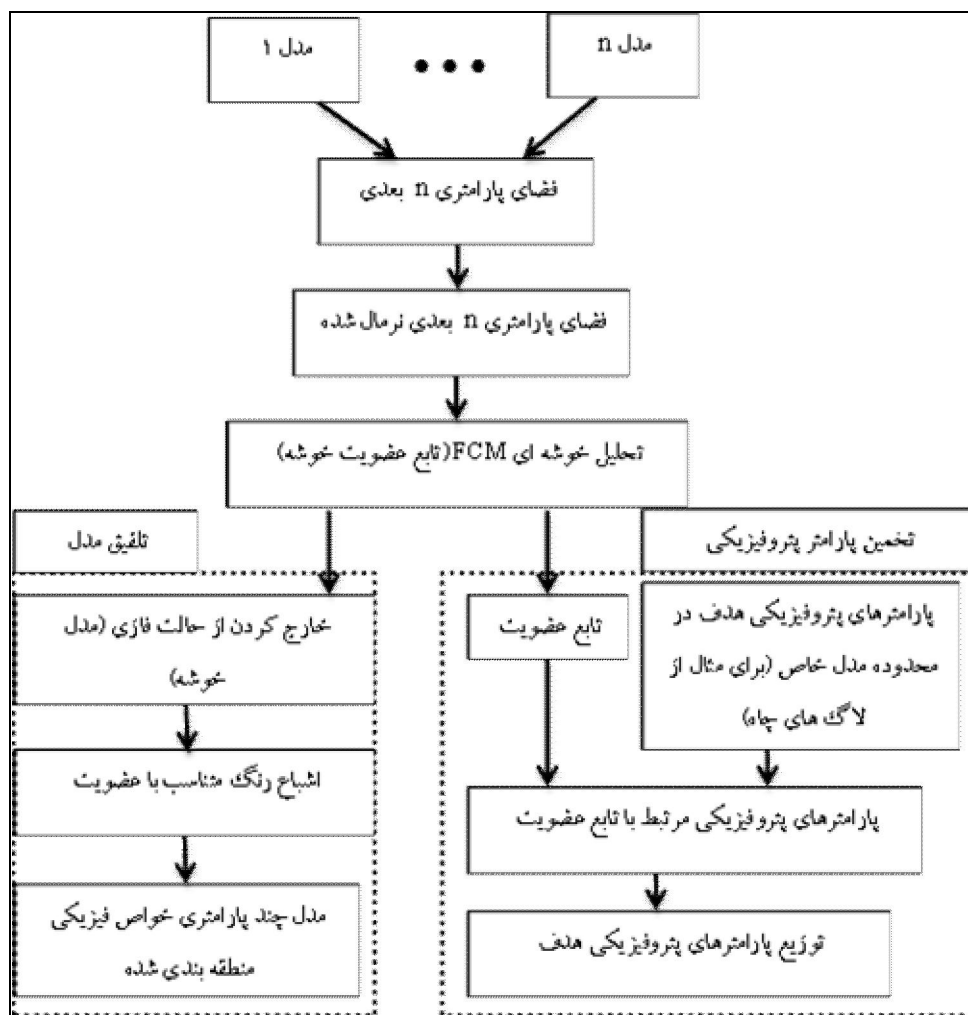
۳-۱ تلفیق مدل‌های توموگرافی سطحی

رهیافت تلفیق مدل‌ها توسط پاشه و همکاران (۲۰۰۶) توسعه پیدا کرد (فلوچارت در شکل ۱-۱ نشان داده شده است) [۷]. هر یک از n مدل ورودی در امتداد یک محور قائم از فضای پارامتری ترسیم شده است. یک نمونه در فضا به وسیله بردار مکانی n -جزء^۱ تعریف می‌شود. تعداد نمونه‌ها در فضای پارامتری برابر تعداد سلول‌های استفاده شده در هر یک مدل‌های ورودی می‌باشد. اگر مش‌های مختلفی در فرآیند تولید مدل استفاده شود، مدل‌ها باید بعد از وارون‌سازی با استفاده از یک مش رایج پردازش شوند تا پیچیدگی اطلاعات به دست آمده از همه مدل‌ها برطرف شود. به عبارتی مدل‌سازی با یک مش مشخص صورت گیرد. در مدل‌های مختلف، مقیاس گذاری محورها در فضای پارامتری نرمالایز^۲ شده است. با این کار همه مدل‌های ورودی بین ۰ و ۱ نرمالایز می‌شود. سپس الگوریتم خوشه‌ای FCM، همه داده‌ها را در فضای پارامتری n بعدی در داخل c زیر مجموعه (خوشه‌ها) و بر

^۱ N-element position vector

^۲ Normalize

اساس فاصله اقلیدوسی^۱ شان نسبت به یکدیگر، گروه بندی می‌کند. نتیجه این، یک توصیف فازی از همه مدل‌های ورودی در تشکیل عضویت جزئی از هر نمونه (سلول مدل) در c خوشه می‌باشد و مقادیر نوعی (میانگین) برای هر کلاس با یک بردار مکان n بعدی تعریف می‌شود (شکل ۱-۱).



شکل ۱-۱- فلوچارت رهیافت تلفیق مدل یکپارچه و تخمین پارامتر هدف. هسته روش، الگوریتم خوشه‌ای FCM است که در پارامتر فضایی چندبعدی به دست آمده از مدل‌های ورودی موجود به کار رفته است [۲]

در بازسازی مدل توموگرافی، به طور خطی در عمق، تباین کاهش می‌یابد. برای مثال اگر دو آنومالی یکسان در دو عمق مختلف قرار گرفته باشند، یعنی در توموگرافی مقاومت‌ویژه الکتریکی سطحی،

¹ Euclidian distance

آنومالی عمیق تر در صورت استفاده از هموارسازی استاندارد و الگوریتم‌های وارون‌سازی مقید شده از نظر میرایی [۲۰]، با تباین پایین به تصویر کشیده می‌شود.

۴-۱ تحلیل خوشه‌ای طیفی

خوشه‌بندی طیفی - مانند دیگر روش‌های خوشه‌بندی - می‌تواند در هر زمینه علمی که با داده‌های تجربی سر و کار دارد و زیر مجموعه‌های داده‌های مشابه را شناسایی می‌کند، به کار برده شود. تحلیل خوشه‌ای طیفی در دهه اخیر محبوبیت زیادی پیدا کرده است و امروزه به طور گسترده‌ای در مسائل مختلف داده‌کاوی^۱ با کاربردهای متفاوت استفاده می‌شود، برای مثال، تقسیم‌بندی تصویر به قطعات مختلف^۲، استخراج ماتریس‌های توصیف کننده ژن^۳، تشخیص دست خط^۴ و صوت یا صدا [۲۱]، [۲۲]، [۲۳]، [۲۴]، [۱۷]، [۲۵].

در به کار بردن روش خوشه‌بندی طیفی برای تلفیق مدل‌های ژئوفیزیکی سطحی، تنها لازم است هسته رهیافت تلفیق مدل جایگزین شود (شکل ۱-۱). هسته جدید مبتنی بر روش خوشه‌بندی طیفی در شکل ۱-۲ نشان داده شده است. پرداختن به مقیاس داده‌ها، به طور کلی آسان‌تر است تا پرداختن به الگوریتم‌های خوشه‌ای تقسیم‌بندی کننده مبتنی بر سنجش فاصله اقلیدوسی مثل FCM. بر اساس n مدل ورودی، یک گراف غیر جهت‌دار^۵ (مثال [۲۶]) تولید می‌شود و به وسیله ماتریس تشابه که ناهمگنی ساختاری همه مدل‌های ورودی را نشان می‌دهد، توصیف می‌شود. به طور پی در پی^۶ یک ماتریس لاپلاسیان گراف^۷ از گراف غیر جهت‌دار، تعیین می‌شود. سپس مقادیر ویژه^۱ (طیف) و

^۱ Data mining

^۲ Image segmentation

^۳ Mining gene-expression matrices

^۴ Handwriting

^۵ Undirected graph

^۶ Consecutively

^۷ Graph Laplacian matrix

متناظر آن بردارهای ویژه^۲ از ماتریس لاپلاسین گراف محاسبه می‌شود. بزرگ‌ترین مقادیر ویژه طیف تعیین می‌شوند و نظیر به نظیر آن، بردارهای ویژه به عنوان ورودی برای تحلیل خوشه‌ای انتخاب می‌شوند. در فضای متعامد از این بردارهای ویژه، نمونه‌هایی که به وسیله بردارهای ویژه تعیین محل می‌شوند، با استفاده از یک الگوریتم خوشه‌ای فازی، خوشه‌بندی می‌شوند. سرانجام، اطلاعات عضویت خوشه در محدوده مدل اصلی دوباره به تصویر درآورده می‌شود؛ تا به یک مدل چند پارامتری منطقه‌بندی شده^۳ دست یافت (شکل ۱-۱ را ببینید).



شکل ۱-۲- هسته اصلاح شده رهیافت تلفیق مدل یکپارچه با استفاده از خوشه‌بندی طیفی (شکل ۱-۱). توجه این که خوشه‌بندی در فضای ویژه از مدل‌های ورودی انجام شده است [۲].

¹ Eigenvalues

² Eigenvectors

³ Zonal

۵-۱ مطالعه مدل مصنوعی

۱-۵-۱ مدل آبخوان

برای بررسی رهیافت تلفیق مدل در یک وضعیت کنترل شده، از مثال مصنوعی یک آبخوان ناپیوسته ناهمگن استفاده شده است. برای شروع، یک مدل تخلخل واقعی با گسترش ۱۲۰ متر در جهت x و ۲۰ متر در جهت z تولید و از یک شبکه همگن با تقسیم‌بندی ۰/۱ متر استفاده شده است. توزیع تخلخل دو بعدی به وسیله یک روش شبیه‌سازی تصادفی غیر شرطی^۱ [۲۷]، [۲۸]؛ با استفاده از یک تابع هم پراش نمایی^۲، به دست آمده است که یک مورد خاص از مدل هم پراش کارمن^۳ است که چندبعدی، مقید شده و مقیاس ثابت می‌باشد. این مدل در بسیاری از موارد در توصیف ساختارهای ناهمسان‌گرد ناهمگن از قبیل آبخوان‌های رسوبی، استفاده شده است [۷]، [۲۹] و [۳۰]. مدل تخلخل دارای یک تابع چگالی احتمال گوسی^۴ است. طول همبستگی افقی و عمودی به ترتیب ۲۰۰ متر و ۲۰ متر در نظر گرفته شده است [۳۱] که یک انتخاب مناسب برای شبیه‌سازی آبخوان ناپیوسته می‌باشد [۳۲].

برای شبیه‌سازی یک آبخوان تشکیل شده از ماسه با مقدار رس متفاوت، زون‌هایی با تخلخل پایین و با محتوای رس بالاتر از ۱۸ درصد، در نظر گرفته شده‌اند. با افزایش محتوای رس به حدود ۲۰ درصد و با توزیع رس در بین فضای خلل و فرج ماسه، تخلخل کاهش می‌یابد [۳۳]. تغییرات فضایی تخلخل و محتوای رس مدل آبخوان مصنوعی مورد نظر که از مدل‌سازی تصادفی یا آماری^۵ به دست آمده در شکل ۳-۱ نشان داده شده است.

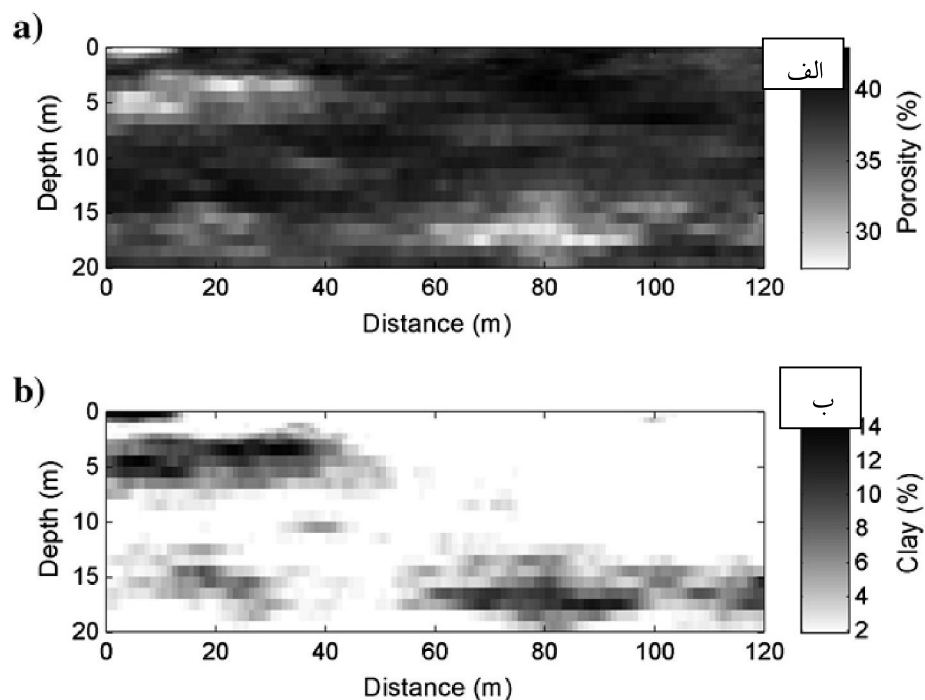
¹ Unconditional stochastic

² Exponential auto-covariance

³ Von Kármán

⁴ Gaussian probability density function

⁵ Stochastic Modeling



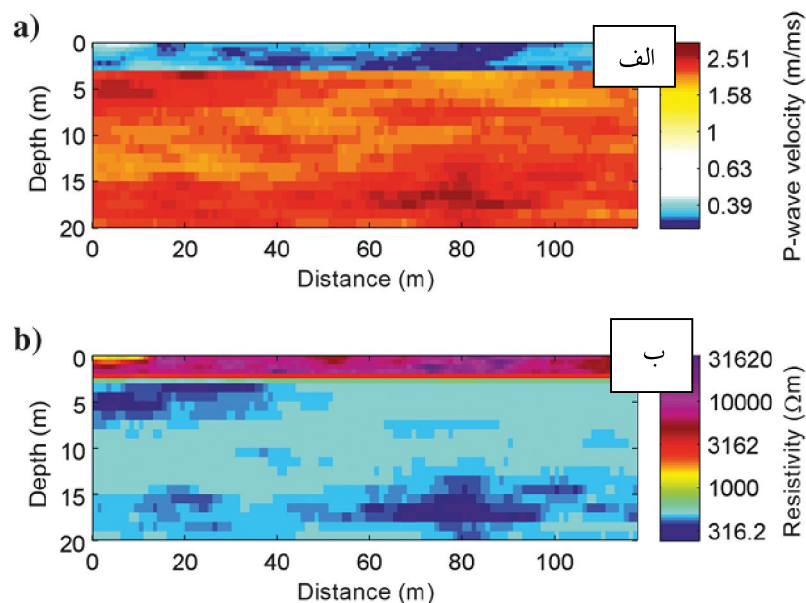
شکل ۱-۳- الف) توزیع تخلخل و ب) محتوای رس یک مدل آبخوان غیر پیوسته مصنوعی [۲]

در یک مثال موردی، روابط پتروفیزیکی برای تبدیل خواص تخلخل و محتوای رس آبخوان به مدل‌های سرعت لرزه‌ای و مقاومت ویژه الکتریکی، به کار برده شده است. زون اشباع نشده در عمق بالای ۲ متر زیر سطح زمین در نظر گرفته شده است. زون اشباع در زیر یک زون متراکم به ضخامت یک متر که محتوای آب آن به طور خطی افزایش می‌یابد، قرار دارد. مدل در عمق ۳ متری به پایین به صورت کاملاً اشباع در نظر گرفته شده است. برای دست یافتن به سرعت لرزه‌ای (شکل ۱-۴ الف)، از معادله تجربی هان^۱ و همکاران (۱۹۸۶) استفاده شد که برای ماسه رسی اشباع تحت فشار ۵ مگا پاسکال و دامنه تغییرات تخلخل بین ۲ تا ۳۰ درصد، معتبر است. از آنجا که رسوبات ناپیوسته هستند؛ از لحاظ تجربی، سرعت محاسبه شده به مقدار ۰/۴ متر بر میلی ثانیه کاهش داده شد تا مقدار سرعت به طور نوعی برای رسوبات اشباع شده سطحی، به دست آید. در مناطقی از مدل با اشباع شدگی کمتر

^۱ Han

از ۹۵ درصد، مقدار سرعت میانگین برابر ۰/۳۵ متر بر میلی ثانیه در نظر گرفته شد تا تباین سرعت نسبی به دست آمده توسط معادله هان و همکاران (۱۹۸۶)، حفظ شود [۳۴].

با استفاده از روابط تجربی نول^۱ (۱۹۹۶) یک مدل مقاومت ویژه الکتریکی مصنوعی (شکل ۴-۱ ب) از پارامترهای تخلخل و محتوای رس آبخوان در حالت‌های کاملاً اشباع و غیر اشباع، به دست آمد [۳۳]. مقاومت‌های ویژه به دست آمده برای زون اشباع نشده در مقایسه با مشاهدات صحرایی غیر واقعی بودند [۳۵]، [۸]. به هر حال نول (۱۹۹۶) از نمونه‌های اتاق خشک^۲ که به اندازه کافی شرایط زون اشباع نشده را نشان نمی‌دادند، استفاده کرد. بنابراین مقادیر مقاومت ویژه در زون تهویه در مدل به اندازه ۱۰ اهم‌متر کاهش داده شد تا یک مدل مقاومت ویژه نزدیک به مقدار واقعی برای آبخوان غیر پیوسته سطحی به دست آید [۲].



شکل ۴-۱- الف) سرعت موج P و ب) مدل مقاومت ویژه الکتریکی به دست آمده از تغییر پتروفیزیکی خواص آبخوان نشان داده شده در شکل ۳-۱ [۲].

¹ Knoll
² Room-dry

۲-۵-۱ آزمایش‌ها و وارون‌سازی داده‌های مصنوعی

یک آزمون توموگرافی لرزه‌ای انکساری مصنوعی با به کارگیری مدل‌سازی شکل موج^۱ بر اساس یک راه حل تفاضل محدود در حوزه زمان (FDTD)^۲ از معادله موج صوتی، شبیه‌سازی شد. منبع پالس یک موجک Ricker با فرکانس مرکزی ۸۰ هرتز و یک پهنای باند بین ۲ تا ۳ اکتاو^۳ (برای بیان محدوده فرکانس یا نسبت فرکانسی استفاده می‌شود) بود. داده‌های لرزه‌ای با موقعیت‌های منبع در هر ۵ متر و موقعیت‌های ژئوفون^۴ در هر متر محاسبه شده است. به صورت متوالی، زمان رسیدن اولین موج P به صورت قله (نوک تیز) می‌باشد. نتیجه مهاجرت‌های زمانی با استفاده از یک رهیافت وارون‌سازی محدود شده هموار بر اساس راه حل تفاضل محدود از معادله ایکونل^۵، وارون‌سازی شده که به طور کامل، انحنای مسیر پرتو در لایه ناهمگن را محاسبه می‌کند [۳۶]. نتیجه مدل سرعت در شکل ۵-۱ الف نشان داده شده است.

یک برداشت مقاومت‌ویژه دو بعدی با استفاده از ۱۰۰ الکتروده با فاصله یک متر در امتداد لبه بالایی مدل آبخوان، شبیه‌سازی شد. آرایه الکترودی شلومبرژه با الکترودهای پتانسیل به فواصل ۱، ۲ و ۳ متر و آرایه دوقطبی-دوقطبی با فاصله الکترودی (طول هر یک از دوقطبی‌ها) ۱، ۲، ۵ و ۸ متر و گام‌های حداکثر ۱۰ برابر فاصله الکترودی بین دوقطبی‌ها، داده‌های مصنوعی مقاومت‌ویژه ظاهری را بر حسب فاصله الکترودی تولید کرده است. مقاومت‌ویژه ظاهری با استفاده از نرم‌افزار DC2DInvRes تولید یا محاسبه شده است. در مجموع، نوفه^۶ تصادفی ۳ درصد و یک نسبت سیگنال به نوفه وابسته به ولتاژ، در شبیه‌سازی برداشت و اندازه‌گیری واقعی، به کار گرفته می‌شود [۳۷]. برای انجام وارون‌سازی بر روی داده‌ها [۳۸]، از نرم‌افزار تجاری RES2DINV استفاده شد؛ تا مقاومت‌ویژه ظاهری

^۱ Waveform modeling

^۲ Finite difference time domain

^۳ Octaves

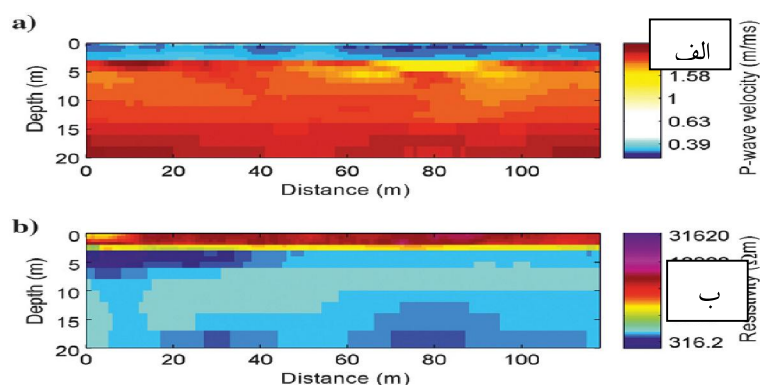
^۴ Geophone

^۵ Eikonal equation

^۶ Noise

با استفاده از یک الگوریتم وارون‌سازی مقید شده هموار گوس- نیوتن وارون‌سازی شود. مدل مقاومت‌ویژه به دست آمده در شکل ۵-۱ ب نشان داده شده است [۲].

با مقایسه مدل‌ها در شکل ۴-۱ و شکل ۵-۱، مشاهده می‌شود؛ که مدل‌های ساخته شده در شکل ۵-۱، بازسازی نامشخص‌تری در آنومالی‌های بزرگ مقیاس نسبت به مدل‌های اصلی دارند. یعنی آنومالی‌های مقاومت‌ویژه پایین در $x=10/z = 4m$ و $x=80/z = 16m$ (شکل ۵-۱ ب)، مشاهده می‌شود که آنومالی عمیق‌تر با تباین پایین‌تری بازسازی شده است. برای مدل سرعت، ساختار آنومالی‌ها در نیمه بالایی منطقه مدل اشباع شده، به وضوح مشخص شده است. در اعماق بالاتر، برخی اثرات مصنوعی، به علت پوشش ضعیف مناطق عمیق‌تر مدل در اثر نفوذ تعداد خیلی محدودی از پرتوها، می‌باشد [۲].



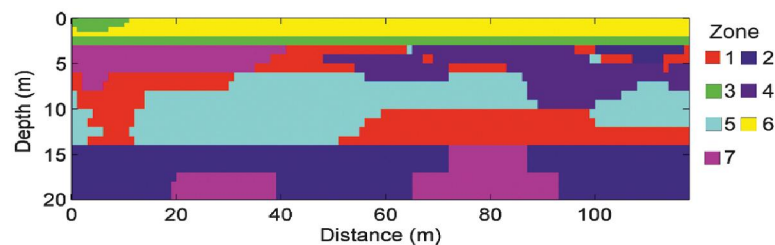
شکل ۵-۱- داده‌های مصنوعی وارون‌سازی شده الف) سرعت موج P حاصل از تست لرزه نگاری انکساری و ب) مدل مقاومت‌ویژه الکتریکی به دست آمده از تست مقاومت ویژه بر روی مدل آبخوان نشان داده شده در شکل ۴-۱. توجه این که به طور سیستماتیکی کیفیت مدل با عمق کاهش می‌یابد (مقایسه با شکل ۴-۱) [۲]

۳-۵-۱ تحلیل خوشه‌ای طیفی

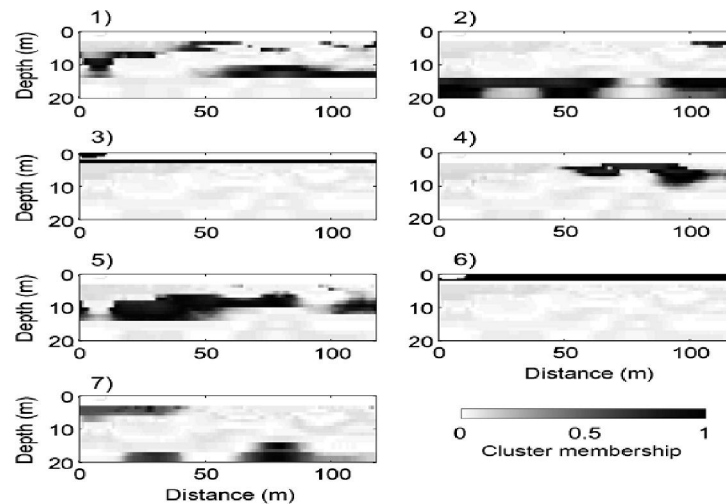
مدل‌های شکل ۵-۱ نرمال شده و تحت تحلیل خوشه‌ای طیفی قرار گرفتند. با محاسبه تشابهات جفت جفت^۱ گوسین یک گراف نزدیک‌ترین همسایگی متقابل k با $k = 320$ و $\sigma = 50$ تولید شد و

^۱ Pairwise similarities

نتیجه در یک ماتریس تشابه قرار گرفت. اطلاعات قدرت تفکیک مدل در زمان محاسبه تشابهات گوسین، در نظر گرفته نمی‌شود. اندازه‌های تشابه مصنوعی وابسته به ترکیب اطلاعات قدرت تفکیک مدل شده است. اولین هفت بردار ویژه از ماتریس لاپلاسین گراف نرمال شده، تحت تحلیل خوشه‌ای FCM قرار گرفته است. سرانجام اطلاعات عضویت می‌تواند برای رسیدن به یک مدل منطقه‌بندی شده از حالت فازی خارج شود؛ که این مدل، تغییرات ساختاری غالب موجود در هر دو مدل ورودی را نشان می‌دهد (شکل ۶-۱). قابلیت اعتماد مربوط به یک سلول مدل در یک خوشه مشخص می‌تواند با تحلیل اطلاعات عضویت خوشه، تعیین شود (شکل ۷-۱) [۳۹].



شکل ۶-۱- مدل منطقه‌بندی به دست آمده از تحلیل خوشه‌ای طیفی از مدل‌های وارون‌سازی شده شکل ۵-۱. زون‌ها (طبقه‌ها) با رنگ مشخص شده‌اند [۲].



شکل ۷-۱- توزیع عضویت خوشه به دست آمده برای هر کلاس از مدل منطقه‌بندی شده (شکل ۶-۱). درجه بالای عضویت، قابلیت اعتماد بالای یک پارامتر مدل (سلول مدل) را در کلاس مشخص، نشان می‌دهد [۲].

¹ Mutual k-nearest neighbor graph

به طور خلاصه می‌توان گفت که رهیافت تحلیل داده‌های ژئوفیزیکی زیر سطحی و تحلیل خوشه‌ای طیفی منجر به یک مدل ژئوفیزیکی چند پارامتری منطقه‌بندی شده می‌شود؛ که در آن ساختارهای متوسط تا بزرگ مقیاس غالب و موجود در مدل آبخوان، مشخص و بارز هستند. به هر حال، قرار دادن پارامتر مناسب از الگوریتم طیفی، تحت‌الشعاع یک ارزیابی تجربی خاص از کاربر قرار می‌گیرد. این نتایج، اشکالات تعداد خوشه بهینه را مطابق با تحلیل خوشه‌ای تقسیم‌بندی کننده استاندارد بر اساس اندازه آماری اعتبار خوشه، از قبیل آنتروپی طبقه‌بندی نرمال شده یا شاخص کسی - بنی^۱، مشخص می‌کند [۴۰]. آن‌ها تنها در مرحله آخر یک الگوریتم خوشه‌ای طیفی مشخص می‌شوند؛ اما هر اطلاعاتی درباره تنظیم مناسب ماتریس تشابه یا انتخاب یک لاپلاسیان گراف مناسب را نمی‌دهد. در آخر می‌تواند به وسیله تحلیل مقدار ویژه ماتریس تشابه تعیین شود. تعداد مقدار ویژه‌های نزدیک به واحد می‌تواند انتخاب یک تعداد مناسب از خوشه‌ها را مشخص کند (برای مثال شکل ۱-۲ ب را ببینید). به هر حال بررسی‌های چشمی نتایج پایانی، اهمیت تحلیل خوشه‌ای تقسیم‌بندی کننده را نشان می‌دهد. در تحلیل خوشه‌ای، ساختارهای خوشه‌بندی شده یا نباید کوچک‌تر از محدودیت قدرت تفکیک داده‌ها باشد یا باید با دیگر اطلاعات موجود سازگاری داشته باشد [۴۰].

۱-۶ الگوریتم TSVM

تئوری یادگیری آماری^۲ توسط وپنیک^۳ [۴۱] پیشنهاد شده است؛ که یک تئوری مربوط به مسئله نمونه کوچک است. ماشین بردار پشتیبان (SVM)^۴ یک تولید جدید از الگوریتم‌های یادگیری بر اساس تئوری یادگیری آماری می‌باشد [۴۲]. در حال حاضر SVM به طور پیوسته توسعه می‌یابد.

^۱ Xie-Beni

^۲ Statistic learning theory

^۳ Vapnik

^۴ Support vector machine

اگرچه SVM اولیه می‌تواند با مجموعه داده‌های ابعاد بالا، غیرخطی و نمونه کوچک، سر و کار داشته باشد، همچنین می‌تواند یک مسئله پیچیده در فرآیند طبقه‌بندی نمونه‌های محدود را حل کند. بنابراین اصل اساسی آن است که وقتی یک مسئله داده شده^۱ حل شد، باید از حل یک مسئله کلی‌تر^۲ در فرآیند میانی^۳ اجتناب کنیم. بر اساس این تئوری، وپنیک ایده تبدیل (ترانس داکتیو)^۴ را پیشنهاد نمود [۴۳] و الگوریتم ماشین بردار پشتیبان تبدیلی (TSVM) را ارائه داد که می‌تواند عمل طبقه‌بندی را به طور مؤثری بهبود بخشد [۴۴].

بر اساس خصوصیات ذاتی طبقه‌بندی SVM، الگوریتم TSVM بر اساس خوشه‌بندی طیفی مطرح می‌شود [۲۳]. این طبقه‌بند (کلاسیفایر) ابتدا از خوشه‌بندی طیفی برای خوشه‌بندی نمونه‌های مشابه بدون برچسب^۵ استفاده می‌کند. سپس همه نمونه‌ها در خوشه یکسان با برچسب مشابه انتخاب می‌شوند. سرانجام کلاسیفایر از روش یادگیری ترانس داکتیو^۶ برای طبقه‌بندی استفاده می‌کند. چون این روش از خوشه‌بندی طیفی در توزیع مؤثر اطلاعات نمونه‌های بدون برچسب، استفاده می‌کند، بنابراین می‌تواند به طور مؤثری عملکرد الگوریتم TSVM را بهبود بخشد [۴۴].

ایده یادگیری ترانس داکتیو و الگوریتم طبقه‌بندی TSVM که توسط جواچیمز^۷ (۱۹۹۹) مطرح شده است [۴۵]، در فصل ۳ آمده است. بر اساس ویژگی‌های کلاسیفایر SVM سنتی و نتایج خوب خوشه‌بندی طیفی، یک الگوریتم یادگیری TSVM بر اساس خوشه‌بندی طیفی مطرح و جزئیات فرآیند الگوریتم در ادامه ارائه خواهد شد [۴۴].

^۱ Given problem

^۲ More general problem

^۳ Middle process

^۴ Transductive

^۵ Non-label

^۶ Transductive

^۷ Joachims

۷-۱ نتایج پردازش و تحلیل مثال موردی

به منظور بررسی عملکرد TSVMSC، یک مجموعه داده از اخبار رویترز^۱ (Reuters-21578) ساخته شد. این مجموعه داده از اخبار رویترز در سال ۱۹۸۷ می‌باشد، که یک مجموعه داده طبقه‌بندی متنی معمول^۲ است [۴۶]. داده‌های آزمایشی با به کارگیری مجموعه داده قبلاً پردازش شده جواچیمز به دست آمدند.

هر کلمه با یک بردار در فضای ابعادی بالا نشان داده شده است. هدف از این مثال موردی آموزش دادن یک کلاسیفایر در پیدا کردن متن‌های مرتبط با اکتساب یکپارچه^۳ می‌باشد. این مثال موردی تابع مبنای شعاعی^۴ (RBF) را به عنوان تابع کرنل^۵ [۴۷] انتخاب می‌کند (معادله (۱-۱)):

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (1-1)$$

که σ یک پارامتر عرض می‌باشد. x و y بردارهای n بعدی در فضای مشخصه اصلی هستند.

در این بحث، مجموعه آموزشی متشکل از ۵ نمونه برچسب خورده مثبت^۶ و ۵ نمونه برچسب خورده منفی^۷ در نظر گرفته شده است. برای تست عملکرد الگوریتم TSVM، تعداد متفاوت از نمونه‌های برچسب نخورده انتخاب می‌شود. برای الگوریتم TSVMSC، k (تعداد کلاس) برابر ۳ و ۵ و ۷ در الگوریتم خوشه‌بندی طیفی تنظیم می‌شود. در این مثال از مجموعه تست یکسان شامل ۳۰۰ نمونه برچسب خورده مثبت و ۳۰۰ نمونه برچسب خورده منفی استفاده شده است. جدول ۱-۱ نتایج یادگیری الگوریتم TSVM و الگوریتم TSVMSC را زمانی که $k = 5$ است را نشان می‌دهد و جدول ۲-۱ نتایج یادگیری الگوریتم TSVMSC را زمانی که k برابر ۳ و ۵ و ۷ است را نشان می‌دهد [۴۴].

¹ Reuters newswire

² Typical text classification data set

³ Corporate acquisition

⁴ Radial Basis Function

⁵ Kernel function

⁶ Positive labeled samples

⁷ Negative labeled samples

در جدول ۱-۱ نشان داده شده است که الگوریتم TSVM خیلی پایدار نیست، زیرا برای الگوریتم TSVM، تعداد نمونه‌های برچسب خورده مثبت، N ، در مجموعه نمونه‌های برچسب نخورده باید به طور مصنوعی تعیین شود. به هر حال مقدار N اغلب برای تخمین قابل قبول مشکل است. اگر N به درستی تخمین زده شود، دقت الگوریتم TSVM به طور نسبی بالاست. اما اگر N به درستی تخمین زده نشود عملکرد الگوریتم می‌تواند به طور زیاد تحت تأثیر قرار گیرد. به هر حال در الگوریتم TSVMSC نیازی به تخمین مقدار N نیست، بنابراین پایداری الگوریتم TSVMSC به طور قابل ملاحظه‌ای از الگوریتم TSVM بهتر و بهبود یافته‌تر است. چون که الگوریتم TSVMSC از خوشه‌بندی طیفی در نمونه‌های مشابه خوشه استفاده می‌کند، و آن‌ها را با برچسب یکسان علامت گذاری می‌کند، بنابراین این الگوریتم عملاً از اطلاعات توزیع قبلی^۱ اطلاعات نمونه‌های برچسب نخورده به طور مؤثری استفاده می‌کند و به این خاطر این الگوریتم می‌تواند به یک سطح دقت بالاتری برسد [۴۴].

جدول ۱-۱- مقایسه بین نتایج یادگیری بر روی مجموعه داده‌های روتیز با استفاده از روش‌های TSVM و

TSVMSC [۴۴]

تعداد نمونه‌های برچسب نخورده	الگوریتم	زمان آموزش (ثانیه)	دقت تست (درصد)
مثبت = ۵۰۰	TSVM	۱۳۳/۶۶	۸۹/۱۶
منفی = ۵۰۰	TSVMSC	۱۰۵/۳۲	۹۰/۲۱
مثبت = ۵۰۰	TSVM	۱۵۶/۰۳	۹۵/۲۶
منفی = ۱۰۰۰	TSVMSC	۱۱۵/۳۵	۹۲/۱۹
مثبت = ۱۰۰۰	TSVM	۱۶۰/۹۶	۷۲/۶۹
منفی = ۵۰۰	TSVMSC	۱۱۵/۳۵	۹۲/۱۹
مثبت = ۱۰۰۰	TSVM	۳۰۰/۵۶	۹۱/۱۰
منفی = ۱۰۰۰	TSVMSC	۲۳۶/۱۱	۹۳/۵۶

جدول ۲-۱ نشان می‌دهد که با افزایش k (تعداد کلاس‌ها) زمان اجرای الگوریتم TSVMSC طولانی‌تر می‌شود، که درک آن بسیار ساده است چون که الگوریتم TSVM نیاز به حل مسئله بهینه‌سازی دارد. با افزایش k دقت الگوریتم TSVMSC روی مجموعه تست افزایش می‌یابد. این امر مهمی است زیرا

^۱ Priori distribution

مسئله بهینه‌سازی با افزایش k می‌تواند با دقت بیشتری حل شود، بنابراین تعمیم توانایی الگوریتم افزایش می‌یابد و به همان نسبت دقت روی مجموعه تست بهبود می‌یابد [۴۴].

جدول ۱-۲- مقایسه نتایج یادگیری TSVMSC با در نظر گرفتن تعداد خوشه مختلف [۴۴]

k	تعداد نمونه‌های برچسب نخورده	زمان آموزش (ثانیه)	دقت تست (درصد)
۳	۱۰۰۰	۷۵/۶۰	۸۶/۳۷
۵		۱۰۵/۳۲	۹۰/۲۱
۷		۱۶۵/۲۳	۹۵/۹۱
۳	۱۵۰۰	۱۰۸/۱۶	۸۸/۱۸
۵		۱۱۵/۳۵	۹۲/۱۹
۷		۱۸۱/۲۲	۹۶/۷۷
۳	۲۰۰۰	۱۹۸/۷۷	۸۹/۶۷
۵		۲۳۶/۱۱	۹۳/۵۶
۷		۳۱۰/۱۶	۹۵/۵۲

۸-۱ ضرورت و اهداف تحقیق

همان طور که قبلاً اشاره شد، تجربه نشان می‌دهد که روش TSVMSC برتری بالاتری نسبت به روش TSVM در عملکرد پایدار و دقت دارد. به دلیل این که این روش یکی از روش‌های نیمه نظارت شده می‌باشد، می‌توان با توجه به دانش‌محور بودن این روش، طبقه‌بندی بهتری از نتایج خروجی معکوس‌سازی شده داده‌های توموگرافی لرزه انکساری و مقاومت ویژه داشت.

در سال‌های اخیر برخی تلاش‌ها در جهت تلفیق مدل‌های توموگرافی سطحی با استفاده از الگوریتم خوشه‌ای تقسیم‌بندی شده بوده است. یعنی الگوریتم گوستافسون-کسل یا FCM که توسط نویسندگان ارائه شده به علت نتایج ضعیف، این مطالعات، مخاطبان زیادی را شامل نشد. مشکلات کاربرد الگوریتم‌های خوشه‌ای جزئی در تلفیق مدل‌های توموگرافی مشتق شده از داده‌های سطحی، در خاصیت روش‌های برداشت مختلف می‌باشد.

با توجه به مطالعات انجام شده توسط هاجمولر و پاشه در سال ۲۰۱۳، به منظور رسیدن به یک نظریه هدف برای تلفیق مدل‌های توموگرافی ناشی از برداشت‌های ژئوفیزیکی سطحی، نظریه پاشه

و همکاران در سال ۲۰۰۶ بازنگری شده و در عمل هسته روش با تغییراتی در الگوریتم خوشه‌ای FCM با یک الگوریتم خوشه‌ای طیفی، بهبود یافت.

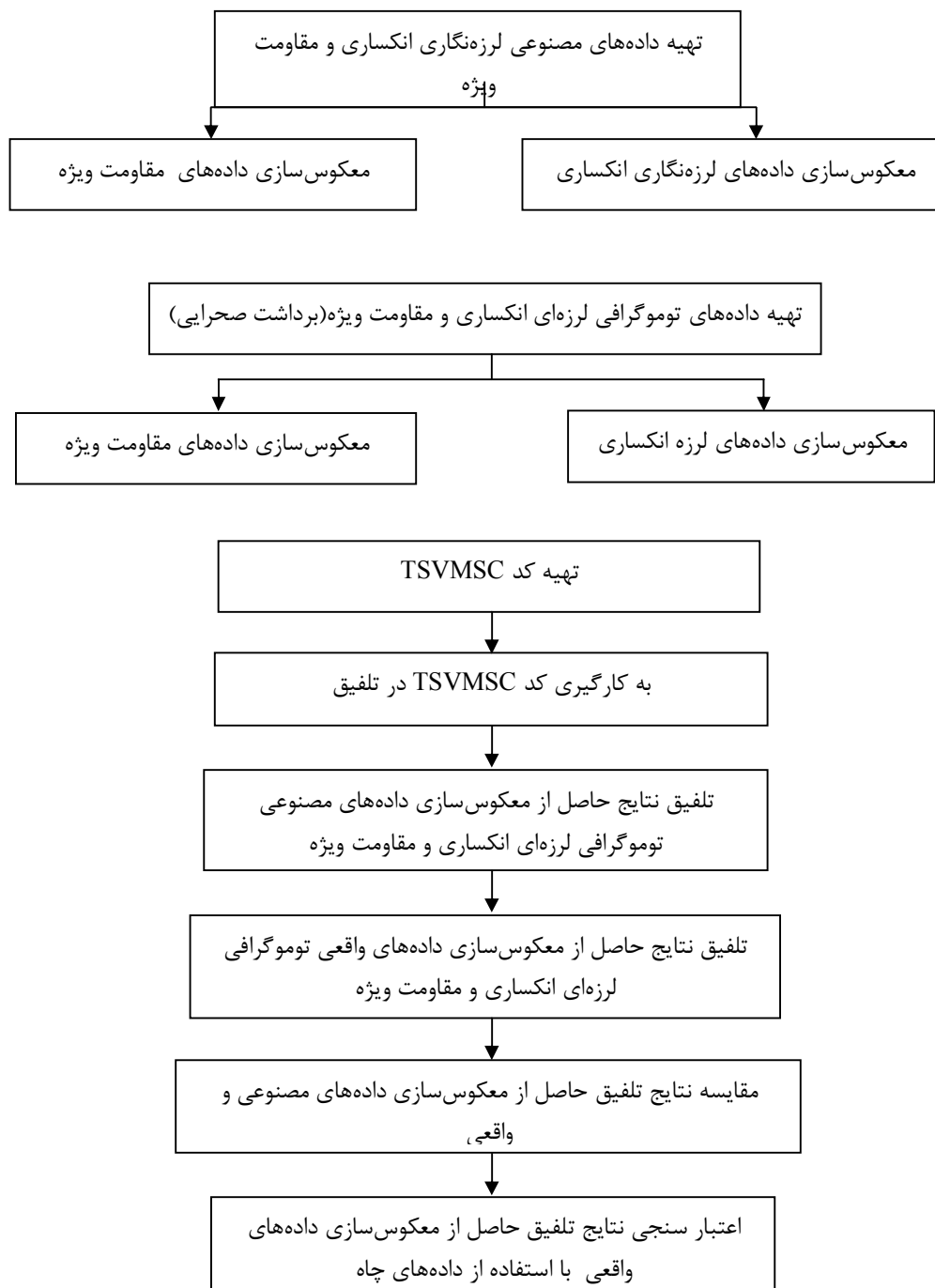
در مطالعه حاضر به جای استفاده از الگوریتم خوشه‌ای طیفی، سعی بر آن است که از الگوریتم ماشین برداری پشتیبان تبدیلی یا ترانس داکتیو بر اساس خوشه‌بندی طیفی (TSVMSC)، به دلیل عملکرد پایدار و دقت بالای آن استفاده شود.

هدف از مطالعه حاضر، بهبود تلفیق نتایج حاصل از معکوس‌سازی داده‌های لرزه‌نگاری انکساری و مقاومت ویژه با استفاده از روش TSVMSC که یک روش دانش محور است، می‌باشد. این الگوریتم، ناهمگنی فضایی در همه مدل‌ها را در بر می‌گیرد و نیز انعطاف‌پذیر بوده و می‌تواند با هر نوع داده‌ای برای شناسایی الگوها و ساختارهای غالب موجود در داده‌های چند متغیره سازگاری داشته باشد.

۹-۱ روش انجام تحقیق

داده‌ها با هماهنگی شرکت توسعه منابع آب و نیروی ایران از شرکت سابیر اخذ شد. بعد از تهیه داده‌های مصنوعی و صحرایی توموگرافی لرزه‌نگاری انکساری و مقاومت ویژه، معکوس‌سازی روی آن‌ها صورت گرفته و نتایج مورد تحلیل قرار گرفت.

روش انجام تحقیق همراه با جزئیات آن در نمودار زیر ارائه شده است:



بعد از تهیه کدهای میانگین k ، میانگین فازی، گوستافسون-کسل و الگوریتم TSVMSC در محیط برنامه نویسی Matlab، تلفیق نتایج حاصل از معکوس‌سازی داده‌های مصنوعی و صحرائی توموگرافی

لرزه‌ای انکساری و مقاومت ویژه بدست آمد و نتایج حاصله، با نگار گمانه موجود مورد ارزیابی و اعتبارسنجی قرار گرفت.

۱-۱۰ سازماندهی مطالب

مجموعه حاضر شامل هفت فصل می‌باشد. در فصل اول کلیاتی از تلفیق مدل‌های توموگرافی سطحی، تحلیل خوشه‌ای طیفی و مثال موردی از کاربرد تحلیل خوشه‌ای طیفی در روش تلفیق FCM ارائه شده است. در این فصل همچنین مقدمه‌ای بر الگوریتم TSVM ارائه شده است. در فصل دوم، مبانی توموگرافی مقاومت‌ویژه و لرزه انکساری مورد بحث قرار گرفته است. در فصل سوم، روش‌های طبقه‌بندی و خوشه‌بندی بررسی می‌شود. در فصل چهارم، فرمولاسیون مسئله مورد بررسی قرار گرفته است. در فصل پنجم، وارون‌سازی داده‌های مصنوعی و صحرایی داده‌های لرزه و مقاومت‌ویژه ارائه و نتایج مورد تجزیه و تحلیل قرار می‌گیرد. در فصل ششم، تلفیق نتایج حاصل از وارون‌سازی داده‌های توموگرافی لرزه‌ای و مقاومت‌ویژه و اعتبارسنجی آن ارائه شده و در نهایت، در فصل هفتم، نتیجه گیری و پیشنهادات مطرح و مورد بررسی قرار گرفته است.

فصل دوم

مبانی تو موکراتانی مقاومت و پیره و لرزه

انکساری

۱-۲ روش توموگرافی الکتریکی

ریشه اصلی واژه توموگرافی^۱ از کلمه یونانی تومو به معنی برش و گراف به معنی تصویر استخراج شده است. اسمی که بعداً به این تکنیک و روش تصویر برداری جدید داده شد، بیشتر به کاربرد آن در علم پزشکی تحت عنوان سی تی اسکن^۲ برمی گردد، زیرا اولین تصاویر یا توموگرام‌های رادیولوژی یک اسکن کامل مربوط به بخش خاصی از بدن می باشند که در آن‌ها از پرتوهای مجهول X به جای پرتوهای لرزه‌ای استفاده شده است. این تصاویر با جابجایی منبع اشعه X و صفحه عکس برداری در جهت‌های مختلف به دست می آیند و رادیولوژیست‌ها قادر هستند، از طریق شدت اشعه X مشاهده شده در محدوده مورد بررسی، وضعیت بافت‌ها و تراکم یا پوکی استخوان‌ها را بررسی کنند. آنچه در ژئوتوموگرافی انجام می شود بسیار شبیه توموگرافی پزشکی می باشد با این تفاوت که در ژئوتوموگرافی به علت وسعت بیشتر محدوده و شرایط خاص زمین شناسی با چالش‌های بیشتری رو به رو هستیم [۴۸].

روش توموگرافی الکتریکی یکی از روش‌های مقاومت‌ویژه با دقت تفسیر بالا می باشد [۴۹]. در این روش برداشت‌های مقاومت‌ویژه در یک شبکه منظم به وسیله یک آرایه الکتروودی بهینه انجام گرفته و سپس با استفاده از مدل سازی، مقطع عمقی منطقه مورد مطالعه در اختیار قرار می گیرد. این روش یک روش نسبتاً جدید با برداشت بهینه می باشد که به منظور بهبود، افزایش دقت، قدرت تفکیک عمقی (قائم) و جانبی (افقی) در شناسایی اهداف زیرزمینی به صورت دو بعدی و سه بعدی کاربرد دارد. هدف از توموگرافی الکتریکی ایجاد تصاویری از خصوصیات الکتریکی زیر سطح زمین می باشد.

^۱ Tomography

^۲ Computed Tomography Scanning

در دهه‌های اخیر برداشت‌های توموگرافی الکتریکی مقاومت‌ویژه^۱ (ERT) به طور گسترده‌ای در بررسی‌های ژئوفیزیکی مورد استفاده قرار گرفته است [۵۰]. این برداشت‌ها برای به نقشه در آوردن مناطق با زمین شناسی زیرسطحی پیچیده که در آن‌ها، برداشت‌های پروفیل زنی و سونداژ زنی معمولی کارایی چندانی ندارد و به منظور تأمین اطلاعات در هر دو جهت قائم و جانبی و تعیین یک حالت کیفی از شکل و عمق اهداف زیرزمینی، مورد استفاده قرار می‌گیرند [۵۱]. این بررسی‌ها عموماً به منظور اکتشاف هیدروژئولوژیکی، معدنی، ژئوتکنیکی و نیز مطالعات زیست محیطی مورد استفاده می‌باشند [۵۲].

۱-۲ مدل سازی داده‌های توموگرافی

فرایند مدل سازی داده‌های توموگرافی که از اهمیت بالایی برخوردار است؛ داده‌های خام به دست آمده از برداشت را به یک نتیجه قابل تفسیر و بررسی تبدیل می‌کند. در این مرحله ضخامت و مقاومت‌ویژه لایه‌های زیرسطحی تخمین زده می‌شوند. مدل سازی دقیق و صحیح، اعتبار تفسیر را بالا می‌برد. از این رو مدل سازی باید با دقت بالایی انجام پذیرد.

در حالت کلی مدل سازی می‌تواند به دو صورت انجام گیرد

الف) مدل سازی مستقیم^۲ (پیشرو)

ب) مدل سازی وارون^۳ (معکوس)

^۱ Electrical resistivity tomography

^۲ Forward modelling

^۳ Inverse modelling

۲-۱-۲ مدل سازی پیشرو

در این نوع مدل سازی زمین و پراکندگی مقاومت ویژه در قسمت های مختلف آن موجود است و داده ها و منحنی های مورد نیاز از این مدل زمین به دست می آید. به عبارت بهتر در این نوع مدل سازی، از زمین به منحنی مقاومت ویژه ظاهری رسیده می شود. در این روش مدل اولیه به صورت اتفاقی تعیین می شود و کنترل خاصی در انتخاب مدل اولیه وجود ندارد. این روش مدل سازی بر مبنای آزمون و خطا^۱ بر روی مدل های زمینی دوبعدی با استفاده از روش اجزاء محدود یا روش تفاضل محدود صورت می پذیرد. مدل اولیه در این نوع از مدل سازی سلیقه ای می باشد و بسته به تجربه کارشناس ژئوفیزیک بهبود می یابد [۵۳]. از نرم افزارهای ژئوفیزیکی که در مدل سازی پیشرو مورد استفاده قرار می گیرد، می توان به نرم افزار Res2Dmod اشاره نمود.

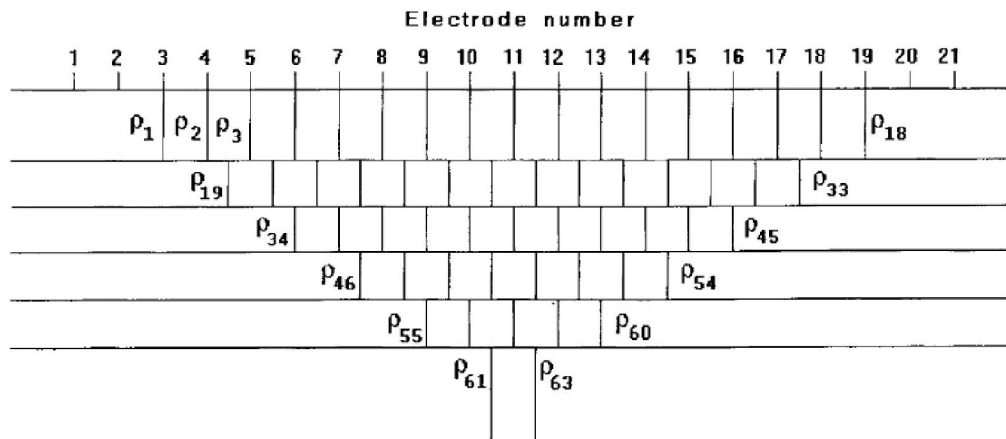
۳-۱-۲ مدل سازی وارون

در روش وارون سازی بر خلاف مدل سازی مستقیم با استفاده از داده های به دست آمده، مدل زمین تخمین زده می شود. مراحل معمول در فرآیند وارون سازی شامل قرائت اولیه داده های مقاومت ویژه، تصحیح پارامترهای وارون سازی در صورت لزوم و انجام وارون سازی می باشد. در مدل سازی قبل از هر چیز به یک مدل فرضی اولیه برای انجام مدل سازی نیاز می باشد [۵۳]. روش مدل سازی وارون معمولاً از یک تکنیک بهینه سازی غیر خطی برای بهبود بخشیدن به یک مدل ساده اولیه استفاده می کنند [۵۴].

در مدل سازی دو بعدی (2D) برای وارون سازی داده های مقاومت ویژه از نرم افزار RES2DINV استفاده می شود. در این نرم افزار مدل مورد استفاده به عنوان مدل اولیه شامل یک سری از بلوک های

^۱ Trial and error

مستطیلی دو بعدی که در شکل ۱-۲ نشان داده شده است. در این مدل تعداد بلوک‌ها با تعداد نقاط داده‌ها در شبه مقطع مقاومت ویژه ظاهری برابر است و به یک شیوه مشابه مرتب شده‌اند. همچنین عمق مراکز بلوک‌ها در عمق میانگین متناظر بازه‌های الکترونی مختلف استفاده شده قرار گرفته‌اند [۵۳].



شکل ۱-۲- ترتیب بلوک‌های مستطیلی مورد استفاده در مدل‌سازی [۵۳]

پس از تخمین مدل اولیه نوبت به تصحیح این مدل با هدف رسیدن به انطباق قابل قبول با داده‌های صحرائی می‌رسد. برای این منظور دو روش مورد استفاده قرار می‌گیرد که عبارتند از روش معمول کمترین مربعات گوس- نیوتن^۱ و روش شبه نیوتن^۲.

۲-۲ روش توموگرافی لرزه‌ای

به تصویر کشیدن ساختار سرعتی زمین با استفاده از وارون‌سازی زمان سیر امواج لرزه‌ای، توموگرافی لرزه‌ای نامیده می‌شود. توموگرافی لرزه‌ای با توجه به آرایش خاص چشمه‌ها و گیرنده‌ها و پوشش کامل منطقه مورد بررسی توسط پرتوهای موج، قابلیت تصویرسازی لایه‌های زیرسطحی را با دقت و قدرت

^۱ Gauss-Newton

^۲ Quasi-Newton

تفکیک بالا دارا می‌باشد. این روش ژئوفیزیکی از توانایی قابل توجهی در تشخیص ویژگی‌های لرزه‌شناسی (سرعت موج تراکمی و برشی) لایه‌های زمین برخوردار است. هدف اصلی در این روش، یافتن توزیع سرعت موج در محدوده مابین چشمه‌ها و گیرنده‌ها با استفاده از زمان اولین رسیدهای موج می‌باشد. مبنای کار بر تأثیرات مختلف توده‌های سنگی روی سرعت انتشار امواج الاستیک در محیط مورد بررسی استوار است. این روش به عنوان ابزاری مفید برای تعیین ویژگی‌های مهندسی ساختگاه‌ها و شناسایی زون‌های شکست خورده، هوازده، ضعیف و کاربردهای ژئوتکنیکی (نقاط مناسب برای حفاری و تزریق و بررسی اثرات تزریق در توده سنگ‌ها)، اکتشافات معدنی، پیش بینی خطرات احتمالی معدن کاری، توصیف مخازن نفتی و غیره مورد استفاده قرار می‌گیرد [۴۸].

۱-۲-۲ مدل سازی مستقیم

مدل سازی مستقیم شبیه سازی پرتو یا موج منتشر شده بر اساس شرایط زمین شناسی و داده‌های ژئوفیزیکی موجود برای یک محیط زیرسطحی می‌باشد. هدف از مدل سازی به دست آوردن لرزه نگاشت^۱ ثبت شده توسط مجموعه‌ای از گیرنده‌ها، برای یک مدل مفروض می‌باشد [۵۵].

مدل سازی بر اساس تئوری پرتو یا تئوری موج می‌باشد که در آن مسیر انتشار پرتوها و زمان سیرها برای یک مدل مشخص می‌شوند. در توموگرافی زمان سیر انجام این کار وابسته به تئوری پرتو می‌باشد اما در توموگرافی جبهه موج بر اساس نظریه موج صورت می‌گیرد [۵۶].

^۱ Seismogram

۲-۲-۲ روش‌های بازسازی تکراری

روش‌های بازسازی تصویر عبارتند از: الگوریتم برگشت به عقب^۱، روش ماتریس وارون^۲، روش‌های بازسازی تحلیلی^۳ و روش‌های بازسازی تکراری^۴، که در این مبحث به روش‌های بازسازی تکراری پرداخته می‌شود.

تعداد زیادی روش بازسازی تکراری با اندکی تفاوت وجود دارند که از جمله آن‌ها می‌توان روش‌های بازسازی جبری^۵ (ART)، روش بازسازی تکرار همزمان^۶ (SIRT) و روش بازسازی جبری چندمنظوره^۷ MART را نام برد. دو روش برجسته در تکنیک بازسازی تکراری یکی روش ART [۵۷] و دیگری روش SIRT [۵۸] می‌باشند. در روش‌های بازسازی تکراری، حل مسئله تا رسیدن به یک مقدار بهینه تکرار می‌شود. اصول این روش‌ها کمینه کردن مجموع تغییرات کندی در سلول‌های مدل می‌باشد. این تفاوت اصلی روش‌های تکراری با روش حداقل مربعات خطاست، که در آن مجموع مربعات اختلاف بین زمان رسیدهای داده‌های مشاهده شده و داده‌های محاسبه شده کمینه می‌گردد [۵۹].

در الگوریتم ART، میزان اصلاح کندی برای سلول‌هایی که اولین پرتو از آن‌ها عبور می‌کند، از معادله (۱-۲) محاسبه می‌شود. این مقدار اصلاح بر روی مقدار کندی تمام سلول‌ها اعمال می‌گردد و این کار به همین ترتیب برای تمام پرتوهای دیگر نیز انجام می‌شود. بنابراین هر دور تکرار شامل بررسی تمام پرتوها می‌باشد و مجموع تغییرات کندی در هر دوره تکرار نسبت به دوره قبل کمینه می‌گردد. عملیات تا رسیدن به یک مقدار بهینه تکرار می‌شود [۵۹].

¹ Back Projection

² Inverse Matrix

³ Analytical Reconstruction

⁴ Iterative Reconstruction

⁵ Algebraic Reconstruction Technique (ART)

⁶ Simultaneous Iterative Reconstruction Technique (SIRT)

⁷ Multiplicative Algebraic Reconstruction Technique (MART)

$$\Delta s_j^k = \frac{\Delta t_k d_{kj}}{\sum_{j=1}^N (d_{kj})^2} \quad (1-2)$$

که در آن Δt_k اختلاف زمان مشاهده شده و زمان محاسبه شده برای پرتو k ام، d_{kj} طول پرتو k ام در سلول j ام و Δs_j^k مقدار تصحیح کندی سلول j ام برای پرتو k ام می‌باشد.

در الگوریتم SIRT، میزان اصلاح کندی هر سلول با در نظر گرفتن تمام پرتوهای عبوری از آن سلول انجام می‌شود. به عبارت دیگر میزان اصلاح کندی سلول‌هایی که یک پرتو از آن عبور می‌کند مانند روش ART محاسبه می‌شود، اما این اصلاح پس از بررسی تمامی پرتوها و در پایان یک دور تکرار بر مقادیر کندی سلول‌ها اعمال می‌شود. بنابراین می‌توان نوشت (معادله (2-2)) [48]:

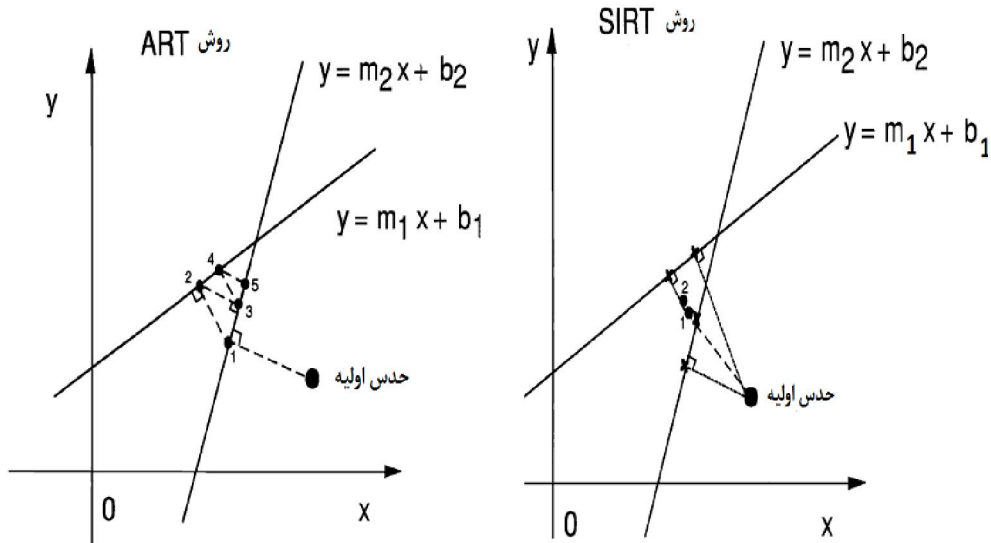
$$\Delta s_j = \frac{1}{N} \sum_{k=1}^N \Delta s_j^k \quad (2-2)$$

که N تعداد پرتوهای عبوری از سلول j ام و Δs_j مقدار اصلاح کندی این سلول می‌باشد.

شکل 2-2 توصیف شماتیک و مقایسه‌ای دو الگوریتم بازسازی تکراری در پیدا کردن جواب دو معادله (خط) را نشان می‌دهد. همان طور که دیده می‌شود در الگوریتم ART هر مرحله از حدس جواب، با در نظر گرفتن یک معادله (خط) انجام می‌شود و مسئله پس از حل برای هر معادله وارد مرحله جدید می‌شود. اما در الگوریتم SIRT حدس جواب برای هر معادله انجام می‌شود ولی پس از محاسبه مقدار میانگین این جواب‌ها، مرحله بعد انجام می‌شود [48].

روش‌های تکراری مفیدترین روش‌ها برای بازسازی تصاویر ژئوفیزیکی می‌باشند. از مزایای آن‌ها زمان نسبتاً کم محاسبات بازسازی است. هم چنین تمام آرایش‌های فرستنده-گیرنده را به راحتی در نظر می‌گیرند، اطلاعات اولیه را به خوبی در سیستم لحاظ می‌کنند و برای بازسازی انحنای پرتو مفید می‌باشند. هر چند روش ART کمتر قادر به کنترل نوفه می‌باشد، ولی یکی از کارآمدترین روش‌های بازسازی است که در هر دوره تکرار بدون هیچ زمان محاسبه اضافی، تعداد زیادی از داده‌های تصحیح

شده را ثبت می‌کند. از معایب این روش تکرار جبری این است که در تمام موارد تکرار، همگرایی به یک مقدار قابل قبول برای آن میسر نیست [۶۰].



شکل ۲-۲- توصیف شماتیک دو الگوریتم ART و SIRT [۴۸]

۳-۲-۲ معادلات ریاضی مربوط به روش SIRT

مبانی الگوریتم SIRT که در توموگرافی پرتوهای لرزه‌ای استفاده شده است، بر مبنای روش کازمرز^۱ [۶۱] برای حل معادلات خطی است. در این روش محاسبات ماتریسی به صورت ردیف به ردیف انجام گرفته و نیازی به ذخیره کردن تمام آرایه‌ها و عناصر ماتریس نمی‌باشد.

ماتریس $T = D.S$ ، شامل بردار T (زمان رسید) (معادله (۳-۲)) با تعداد M زمان رسید و بردار

S یا کندی و (معادله (۴-۲)) به عنوان تابع مدل برای N سلول است [۶۰].

^۱ Kaczmarz method

$$T = \begin{bmatrix} t_1 \\ t_2 \\ \dots \\ t_M \end{bmatrix} \quad (3-2)$$

$$S = \begin{bmatrix} s_1 \\ s_2 \\ \dots \\ s_N \end{bmatrix} \quad (4-2)$$

هم چنین D ماتریسی $M \times N$ است که آرایه‌های آن طول پرتوها در هر سلول می‌باشند (معادله (5-2)).

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ \dots & \dots & \dots & \dots \\ d_{M1} & d_{M2} & \dots & d_{MN} \end{bmatrix} \quad (5-2)$$

که d_{MN} طول پرتو Mام در سلول Nام می‌باشد.

ماتریس D در حقیقت عملگری خطی است که با عمل بر روی بردار S مربوط به مقادیر کندی برآورده شده، داده‌های اندازه‌گیری شده بردار T را ارائه می‌دهد. معادله (6-2) برای انجام توموگرافی به صورت زیر نوشته می‌شود [60].

$$T^{meas} = D S^{true} \quad (6-2)$$

این معادله به صورت مستقیم حل نمی‌شود. با مشخص بودن مقادیر T^{means} و D و با استفاده از روش‌های وارون‌سازی، مقادیر S^{true} طی یک فرایند تکراری نسبت به برآورد اولیه بهبود می‌یابند. کار

اصلی مشخص کردن یک عملگر وارون سازی تعمیم یافته^۱ است. در عمل به دو دلیل تعیین عملگر وارون سازی تعمیم یافته مشکل می باشد [۶۰].

اول اینکه ماتریس D بسیار بزرگ و پراکنده است. دوم اینکه با توجه به نحوه توزیع پرتوها، ماتریس D بد شرط^۲ است. این امر باعث می شود محاسبات وارون سازی D بسیار ناپایدار^۳ باشند. روش کازمرز [۶۱] این مشکلات را برای وارون سازی ماتریس های بزرگ و پراکنده برطرف کرده و یک ابزار موثر و مفید در تعیین یک راه حل تقریبی برای معادله (۲-۶) با استفاده از یک فرایند تکراری ارائه می دهد.

در الگوریتم SIRT که از روش کازمرز پیروی می کند، بر خلاف الگوریتم ART که اصلاح مدل بعد از ردیابی هر پرتو انجام می شود، ابتدا تمام پرتوها در سلول ها ردیابی و مقادیر اصلاح محاسبه می شوند و پس از پایان یک دور تکرار روی مدل اعمال می گردد. استراتژی اصلی روش های تکراری اعمال اصلاح به دست آمده، از اختلاف زمان سیر مشاهده شده نسبت به زمان سیر محاسبه شده برای هر پرتو، بر روی مقدار کندی اولیه سلول ها می باشد. این کار تا رسیدن به یک مقدار بهینه ادامه می یابد.

شکل ۲-۳ فلوچارت مربوط به روش SIRT را نشان می دهد که شامل سه مرحله اساسی است. روش کار بدین صورت است که ابتدا یک سری مقادیر اولیه برای بردار کندی انتخاب می شود (S^0). این مقادیر می توانند یک مقدار کندی ثابت و یا مقادیر حاصل از روش برگشت به عقب یا BP باشند. گام اول: در این مرحله متغیرهای مدل با استفاده از مقادیر اولیه محاسبه می شوند. بدین ترتیب که پس از ردیابی پرتو با استفاده از روش های موجود، زمان محاسبه شده برای هر پرتو با انجام

¹ Generalised inverse operator

² Ill conditioned

³ Unstable

مدل سازی مستقیم و با استفاده از رابطه مشابه معادله (۷-۲) به دست می آید. این کار برای تمام پرتوها انجام می شود [۶۰].

$$t_k^{calc} = \sum_{j=1}^N d_{kj} s_j^q \quad (۷-۲)$$

که در آن t_k^{calc} زمان سیر محاسبه شده برای پرتو k ام، N تعداد اجزای مدل، d_{kj} طول مسیر پرتو k ام در سلول j ام و s_j مقدار کندی مربوط به سلول j ام در مرحله تکرار q ام می باشد. گام دوم: مقدار تصحیح کندی در هر سلول برای تمام پرتوهای عبوری از آن سلول محاسبه می شود و مقدار میانگین آن از معادله (۲-۲) محاسبه می شود.

مقادیر تصحیح به دو دسته تصحیحات جمعی و ضربی تقسیم بندی می شوند. در نوع جمعی به هر سلول یک ضریب وزنی اختصاص داده می شود. ضرایب وزنی مربوطه، تابعی از طول مسیر پرتو عبوری از هر سلول (d_{kj}) و زمان باقیمانده برای پرتو k ام (Δt_k) می باشند. زمان باقیمانده اختلاف زمان مشاهده شده با زمان محاسبه شده در مرحله تکرار q ام است (معادله (۸-۲) و (۹-۲)) [۶۲].

$$\Delta s_j^k = \Delta t_k \cdot \frac{d_{kj}}{\sum_{j=1}^N (d_{kj})^2} \quad (۸-۲)$$

$$\Delta t_k = t_k^{meas} - t_k^{calc, q} \quad (۹-۲)$$

در تصحیح ضربی مقدار تصحیح از معادله (۱۰-۲) محاسبه می شود [۶۰]:

$$\Delta c_j^q = \frac{t_k^{meas}}{t_k^{calc, q}} \quad (۱۰-۲)$$

الگوریتم ART میزان اصلاح کندی برای هر پرتو را به محض محاسبه به کار می برد. به عنوان مثال به محض اینکه t_k^{calc} برای مسیر پرتو k ام محاسبه شد، مقدار تصحیح حاصل از آن برای تمام سلول هایی

که این پرتو از آن‌ها عبور می‌کند، اعمال می‌شود. اما الگوریتم SIRT تا زمانی که تمام تصحیحات مربوط به تمام پرتوهای عبوری از تمام سلول‌ها محاسبه نشوند، هیچ تصحیحی روی مدل اعمال نمی‌کند. در اینجا مقدار متوسط تصحیحات به دست آمده از تمام پرتوهای عبوری در هر سلول مشخص می‌شود و مقدار کندی هر سلول با این مقدار میانگین تصحیح می‌شود.

گام سوم: تصحیحات به دست آمده به مقدار کندی در جریان (s_j^q) مربوط به سلول زام اضافه می‌شوند و مقدار جدید کندی (s_j^{q+1}) برای مرحله تکرار بعد $(q+1)$ به عنوان مقدار شروع اولیه استفاده می‌شود. کندی جدید، با استفاده از تصحیح جمعی و تصحیح ضربی، به ترتیب به صورت معادله $(11-2)$ و $(12-2)$ به دست می‌آید [۶۰].

$$s_j^{q+1} = s_j^q + \Delta s_j \quad (11-2)$$

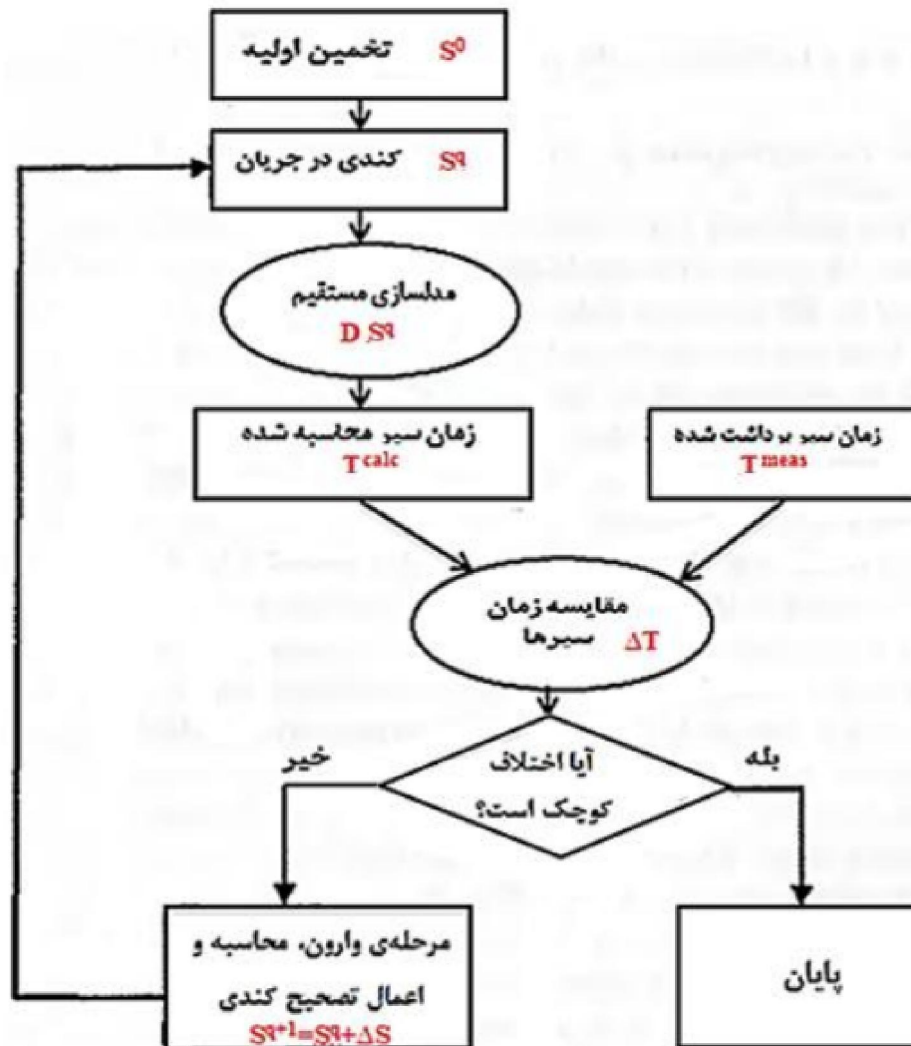
$$s_j^{q+1} = s_j^q . \Delta c_j \quad (12-2)$$

بنابراین اگر مقدار زمان سیر محاسبه شده پرتو در مقایسه با مقدار زمان محاسبه شده خیلی کوچک باشد، در این صورت مقدار کندی در سلولی که این پرتو از آن عبور کرده باشد با یک مقدار مناسب افزایش می‌یابد. پس از انجام مراحل ذکر شده برای تمام سلول‌ها و تمام پرتوها، اولین مرحله تکرار کامل شده است.

برای مرحله تکرار q ام مقدار $t_k^{calc,q}$ مانند قبل از معادله $(7-2)$ به دست می‌آید که این مقدار با مقدار زمان مشاهده برای هر پرتو مقایسه می‌شود. تمام اختلاف‌های بین مقادیر این زمان‌رسیدها برای تصحیح مقدار کندی در جریان (s_j^q) استفاده می‌شود. بنابراین در تکرار بعدی مقدار $t_k^{calc,q+1}$ به مقدار واقعی نزدیک‌تر خواهد شد.

عملیات تکرار تا رسیدن به دقت مورد نظر ادامه می‌یابد. پیدا کردن یک معیار توقف مناسب برای این تکنیک بازسازی ساده نیست. گاهی اوقات این معیار از روش سعی و خطا مشخص می‌شود. رویکرد دیگر این است که یک حد آستانه برای خطای باقیمانده در نظر گرفته شود و یا ارزیابی نرخ

کاهش خطا ملاک قرار گیرد. به عبارت دیگر اختلاف زمان مشاهده شده و مقدار زمان محاسبه شده از یک مرحله به بعد در یک محدوده تقریباً ثابت قرار بگیرد.



شکل ۳-۲- فلوچارت مربوط به روش SIRT. پارامتر S^0 تخمین اولیه کنونی، S^q کنونی در جریان که تصحیحات روی آن انجام می شود، T^{calc} زمان رسید محاسبه شده و T^{meas} زمان رسید مشاهده شده است. پارامتر S^q به طور مرتب اصلاح می شود تا جایی که اختلاف T^{calc} و T^{meas} به حد مجاز برسد [۶۰].

۴-۲-۲ وارون سازی توموگرافی

مدل اولیه مربوط به ساختار مورد بررسی، با استفاده از شواهد زمین شناسی و داده‌های ژئوفیزیک به دست می‌آید و زمان سیر محاسبه شده و مشاهده شده با یکدیگر مقایسه می‌شوند. هدف از وارون سازی به حداقل رساندن اختلاف زمان سیر حاصل از مدل اولیه و زمان سیر مربوط به داده‌های واقعی (برداشت شده) طی یک فرایند تکراری می‌باشد [۶۳]، [۶۴]. زمان سیر از معادله (۲-۱۳) به دست می‌آید:

$$t_i = \int_{l_i(z)} s \cdot dl_i(z) \quad (2-13)$$

که در آن t_i زمان سیر مربوط به پرتو i ام، s مقدار کندی یا عکس سرعت و l طول پرتو i ام است که با عمق z تغییر می‌کند.

۵-۲-۲ توموگرافی اولین زمان رسید

توموگرافی اولین زمان رسید با استفاده از اولین زمان رسیدهای ثبت شده انجام می‌شود. در این روش مدل اولیه به تعدادی سلول تقسیم می‌شود و در نهایت حل مسئله وارون با استفاده از یک الگوریتم مانند الگوریتم SIRT و با معادله (۲-۱۴) انجام می‌شود.

$$dt_i = M \cdot ds \quad (2-14)$$

۳-۲ جمع بندی و نتیجه گیری

در این فصل مبانی توموگرافی مقاومت ویژه الکتریکی و لرزه انکساری مورد بحث قرار گرفته است. در مدل سازی وارون داده‌های مقاومت ویژه، روش کمترین مربعات گوس- نیوتن و در مدل سازی وارون

داده‌های لرزه انکساری، روش بازسازی تکرار همزمان یا SIRT مدنظر بوده و تحلیل‌ها روی این موارد صورت گرفته است.

فصل سوم

روش‌های طبقه‌بندی و خوشه‌بندی

۳-۱ روش‌های طبقه‌بندی

روش‌های یادگیری نظارت شده هستند. در طبقه‌بندی، تعداد کلاس‌ها معلوم، نمونه‌ها هم معلوم است که مربوط به کدام کلاس‌ها هستند (برچسب کلاس مشخص است) و ویژگی‌ها یا مشخصه‌های نمونه‌ها نیز مشخص است. برای مثال یک دکتر ممکن است پیش بینی کند که یک بیمار کم خون^۱ (کلاس یک) است یا نه (کلاس دو)، بر اساس برخی از فاکتورهای پیشگو از قبیل سن، وزن و غیره. بر اساس این مثال هر شیء فرض می‌شود که تنها به یک کلاس معلوم تعلق دارد، یعنی یک تابع هدف برای قرار دادن اشیاء در کلاس مناسب تعریف می‌شود [۶۵]. در این چارچوب، روش‌های طبقه‌بندی در مطالعه و توسعه روابط بین کلاس‌های از قبل تعیین شده و یک مجموعه از مشخصه‌های وابسته بیان می‌شوند، از قبیل طبقه‌بندی ماشین بردار پشتیبان (SVM)، ماشین بردار پشتیبان تبدیلی (TSVM) و...

۳-۱-۱ ماشین بردار پشتیبان

الگوریتم ماشین بردار پشتیبان یا SVM اولیه در ۱۹۶۳ توسط ولادیمیر وپنیک ابداع شد و در سال ۱۹۹۵ توسط وپنیک و کرینا کورتس^۲ [۴۱] برای حالت غیرخطی تعمیم داده شد. SVM یکی از روش‌های یادگیری با نظارت^۳ است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. در این روش، با فرض این که دسته‌ها به صورت خطی جداپذیر باشند، ابرصفحه‌هایی با حداکثر حاشیه^۴ را به دست می‌آورد که دسته‌ها را جدا کنند و در مسائلی که داده‌ها به صورت خطی جداپذیر نباشند،

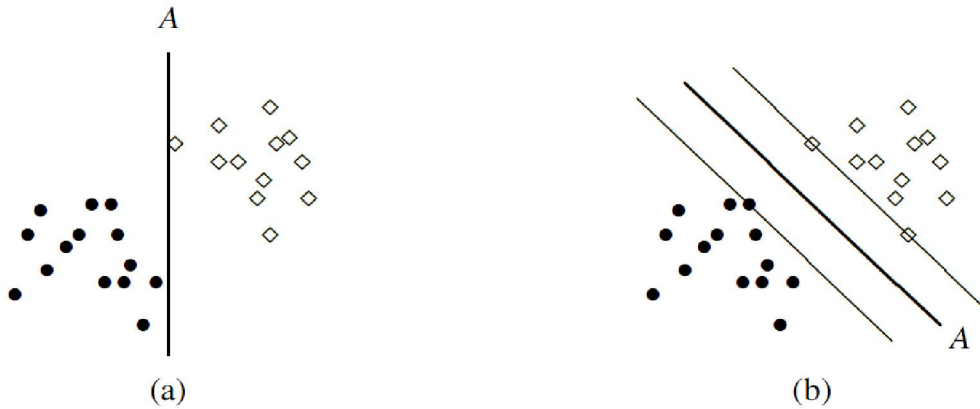
^۱ Anemic

^۲ Corinna Cortes

^۳ Supervised learning

^۴ Maximum margin

داده‌ها به فضای با ابعاد بیشتر^۱ نگاشت پیدا می‌کنند تا بتوان آن‌ها را در این فضای جدید به صورت خطی جدا نمود (شکل ۱-۳) [۶۶].



شکل ۱-۳- دو مجموعه مجزا به صورت خطی با ابرصفحه‌های جداکننده. ابرصفحه جداکننده در سمت راست (خط ضخیم) در فاصله بیشینه از نزدیک‌ترین نقاط قرار می‌گیرد. خطوط نازک مرز یا حاشیه را مشخص می‌کنند [۶۶].

۲-۱-۳ جداسازی داده‌ها به صورت خطی

طبقه‌بندی دوتایی^۲ با یک مجموعه الگوهای آموزشی $(x_i, i = 1, \dots, n)$ در یکی از دو کلاس w_1 و w_2 با برچسب‌های $y_i = \pm 1$ تعریف شده است.

تابع جدا کننده خطی^۳ به صورت معادله (۱-۳) نشان داده می‌شود:

$$g(x) = w^T x + w_0 \quad (1-3)$$

که در آن w بردار وزنی و w_0 بردار آستانه ای است. قاعده تصمیم^۴ بصورت معادله (۲-۳) می‌باشد:

$$w^T x + w_0 \begin{cases} > 0 \\ < 0 \end{cases} \Rightarrow x \in \begin{cases} w_1 & \text{with corresponding numeric value } y_i = +1 \\ w_2 & \text{with corresponding numeric value } y_i = -1 \end{cases} \quad (2-3)$$

¹ High-dimensional

² Binary classification

³ Linear discriminant function

⁴ Decision rule

بنابراین همه نقاط آموزشی به طور صحیح طبقه‌بندی می‌شوند، اگر (معادله (۳-۳))

$$y_i (w^T x_i + w_0) > 0 \text{ for all } i \quad (3-3)$$

شکل ۱-۳ الف دو مجموعه مجزا را با ابرصفحه‌های جداکننده^۱ (A) نشان می‌دهد. به طور واضح ابرصفحه‌های جداکننده زیادی وجود دارد. طبقه‌بندی کننده حداکثر حاشیه، ابرصفحه‌ای را با بزرگ‌ترین فاصله تعیین می‌کند (شکل ۱-۳ ب). فرض حداکثر حاشیه خطا را پایین می‌آورد.

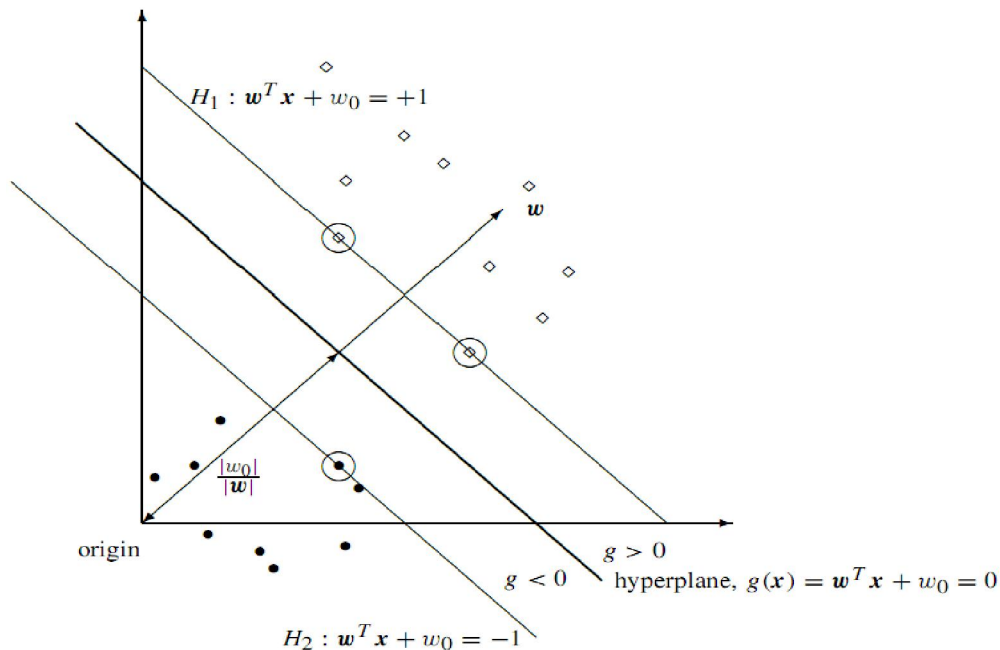
نکته قابل توجه این است که اگر داده‌های تعلیمی به صورت خطی تفکیک‌پذیر باشند، می‌توان دو ابرسطح مرزی را به گونه‌ای انتخاب کرد که هیچ داده‌ای بین آن‌ها نباشد و سپس فاصله بین این دو ابرسطح موازی را به حداکثر رساند. همچنین باید از قرار گرفتن نقاط داده در ناحیه درون مرز جلوگیری کرد، برای این کار یک محدودیت ریاضی به تعریف رابطه اضافه می‌شود، پس داریم (معادله (۴-۳)) [۶۶]:

$$\begin{aligned} w^T x_i + w_0 &\geq +1 \text{ for } y_i = +1 \\ w^T x_i + w_0 &\leq -1 \text{ for } y_i = -1 \end{aligned} \quad (4-3)$$

فاصله بین هر یک از این دو ابرصفحه و ابرصفحه جداکننده، $1/|w|$ است و حاشیه نامیده می‌شود. شکل ۲-۳ ابرصفحه جداکننده و ابرصفحه کانونی^۲ را برای دو مجموعه داده را نشان می‌دهد. نقاطی که روی ابرصفحه کانونی قرار گرفته‌اند بردارهای پشتیبان نامیده می‌شوند (دایره‌ها در شکل ۲-۳) یا به عبارتی نزدیک‌ترین داده‌های آموزشی به ابرصفحه‌های جدا کننده بردار پشتیبان نامیده می‌شوند [۶۶].

¹ Separating hyperplane

² Canonical hyperplanes



شکل ۳-۲- خطوط H_1 و H_2 ابرصفحه‌های کانونی هستند. حاشیه، فاصله عمودی بین ابرصفحه جداکننده ($g(x)=0$) و یک ابرصفحه گذرنده از نزدیک‌ترین نقاط (با دایره مشخص شده‌اند) است. این‌ها بردارهای پشتیبان نامیده می‌شوند [۶۶].

بنابراین بیشینه کردن حاشیه به این معنی است که یک راه حل برای کمینه کردن $|w|$ تحت شرایط معادله (۵-۳) پیدا کرد [۶۶]:

$$C1: \quad y_i (w^T x_i + w_0) \geq 1 \quad i = 1, \dots, n \quad (5-3)$$

یک رهیافت استاندارد در مسئله بهینه‌سازی با محدودیت‌های معادله و نامعادله^۱، فرمول لاگرانژ^۲ است [۶۷] که فرم اولیه تابع هدف را منجر می‌شود، L_p به صورت معادله (۶-۳) می‌باشد:

$$L_p = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i (y_i (w^T x_i + w_0) - 1) \quad (6-3)$$

¹ Equality and inequality constraints

² Lagrange formalism

که $\{\alpha_i, i = 1, \dots, n; \alpha_i \geq 0\}$ ضرایب لاگرانژ هستند. پارامترهای اولیه w و w_0 هستند و تعداد پارامترها برابر $p+1$ است که p بعد فضای مشخصه است.

راه حل مسئله کمینه کردن $w^T w$ تحت شرایط معادله (۳-۵) در حکم تعیین نقطه تعادل تابع L_p می‌باشد، در اینکه L_p با توجه به w و w_0 مینیم و با توجه به α_i بیشینه شود. مشتق L_p با توجه به w و w_0 برابر صفر قرار دادن حاصل می‌شود (معادله (۳-۷)) [۶۶]:

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (۳-۷)$$

$$w = \sum_{i=1}^n \alpha_i y_i x_i$$

با جایگذاری در معادله (۳-۶) فرم دوتایی^۱ لاگرانژ را می‌دهد (معادله (۳-۸)) [۶۶]:

$$L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (۳-۸)$$

که با توجه به α_i و تحت شرایط معادله (۳-۹) بیشینه می‌شود [۶۶]:

$$\alpha_i \geq 0 \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (۳-۹)$$

اهمیت فرم دوتایی (دوگان) در این است که معیار بهینه‌سازی را به عنوان حاصل الگوهای x_i نشان می‌دهد. این یک مفهوم کلیدی است و برای ماشین بردار پشتیبان غیرخطی اهمیت ویژه‌ای دارد. متغیرهای دوگان، ضرایب لاگرانژ، α_i و تعداد متغیرها n هستند [۶۶].

^۱ Dual form

۳-۱-۳ شرایط کاراش - کان - تاکر^۱

در بالا، مسئله اولیه در یک فرم دوگان دیگر که حل عددی آسان تری دارد، فرمول بندی مجدد می شود. شرایط کان - تاکر، شرایط لازم و کافی را برای کمینه کردن تابع هدف با محدودیت های معادله و نامعادله، مهیا می کند. در فرم اولیه تابع هدف داریم (معادله (۳-۱۰) [۶۶]:

$$\begin{aligned}\frac{\partial L_p}{\partial w} &= w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \\ \frac{\partial L_p}{\partial w_0} &= - \sum_{i=1}^n \alpha_i y_i = 0 \\ y_i (x_i^T w + w_0) - 1 &\geq 0 \\ \alpha_i &\geq 0 \\ \alpha_i (y_i (x_i^T w + w_0) - 1) &= 0\end{aligned}\quad (۳-۱۰)$$

در حالت خاص، شرایط $\alpha_i (y_i (x_i^T w + w_0) - 1) = 0$ (شرایط مکمل کاراش - کان - تاکر - حاصل ضرایب لاگرانژ و محدودیت نامعادله) اشاره دارد بر اینکه برای محدودیت های فعال (راه حل به صورت $y_i (x_i^T w + w_0) - 1 = 0$ می باشد) سپس $\alpha_i \geq 0$ ؛ در غیر این صورت برای محدودیت های غیرفعال $\alpha_i = 0$. برای محدودیت های فعال، ضرایب لاگرانژ حساسیت مقدار بهینه L_p را نشان می دهند [۶۸]. این نقاط داده با ضرایب لاگرانژ غیر صفر بر روی ابر صفحه کانونی قرار گرفته است.

۳-۱-۴ طبقه بندی

طراحی مجدد مسئله بهینه سازی مقید در فرم دوگان حل کننده برنامه ریزی غیر خطی عددی را به کار می گیرد. ضرایب لاگرانژ به دست آمده و مقدار w_0 از معادله (۳-۱۱) به دست می آید [۶۶]:

$$\alpha_i (y_i (x_i^T w + w_0) - 1) = 0 \quad (۳-۱۱)$$

^۱ Karush-Kuhn-Tucker conditions

با استفاده از هر بردار پشتیبان (الگوها برای $\alpha_i \neq 0$) یا یک میانگینی از همه بردارهای پشتیبان (معادله (۱۲-۳)) [۶۶].

$$n_{sv} w_0 + w^T \sum_{i \in SV} x_i = \sum_{i \in SV} y_i \quad (12-3)$$

که n_{sv} تعداد بردارهای پشتیبان و جمع بندی‌ها روی بردارهای پشتیبان، SV ، هستند. راه حل برای محاسبه w در معادله (۷-۳) به صورت معادله (۱۳-۳) می‌شود [۶۶]:

$$w = \sum_{i \in SV} \alpha_i y_i x_i \quad (13-3)$$

از این رو $\alpha_i = 0$ برای دیگر الگوها. بنابراین، بردارهای پشتیبان، ابرصفحه جداکننده را تعریف می‌کنند.

یک الگو جدید، x ، بر اساس معادله (۱۴-۳) طبقه‌بندی می‌شود [۶۶].

$$w^T x + w_0 \quad (14-3)$$

جایگذاری برای w و w_0 جداکننده خطی را می‌دهد: اختصاص x به w_1 اگر (معادله (۱۵-۳)) [۶۶].

$$\sum_{i \in SV} \alpha_i y_i x_i^T x - \frac{1}{n_{sv}} \sum_{i \in SV} \sum_{j \in SV} \alpha_i y_i x_i^T x_j + \frac{1}{n_{sv}} \sum_{i \in SV} y_i > 0 \quad (15-3)$$

۳-۱-۵ داده‌های جدانشدنی به طور خطی^۱

در بیشتر مسائل عملی و واقعی مرزهای غیرخطی برای جدا کردن کلاس‌ها وجود دارد و مسئله جستجو برای ابرصفحه جداکننده بهینه با مشکلاتی همراه است. اگر هم از بردارهای مشخصه

^۱ Linearly non-separable data

مصنوعی ($\phi(x)$) استفاده شود، برای انتقال داده به یک فضای مشخصه با ابعاد بالا در اینکه کلاس‌ها

به طور خطی جداپذیر شوند، به یک بیش برآزش داده و بنابراین توانایی تعمیم پایین می‌رسیم.

به هر حال ایده بالا را می‌توان به داده غیر قابل تفکیک با کم کردن محدودیت‌های معادله

(۳-۴) تعمیم داد. این موضوع توسط متغیرهای کمکی $\xi_i, i = 1, \dots, n$ با محدودیت‌های معادله

(۳-۱۶) بیان می‌شود [۶۶]:

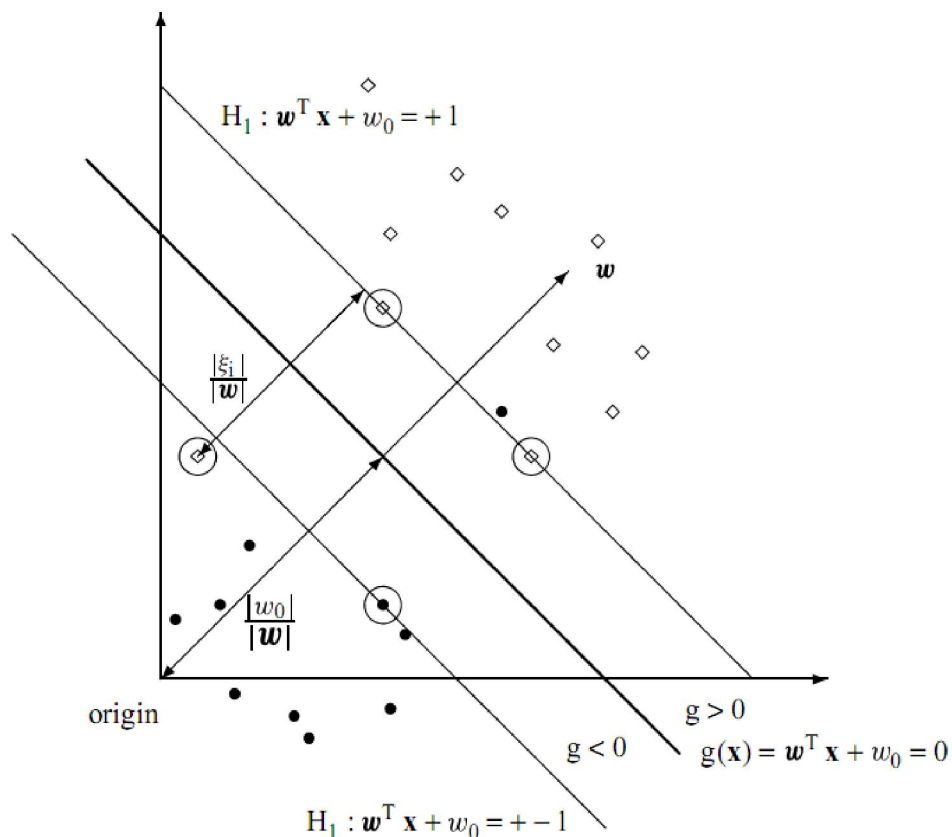
$$w^T x_i + w_0 \geq +1 - \xi_i \quad \text{for } y_i = +1$$

$$w^T x_i + w_0 \leq -1 + \xi_i \quad \text{for } y_i = -1$$

(۳-۱۶)

$$\xi_i \geq 0 \quad i = 1, \dots, n$$

برای یک نقطه با طبقه‌بندی اشتباه به وسیله ابرصفحه جداکننده، داریم $\xi_i > 1$ (شکل ۳-۳)



شکل ۳-۳- ابرصفحه جدا کننده خطی برای داده غیر قابل تفکیک [۶۶]

یک راه مناسب برای ترکیب کردن هزینه اضافی به علت تفکیک ناپذیری، نشان دادن یک عبارت هزینه اضافی در تابع هزینه با جایگزینی $w^T w / 2 + C \sum_i \xi_i$ با $w^T w / 2$ می باشد، که C ، پارامتر منظم سازی^۱ است. عبارت $C \sum_i \xi_i$ می تواند از اندازه گیری مقدار اندکی از طبقه بندی نادرست^۲ استدلال شود- مقدار کمتر از C ، پناستی^۳ کوچک تر برای نقاط دورافتاده و حاشیه نرم تر می باشد.

عبارت های دیگر پناستی هم ممکن است، مثل $C \sum_i \xi_i^2$

بنابراین معادله (۱۷-۳) را کمینه می کنیم [۶۶]:

$$\frac{1}{2} w^T w + C \sum_i \xi_i \quad (17-3)$$

با محدودیت معادله (۱۶-۳)، فرم دوگان لاگرانژ (معادله (۶-۳)) می شود (معادله (۱۸-۳)) [۶۶]:

$$L_p = \frac{1}{2} w^T w + C \sum_i \xi_i - \sum_{i=1}^n \alpha_i (y_i (w^T x_i + w_0) - 1 + \xi_i) - \sum_{i=1}^n r_i \xi_i \quad (18-3)$$

که $\alpha_i \geq 0$ و $r_i \geq 0$ ضرایب لاگرانژ هستند؛ r_i ها برای نشان دادن مثبت بودن ξ_i هستند.

مشتق گیری با توجه به w و w_0 نتایج زیر را نشان می دهد (معادله (۱۹-۳)) [۶۶]:

$$\begin{aligned} \sum_{i=1}^n \alpha_i y_i &= 0 \\ w &= \sum_{i=1}^n \alpha_i y_i x_i \end{aligned} \quad (19-3)$$

و مشتق گیری بر اساس ξ_i می دهد (معادله (۲۰-۳)) [۶۶]:

$$C - \alpha_i - r_i = 0 \quad (20-3)$$

¹ Regularisation parameter

² Misclassification

³ Penalty

جایگزینی نتایج معادله (۱۹-۳) در فرم اولیه معادله (۱۸-۳) و استفاده از معادله (۲۰-۳) فرم دوگان لاگرانژ را می‌دهد (معادله (۲۱-۳)) [۶۶]:

$$L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (۲۱-۳)$$

که فرم مشابه طبقه‌بندی کننده حداکثر حاشیه معادله (۸-۳) می‌باشد. این با توجه به α_i با شرایط زیر بیشینه شده است (معادله (۲۲-۳)) [۶۶]:

$$\begin{aligned} \sum_{i=1}^n \alpha_i y_i &= 0 \\ 0 \leq \alpha_i &\leq C \end{aligned} \quad (۲۲-۳)$$

شرط بعدی از معادله (۲۰-۳) و $r_i \geq 0$ تبعیت می‌کند. بنابراین تنها تغییر در مسئله بیشینه کردن، مرز بالا در α_i است.

شرایط مکمل کاراش-کان-تاکر به صورت معادله (۲۳-۳) می‌باشد [۶۶]:

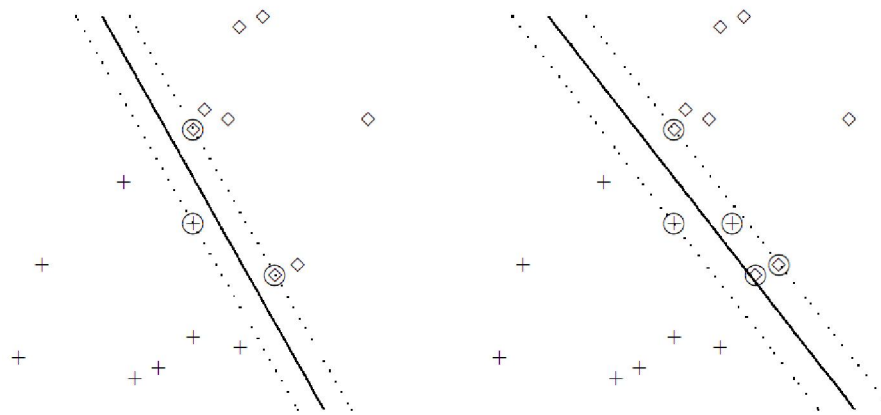
$$\begin{aligned} \alpha_i (y_i (x_i^T w + w_0) - 1 + \xi_i) &= 0 \\ r_i \xi_i = (C - \alpha_i) \xi_i &= 0 \end{aligned} \quad (۲۳-۳)$$

الگوها برای $\alpha_i > 0$ بردارهای پشتیبان نامیده می‌شوند. این‌ها در شرایط $0 < \alpha_i < C$ و $\xi_i = 0$ هستند و در یک فاصله $1/|w|$ از ابرصفحه جداکننده قرار می‌گیرند (به بردارهای پشتیبان، بردارهای حاشیه نیز گفته می‌شود). متغیرهای کمکی غیر صفر می‌تواند تنها در زمانی که $\alpha_i = 0$ است، باشند. در این مورد نقاط x_i در طبقه‌بندی نادرست هستند، اگر $\xi_i > 1$ باشد. اگر $\xi_i < 1$ باشد، طبقه‌بندی درست است، اما نزدیک‌تر از $1/|w|$ به ابرصفحه جداکننده قرار گرفته است. به مانند مورد تفکیک‌پذیر، مقدار w_0 با استفاده از شرط اول بالا و هر بردار پشتیبان تعیین شده است و یا به وسیله مجموع نمونه‌ها برای $0 < \alpha_i < C$ (برای $\xi_i = 0$) (معادله (۱۲-۳)) و w با معادله (۱۳-۳) داده شده است. این می‌دهد (معادله (۲۴-۳)) [۶۶]:

$$w_0 = \frac{1}{N_{\mathcal{S}V}} \left\{ \sum_{i \in \mathcal{S}V} y_i - \sum_{i \in \mathcal{S}V, j \in \mathcal{S}V} \alpha_i y_i x_i^T x_j \right\} \quad (24-3)$$

که $\mathcal{S}V$ مجموعه‌ای از بردارهای پشتیبان با مقادیر پیوسته از α_i با شرط $0 < \alpha_i < C$ و $\frac{1}{\mathcal{S}V}$ مجموعه‌ای از بردارهای پشتیبان $N_{\mathcal{S}V}$ با شرط $0 < \alpha_i < C$ هستند (در فاصله هدف $1/|w|$ از ابرصفحه جداکننده).

شکل ۳-۴ ابرصفحه جداکننده بهینه را برای داده تفکیک‌پذیر به طور خطی و داده تفکیک ناپذیر نشان می‌دهد. بردارهای پشتیبان ($\alpha_i > 0$) با دایره مشخص شده‌اند. همه نقاطی که بردار پشتیبان نیستند در بیرون از نوار حاشیه قرار گرفته‌اند ($\alpha_i = 0, \xi_i = 0$). در سمت راست عکس یکی از بردارهای پشتیبان (از کلاس +) به طور نادرست طبقه‌بندی شده است ($\xi_i > 1$) [۶۶].



شکل ۳-۴- داده تفکیک‌پذیر به طور خطی (چپ) و داده تفکیک ناپذیر (راست)، $C=20$ که C پارامتر منظم‌سازی

است [۶۶]

۳-۱-۶ ماشین‌های بردار پشتیبان چند کلاسی

ماشین‌های بردار پشتیبان می‌توانند در مسائل چند کلاسی^۱ به صورت‌های زیر به کار روند: الف- طبقه‌بندی کننده دوتایی در وضعیت یک در برابر همه^۲ یا، ب- وضعیت یک در برابر یک^۳ و یا استفاده از، ج- توابع جداکننده خطی به صورت همزمان [۶۹].

توابع جداکننده خطی به صورت معادله (۳-۲۵) می‌باشد [۶۶]:

$$g_k(x) = (w^k)^T x + w_0^k \quad k = 1, \dots, C \quad (۳-۲۵)$$

یک راه حل برای $\{(w^k, w_0^k), k = 1, \dots, C\}$ جستجو می‌شود که قاعده تصمیم به صورت معادله (۳-۲۶) باشد: اختصاص x به کلاس w_i اگر [۶۶]

$$g_i(x) = \max_j g_j(x) \quad (۳-۲۶)$$

داده آموزشی بدون خطا را جدا کند. راه حل‌هایی برای $\{(w^k, w_0^k), k = 1, \dots, C\}$ وجود دارد، به طوری که برای همه $k = 1, \dots, C$ داشته باشیم (معادله (۳-۲۷)) [۶۶]

$$(w^k)^T x + w_0^k - ((w^j)^T x + w_0^j) \geq 1 \quad (۳-۲۷)$$

برای همه $x \in w_k$ و برای همه $j \neq k$. این به این معنی است که هر جفت کلاس قابل تفکیک هستند. اگر یک راه حل ممکن باشد، راه حلی جستجو می‌شود برای اینکه (معادله (۳-۲۸)) [۶۶]

$$\sum_{k=1}^C (w^k)^T w^k \quad (۳-۲۸)$$

^۱ Multiclass

^۲ One-against-all

^۳ One-against-one

کمینه شود. اگر داده آموزشی نتواند جدا شود، متغیرهای کمکی مطرح می‌شوند و تابع (۲۹-۳) را کمینه می‌کنیم [۶۶]:

$$L = \sum_{k=1}^C (w^k)^T w^k + C \sum_{i=1}^n \xi_i \quad (29-3)$$

با محدودیت زیر (معادله (۳۰-۳)) [۶۶]:

$$(w^k)^T x_i + w_0^k - ((w^j)^T x_i + w_0^j) \geq 1 - \xi_i \quad (30-3)$$

برای همه x_i ها (که $x_i \in w_k$)، و برای همه $j \neq k$. فرآیند کمینه کردن کمی L با شرط محدودیت‌های نامعادله، مشابه آنچه که برای مورد دو کلاسی توسعه پیدا کرده است، می‌باشد [۶۶].

۷-۱-۳ ماشین‌های بردار پشتیبان غیرخطی

برای داده‌هایی که به طور خطی تفکیک‌پذیر نیستند، تابع $\phi(x)$ برای انتقال فضای مشخصه به کار می‌رود.

برای مسئله طبقه‌بندی دوتایی، یک تابع تشخیص (جداکننده) به فرم زیر جستجو می‌شود (معادله (۳۱-۳)) [۶۶]:

$$g(x) = w^T \phi(x) + w_0 \quad (31-3)$$

با قاعده تصمیم (معادله (۳۲-۳)) [۶۶]:

$$w^T \phi(x) + w_0 \begin{cases} > 0 \\ < 0 \end{cases} \Rightarrow x \in \begin{cases} w_1 \text{ with corresponding} \\ \text{numeric value } y_i = +1 \\ w_2 \text{ with corresponding} \\ \text{numeric value } y_i = -1 \end{cases} \quad (32-3)$$

فرآیند SVM حداکثر حاشیه را از طریق بیشینه کردن لاگرانژ تعیین می‌کند. فرم دوگان لاگرانژ

(معادله (۳۱-۳)) می‌شود (معادله (۳۳-۳)) [۶۶]:

$$L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \phi^T(x_i) \phi(x_j) \quad (۳۳-۳)$$

که $y_i = \pm 1, i = 1, \dots, n$ مقادیر شاخص کلاس هستند و $\alpha_i, i = 1, \dots, n$ ضرایب لاگرانژ با شرایط زیر

می‌باشند (معادله (۳۴-۳)) [۶۶]:

$$\begin{aligned} 0 &\leq \alpha_i \leq C \\ \sum_{i=1}^n \alpha_i y_i &= 0 \end{aligned} \quad (۳۴-۳)$$

برای یک پارامتر منظم‌سازی C. بیشینه کردن معادله (۳۳-۳) به شرط محدودیت‌های معادله (۳۴-۳)

منجر به بردارهای پشتیبان یکی شده با مقادیر غیر صفر α_i می‌شود.

راه حل برای w به صورت معادله (۳۵-۳) می‌باشد [۶۶]:

$$w = \sum_{i \in SV} \alpha_i y_i \phi(x_i) \quad (۳۵-۳)$$

و طبقه‌بندی یک نمونه داده جدید x به صورت معادله (۳۶-۳) و (۳۷-۳) می‌باشد [۶۶]:

$$g(x) = \sum \alpha_i y_i \phi^T(x_i) \phi(x) + w_0 \quad (۳۶-۳)$$

که

$$w_0 = \frac{1}{N_{\tilde{SV}}} \left\{ \sum_{i \in \tilde{SV}} y_i - \sum_{i \in SV, j \in \tilde{SV}} \alpha_i y_i \phi^T(x) \phi(x_j) \right\} \quad (۳۷-۳)$$

که SV مجموعه‌ای از بردارهای پشتیبان با مقادیر پیوسته از α_i با شرط $0 < \alpha_i < C$ و $\frac{1}{\sqrt{N}}$ مجموعه‌ای از بردارهای پشتیبان $N_{\frac{1}{\sqrt{N}}}$ با شرط $0 < \alpha_i < C$ هستند (در فاصله هدف $1/|w|$ از ابرصفحه جداکننده).

بهینه‌سازی L_D معادله (۳۳-۳) و طبقه‌بندی متوالی یک نمونه (معادله (۳۶-۳) و (۳۷-۳)) تنها روی حاصل ضرب اسکالر^۱ بین بردارهای مشخصه انتقال یافته تکیه دارد، که می‌تواند با تابع کرنل^۲ جایگزین شود (معادله (۳۸-۳)) [۶۶]:

$$k(x, y) = \phi^T(x) \phi(y) \quad (38-3)$$

بنابراین می‌توان از محاسبه تبدیل $\phi(x)$ اجتناب کرد و به جای حاصل ضرب اسکالر، تابع کرنل $K(x, y)$ را جایگزین کرد. تابع تشخیص معادله (۳۶-۳) می‌شود (معادله (۳۹-۳)) [۶۶]:

$$g(x) = \sum_{i \in SV} \alpha_i y_i k(x_i, x) + w_0 \quad (39-3)$$

مزیت ارائه کرنل در این است که تنها نیاز به استفاده از K در الگوریتم آموزشی می‌باشد و نیازی به دانستن ϕ به طور صریح نیست. تابع کرنل می‌تواند به صورت ضرب داخلی نشان داده شود. در بعضی از موارد (برای مثال، کرنل نمایی)، فضای مشخصه بی‌نهایت بعدی است و استفاده از کرنل مؤثرتر می‌باشد.

۳-۱-۸ انواع کرنل

انواع زیادی از کرنل‌ها وجود دارد که می‌تواند در SVM مورد استفاده قرار گیرد. جدول ۳-۱ بعضی از فرم‌های استفاده شده عمومی را نشان می‌دهد. کرنل‌های قابل قبول باید قابل نشان دادن به عنوان

^۱ Scalar products

^۲ Kernel function

یک ضرب داخلی در فضای مشخصه باشند [۶۹] و [۷۰]: یک کرنل $K(x, y), x, y \in \mathbb{R}^p$ ، یک ضرب

داخلی در برخی فضای مشخصه است، یا $K(x, y) = \phi^T(x)\phi(y)$ اگر و تنها اگر

$$K(x, y) = K(y, x) \quad \text{و (معادله (۳-۴۰)) [۶۶]}$$

$$\int k(x, z) f(x) f(z) dx dz \geq 0 \quad (۳-۴۰)$$

جدول ۳-۱- کرنل‌های ماشین بردار پشتیبان [۶۶]

رابطه غیرخطی	فرم ریاضی $K(x, y)$
چند جمله‌ای	$(1 + x^T y)^d$
گوسی	$p(- x - y ^2 / \sigma^2)$
سیگموئیدی	$\tanh(kx^T y - \delta)$

برای همه توابع f با شرط (معادله (۳-۴۱)) [۶۶]

$$\int f^2(x) dx < \infty \quad (۳-۴۱)$$

که $K(x, y)$ ممکن است به صورت معادله (۳-۴۲) بسط داده شود [۶۶]:

$$k(x, y) = \sum_{j=1}^{\infty} \lambda_j \hat{\phi}_j(x) \hat{\phi}_j(y) \quad (۳-۴۲)$$

که λ_j و $\hat{\phi}_j(x)$ مقادیر ویژه و توابع ویژه با شرایط معادله (۳-۴۳) می‌باشد [۶۶]:

$$\int k(x, y) \phi_j(x) dx = \lambda_j \phi_j(x) \quad (43-3)$$

و $\hat{\phi}_j(x)$ نرمالایز شده، چنانچه $\int \hat{\phi}_j^2(x) dx = 1$.

به عنوان یک مثال، کرنل $K(x, y) = (1 + x^T y)^d$ را برای $d=2$ و $x, y \in \mathbb{R}^2$ در نظر بگیرید.

بسط زیر به دست می‌آید (معادلات (3-44) و (3-45) [66]:

$$\begin{aligned} (1 + x_1 y_1 + x_2 y_2)^2 = \\ 1 + 2x_1 y_1 + 2x_2 y_2 + 2x_1 x_2 y_1 y_2 + x_1^2 y_1^2 + x_2^2 y_2^2 = \phi^T(x) \phi(y) \end{aligned} \quad (44-3)$$

که

$$\phi(x) = (1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2, x_1^2, x_2^2) \quad (45-3)$$

3-1-9 انتخاب مدل

درجات آزادی^۱ مدل SVM، انتخاب کرنل، پارامترهای کرنل و انتخاب پارامتر منظم‌سازی، C ، می‌باشد. برای انواع کرنل، پیدا کردن مقدار پارامترهای آن برای کلاس‌های قابل تفکیک، ممکن است. به هر حال این یک استراتژی قابل پیش بینی نیست و منتهی به بیش برازش^۲ داده آموزشی و تعمیم ضعیف به داده مشاهده نشده می‌شود.

ساده‌ترین رهیافت در انتخاب مدل اختصاص دادن یک مجموعه ارزیابی است که در نمایش عملکرد مثل پارامترهای مدل متنوع هستند. گزینه‌های دیگر که استفاده بهتری از داده دارند، روش‌های باز نمونه‌گیری داده از قبیل واریسی اعتبار و خودگردان‌سازی هستند [66].

^۱ Degrees of freedom

^۲ Over-fitting

۳-۱-۱۰ مروری بر مبانی روش TSVM

استنتاج تبدیلی (ترانس داکتیو) اولین بار توسط وپنیک در سال ۱۹۹۸ مطرح شد [۶۹]. تفاوت بین یادگیری تبدیلی و یادگیری سنتی این است که کلاسیفایر سنتی ابتدا توابع را بر اساس آموزش مجموعه نمونه تولید می‌کند. در حالی که کلاسیفایر ترانس داکتیو به طور مستقیم آموزش مجموعه نمونه‌ها را برای تشخیص برچسب‌های کلاس نمونه‌های تست استفاده می‌کند [۴۴].

برای یک کار یادگیری $P(\vec{x}, y) = P(y | \vec{x})P(\vec{x})$ یاد گیرنده L یک فرضیه فضایی H از توابع $h: X \rightarrow \{-1, 1\}$ می‌باشد. نمونه S_{train} از n مثال آموزشی است (معادله (۳-۴۶)) [۴۵].

$$(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_n, y_n) \quad (۳-۴۶)$$

هر مثال یادگیری شامل یک بردار سند $\vec{x} \in X$ و یک علامت دو تایی $y \in \{-1, +1\}$ است. نمونه S_{test} از k مثال تست می‌باشد (معادله (۳-۴۷)) [۴۵]:

$$\vec{x}_1^*, \vec{x}_2^*, \dots, \vec{x}_k^* \quad (۳-۴۷)$$

یاد گیرنده ترانس داکتیو L یک تابع $h_L = L(S_{\text{train}}, S_{\text{test}})$ از H با استفاده از S_{train} و S_{test} انتخاب می‌کند و تعداد مورد انتظار خطاهای پیش بینی شده به صورت معادله (۳-۴۸) است [۴۵]:

$$R(L) = \int \frac{1}{k} \sum_{i=1}^k \Theta(h_L(\vec{x}_i^*, y_i^*) - 1) dP(\vec{x}_1, y_1) \dots dP(\vec{x}_k^*, y_k^*) \quad (۳-۴۸)$$

معادله (۳-۴۸) روی مثال‌های تست کمینه است. اگر $a=b$ باشد $\Theta(a, b)$ صفر و در غیر این صورت یک است. وپنیک (۱۹۹۸) مرز انحراف نسبی یکسان آموزش خطا را مشخص کرد (معادله (۳-۴۹)) [۴۵]:

¹ Learner

$$R_{train}(h) = \frac{1}{n} \sum_{i=1}^n \Theta(h(\vec{x}_i), y_i) \quad (49-3)$$

و برای خطای تست (معادله (3-50)) [45]:

$$R_{test}(h) = \frac{1}{k} \sum_{i=1}^k \Theta(h(\vec{x}_i^*), y_i^{true}) \quad (50-3)$$

با احتمال $1-\eta$ (معادله (3-51)) [45]:

$$R_{test}(h) \leq R_{train}(h) + \Omega(n, k, d, \eta) \quad (51-3)$$

که فاصله اطمینان $\Omega(n, k, d, \eta)$ وابسته به تعداد مثال‌های آموزش n ، تعداد مثال‌های تست k و دیمانسیون VC پارامتر d از H می‌باشد (برای جزئیات بیشتر، [69]).

این مسئله استنباط ترانس داکتیو، به نظر نمی‌رسد که تفاوت زیادی از حالت استنتاج معمولی برای یادگیری ماشین، داشته باشد. یاد دادن می‌تواند یک نقش تصمیم بر اساس داده‌های آموزشی و کاربرد آن در تست داده‌ها داشته باشد. با این حال برای حل کردن مسئله تخمین مقادیر دوتایی k $(y_1^*, \dots, y_k^*)$ نیاز به حل کردن مسئله پیچیده‌تر تخمین یک تابع روی یک فضای پیوسته ممکن می‌باشد. این رویه، زمانی که اندازه n در نمونه آموزشی کوچک است بهترین راه حل نیست [45].

چه اطلاعاتی از مطالعه نمونه تست به دست می‌آید و چگونه می‌توان از آن استفاده کرد؟ آموزش و نمونه تست، فضای فرضی H^1 در یک تعداد محدود از کلاس‌های هم‌ارز H'^2 را از هم جدا می‌کند. دو تابع H در صورتی که هر دو طبقه‌بندی آموزش و نمونه تست در یک مسیر مشابه باشند، به یک کلاس هم‌ارز تعلق دارند. این، مشکل یادگیری برای پیدا کردن یک تابع در مجموعه نامحدود H را با پیدا کردن یکی از کلاس‌های هم‌ارز محدود H' ، کاهش می‌دهد. با استفاده از این کلاس‌های

¹ Hypothesis space

² Equivalence classes

هم ارز می توان یک ساختار صعودی دیمانسیون- VC برای کمینه کردن ساختار ریسک ایجاد کرد

[۶۹] (معادله (۳-۵۲))

$$H'_1 \subset H'_2 \subset \dots \subset H' \quad (۳-۵۲)$$

بر خلاف حالت استنتاجی، موقعیت نمونه تست را در زمانی که ساختار تعریف می شود، می توان مطالعه کرد. با استفاده از دانش قبلی درباره طبیعت $P(\mathbf{x}, \mathbf{y})$ می توان یک ساختار مناسب تر و یک یادگیری سریع تر ساخت. در حالت خاص، می توان ساختار را بر اساس مرز ابر صفحه مجزا بر روی هر دوی داده های آموزش و تست ساخت. وپنیک نشان داد که با اندازه مرز می توان تعداد بیشینه کلاس های هم ارز را کنترل کرد (یعنی بعد- VC) [۴۵].

ابر صفحه $h(\mathbf{x}) = \text{sign}\{\mathbf{x}^T \mathbf{w} + b\}$ به عنوان فضای فرضی H در نظر گرفته می شود. اگر بردارهای مشخصه یک نمونه آموزشی و یک نمونه تست در یک توپ با قطر D باشند، سپس حداکثر

$$N_r < \exp(d(\frac{n+k}{d} + 1)), d = \min(a, [\frac{D^2}{\rho^2}] + 1) \quad (۳-۵۳)$$

کلاس های هم ارز وجود دارد که شامل یک ابر صفحه مجزا با معادله (۳-۵۴) می باشد [۴۵].

$$\forall_{i=1}^n \left| \frac{\vec{w}}{\|\vec{w}\|} \cdot \vec{x}_i + b \right| \geq \rho \quad \forall_{j=1}^k \left| \frac{\vec{w}}{\|\vec{w}\|} \cdot \vec{x}_j^* + b \right| \geq \rho \quad (۳-۵۴)$$

(یعنی، مرز بزرگ تر یا مساوی ρ). پارامتر a بعدیت فضایی و $[b]$ قسمت صحیح b است.

توجه این که بعد VC لزوماً وابسته به تعداد مشخصه ها نیست، اما می تواند خیلی پایین تر از بعدیت^۱ فضا باشد. این ساختار، بر اساس مرز ابر صفحه مجزا استفاده می شود. کمینه کردن ساختار ریسک نتیجه می دهد که می توان کوچک ترین مرز در خطای تست را گرفت، در صورتی که کلاس هم

^۱ Dimensionality

ارز از عنصر ساختاری H'_i کمینه شود. برای مسائل مجزا به طور خطی، مسئله بهینه‌سازی زیر ناشی می‌شود [۶۹] (معادله (۵۵-۳)).

OP 1(TSVM (lin.sep.case))

Minimize over $(y_1^*, \dots, y_n^*, \vec{w}, b):$

$$\frac{1}{2} \|\vec{w}\|^2$$

(۵۵-۳)

subject to $\forall_{i=1}^n : y_i [\vec{w} \cdot \vec{x}_i + b] \geq 1$

$$\forall_{j=1}^k : y_j^* [\vec{w} \cdot \vec{x}_j^* + b] \geq 1$$

حل این مسئله، پیدا کردن یک برجسب زنی $y_1^*, \dots, y_k^*$ از داده‌های تست و یک ابر صفحه $\langle \vec{w}, b \rangle$ می‌باشد. بنابراین ابر صفحه داده‌های آموزشی و داده‌های تست را با حداکثر مرز از هم جدا می‌کند. شکل ۵-۳ این موضوع را نشان می‌دهد. برای داده‌های تفکیک ناپذیر، می‌توان متغیر کمکی ξ_i را به طور مشابه با *SVM* استقرایی انجام داد (معادله (۵۶-۳)) [۴۵].

OP 2 (TSVM (non-sep.case))

Minimize over $(y_1^*, \dots, y_n^*, \vec{w}, b, \xi_1, \dots, \xi_n, \xi_1^*, \dots, \xi_k^*):$

$$\frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^n \xi_i + C^* \sum_{j=1}^k \xi_j^*$$

subject to $\forall_{i=1}^n : y_i [\vec{w} \cdot \vec{x}_i + b] \geq 1 - \xi_i$

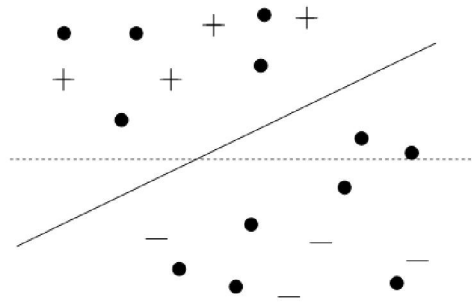
(۵۶-۳)

$$\forall_{j=1}^k : y_j^* [\vec{w} \cdot \vec{x}_j^* + b] \geq 1 - \xi_j^*$$

$$\forall_{i=1}^n : \xi_i > 0$$

$$\forall_{j=1}^k : \xi_j^* > 0$$

پارامترهای C و C^* توسط کاربر تعیین می‌شوند. آن‌ها اجازه می‌دهند که اندازه مرز بر خلاف طبقه‌بندی نادرست مثال‌های آموزشی یا به استثنای مثال‌های تست مشخص شود [۴۵].



شکل ۳-۵- بیشینه مرز ابر صفحه. مثال‌های مثبت و منفی با $+/-$ و مثال‌های تست با نقطه علامت‌گذاری شده‌اند. خط تیره حل‌های SVM است. خطوط پر، طبقه‌بندی ترانس داکتیو را نشان می‌دهد [۴۵].

برای یک تعداد کمی از نمونه‌ها، این فرآیند می‌تواند کامل شود. برای یک تعداد بزرگی از نمونه‌های برچسب نخورده، ضروری است که یک تنوعی از الگوریتم بهینه‌سازی تقریبی برای رسیدن به راه حل، استفاده کرد [۴۴].

در حال حاضر، مهم‌ترین نتیجه در این رابطه، TSVM مطرح شده توسط جواچیمز می‌باشد. در زیر اصول و مفهوم الگوریتم TSVM بیان می‌شود.

فرآیند آموزش TSVM حل مسئله بهینه‌سازی بالاست. الگوریتم آموزشی می‌تواند به مراحل زیر تقسیم شود:

مرحله اول: تنظیم کردن پارامتر C و C^* ، و استفاده از SVM استقرایی روی آموزش داده و به دست آوردن کلاسیفایر اولیه. تنظیم کردن تعدادی از نمونه‌ها با برچسب مثبت، N ، بر اساس یک قاعده [۴۴].

مرحله دوم: محاسبه مقادیر تابع تصمیم همه نمونه‌های برچسب نخورده با کلاسیفایر اولیه. نمونه‌های برچسب N با بزرگ‌ترین مقادیر تابع تصمیم به عنوان مثبت و بقیه به عنوان منفی. تنظیم یک فاکتور اثر موقت C_{tmp}^* [۴۴].

مرحله سوم: آموزش دوباره SVM روی همه نمونه‌ها. برای کلاسیفایر حاصل شده جدید، تغییر برچسب‌ها از یک جفت نمونه برچسب متفاوت و برچسب نخورده با استفاده از یک قاعده معین در

ساختن مقدار تابع هدف در کاهش هرچه بیشتر. این مرحله تکرار می‌شود تا هیچ جفت نمونه‌ای با تغییر شرایط عوض نشود [۴۴].

مرحله ۴: افزایش مقدار C_{tmp}^* به طور اندک و بازگشت به مرحله ۳. زمانی که $C_{tmp}^* > C^*$ ، الگوریتم تمام شده و نتیجه آن خروجی مورد نظر است [۴۴].

۳-۱-۱۱ الگوریتم TSVM بر اساس خوشه‌بندی طیفی (TSVMSC)

چون الگوریتم‌های TSVM از ایده یادگیری ترانس داکتیو به طور مؤثر استفاده می‌کند، بنابراین آن می‌تواند توزیع اطلاعات نمونه‌های برچسب نخورده و آموزش نمونه‌ها را به خوبی ترکیب کند. بنابراین در مقایسه با الگوریتم SVM سنتی، الگوریتم TSVM دقت طبقه‌بندی بالاتری دارد [۴۴].

به هر حال الگوریتم TSVM هنوز یک سری نواقص دارد، از قبیل تعداد N از نمونه‌های برچسب خورده مثبت در نمونه‌های برچسب نخورده برای الگوریتم TSVM باید به طور مصنوعی تعیین شوند، اما تخمین قابل قبول برای مقدار N معمولاً دشوار است [۴۴].

الگوریتم TSVM از یک روش ساده برای تخمین مقدار N استفاده می‌کند که به صورت تخمین نسبت نمونه‌ها با برچسب‌های مثبت به همه نمونه‌های برچسب نخورده بر اساس نسبت نمونه‌ها با برچسب مثبت به همه نمونه‌های برچسب خورده می‌باشد. به هر حال زمانی که تعداد نمونه‌ها با برچسب‌ها کوچک است، تخمین یک مقدار دقیق از N مشکل است. وقتی که مقدار پیش تنظیم N به طور کامل از تعداد نمونه‌های با برچسب‌های مثبت متفاوت است، عملکرد الگوریتم TSVM خیلی ناپایدار خواهد شد و علاوه بر این دقت طبقه‌بندی الگوریتم نمی‌تواند به طور مؤثر مورد اطمینان قرار گیرد [۴۴].

چون عملکرد الگوریتم TSVM خیلی ناپایدار است، بنابراین، TSVM بر اساس خوشه‌بندی طیفی مطرح شده است که به اختصار TSVMSC نامیده شده است.

اساس TSVMSC به شرح زیر است. اول از همه، خوشه کردن نمونه‌های برچسب نخورده مشابه بر اساس عملکرد خوب خوشه‌بندی طیفی می‌باشد. سپس همه نمونه‌ها در خوشه یکسان با برچسب مشابه علامت‌گذاری می‌شوند. سرانجام با TSVM روی همه نمونه‌های برچسب خورده و برچسب نخورده یادگیری را انجام می‌دهند که پیدا کردن حل یک مسئله بهینه‌سازی می‌باشد. در زیر مراحل TSVMSC به صورت جزئی آمده است:

مرحله ۱: تعیین تعداد خوشه‌بندی k (معمولاً $3 < k < 7$) و تعیین پارامترهای C و C^* .

مرحله ۲: استفاده از الگوریتم خوشه‌بندی طیفی در خوشه کردن مجموعه نمونه و تقسیم آن به k خوشه.

مرحله ۳: علامت‌گذاری همه نمونه‌ها در خوشه یکسان با برچسب مشابه و جستجو 2^k مورد در مجموعه نمونه در حل مسئله بهینه‌سازی.

مرحله ۴: خروجی برچسب‌های کلاس (طبقه) از نمونه‌های برچسب نخورده [۴۴].

۲-۳ روش‌های خوشه‌بندی

روش‌های خوشه‌بندی برای داده کاوی و تهیه پیش‌الگوها برای طبقه‌کننده‌های نظارت شده مورد استفاده قرار می‌گیرند. خوشه‌بندی را می‌توان به عنوان مهم‌ترین مسئله در یادگیری بدون نظارت در نظر گرفت. خوشه‌بندی با یافتن یک ساختار درون یک مجموعه از داده‌های بدون برچسب درگیر است. خوشه به مجموعه‌ای از داده‌ها گفته می‌شود که به هم شباهت داشته باشند. در خوشه‌بندی سعی می‌شود تا داده‌ها به خوشه‌هایی تقسیم شوند که شباهت بین داده‌های درون هر خوشه حداکثر و شباهت بین داده‌های درون خوشه‌های متفاوت حداقل شود.

برای تحلیل خوشه‌ها رویه یکتایی وجود ندارد و می‌توان با استفاده از الگوریتم‌های مختلفی اشیاء را خوشه‌بندی کرد. اصطلاح خوشه را نمی‌توان به طور دقیق تعریف کرد و معنای دقیق آن به

محل استفاده آن مرتبط است و یک از دلایلی که تعداد زیادی الگوریتم خوشه‌بندی وجود دارد همین است. خوشه‌بندی به یافتن مجموعه‌ای از خوشه‌ها گفته می‌شود که در بیشتر مواقع همه اشیاء را شامل می‌شوند. خوشه‌بندی را می‌توان به دو نوع تقسیم‌بندی کرد [۷۱]:

خوشه‌بندی سخت: هر شیء متعلق به یک خوشه است یا نه.

خوشه‌بندی نرم (فازی): هر شیء با درجه خاصی متعلق به یک خوشه است.

تقسیم‌بندی‌های دیگری هم برای خوشه‌بندی وجود دارد، به عنوان مثال:

خوشه‌بندی محض: هر شیء دقیقاً متعلق به یک خوشه است.

خوشه‌بندی محض با اشیاء پرت: اشیاء می‌توانند متعلق به هیچ خوشه‌ای نباشند و در این صورت، شیء پرت نامیده می‌شوند.

خوشه‌بندی همراه با هم‌پوشانی: اشیاء می‌توانند به بیش از یک خوشه تعلق داشته باشند.

خوشه‌بندی سلسله مراتبی: اشیائی که به خوشه فرزند تعلق دارند، به خوشه پدر هم تعلق دارند.

۳-۲-۱ قواعد کلی خوشه‌بندی

با این که تعداد زیادی الگوریتم خوشه‌بندی وجود دارند، بهتر است قواعدی کلی مستقل از الگوریتمی خاص برای خوشه‌بندی در نظر بگیریم. در هر مسئله خوشه‌بندی، تعدادی داده داریم که می‌خواهیم بر اساس معیاری خاص آن‌ها را خوشه‌بندی کنیم. آن معیار را فاصله می‌نامیم به این معنی که بین هر دو داده، یک مقدار فاصله وجود دارد. تعداد داده‌ها را n در نظر می‌گیریم. داده‌ها را با استفاده از مجموعه $\{1, \dots, n\}$ بر چسب گذاری می‌کنیم و این مجموعه را S می‌نامیم. فاصله بین نقاط به ما یک تابع فاصله $d: S \times S \rightarrow R$ می‌دهد. برای هر $i, j \in S$ ، داریم [۷۱]:

$$d(i, j) \geq 0, \quad d(i, j) = d(j, i), \quad d(i, j) = 0 \Leftrightarrow i = j$$

یعنی باید فاصله‌ها تقارنی و نامنفی باشند و فاصله دو نقطه فقط زمانی صفر است که آن دو نقطه در واقع یک نقطه باشند. خوشه‌بندی باید نسبت به تغییر مقیاس، تغییر نکند. یعنی اگر خوشه‌بندی خاصی برای یک مجموعه داده با تابع فاصله d را در نظر بگیرید و اسکالر $\alpha > 0$ را در تابع فاصله ضرب کرده و تابع فاصله جدید $d' = \alpha d$ به دست آید، همان خوشه‌بندی قبلی با تابع فاصله d ، برای داده‌های مسئله با تابع فاصله d' هم یک خوشه‌بندی مناسب باشد. این ویژگی به این معنی است که تابع نسبت به کاهش یا افزایش خطی فواصل بین نقاط تغییری نمی‌کند که باعث می‌شود تابع مستقل از مقیاس اندازه‌گیری فاصله باشد [۷۱].

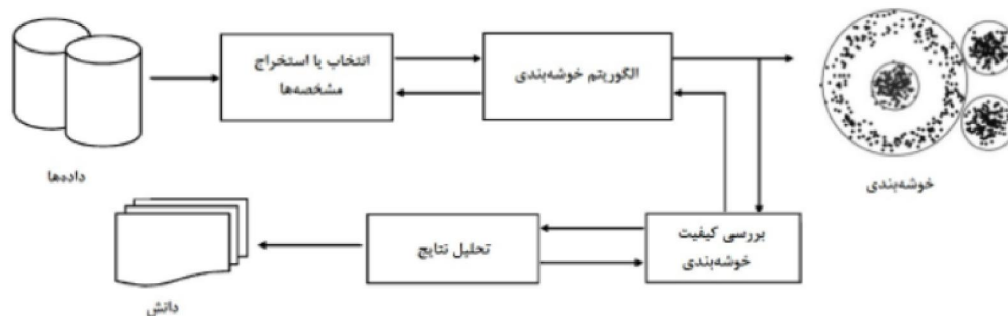
ویژگی دیگر این است که در خوشه‌بندی، نباید تناقض وجود داشته باشد و این بدان معنی است که اگر فاصله بین نقاط درون یک خوشه را کاهش و فاصله بین نقاط در خوشه‌های متفاوت را افزایش دهیم، خوشه‌بندی تحت تابع فاصله جدید همان خوشه‌بندی قبلی باشد. به عبارت دیگر یک خوشه‌بندی روی داده‌های مجموعه S با تابع فاصله d را در نظر بگیرید. حال تابع فاصله d' را به صورت زیر تعریف می‌کنیم:

برای هر $i, j \in S$ که متعلق به یک خوشه هستند: $d'(i, j) \leq d(i, j)$

برای هر $i, j \in S$ که متعلق به خوشه‌های متفاوتی هستند: $d'(i, j) \geq d(i, j)$

تناقض نداشتن خوشه‌بندی یعنی همان خوشه‌بندی قبلی با تابع فاصله d ، برای داده‌های

مسئله با تابع فاصله d' هم یک خوشه‌بندی مناسب باشد [۷۱].



شکل ۳-۶- بخش‌های معمول یک سیستم خوشه‌بندی [۷۲]

به طور سنتی یک سیستم خوشه‌بندی شامل بخش‌های نشان داده در شکل ۳-۶ است [۷۲].

۲-۲-۳ روش خوشه‌بندی k-means

این الگوریتم پارامتر k را به عنوان ورودی گرفته و مجموعه n شیء را به k خوشه افراز می‌کند. به طوری که سطح شباهت داخلی خوشه‌ها بالا بوده و سطح شباهت اشیاء بیرون خوشه‌ها پایین باشد. شباهت هر خوشه نسبت به متوسط اشیاء آن خوشه سنجیده شده که این متوسط مرکز خوشه نامیده می‌شود. این الگوریتم به صورت زیر کار می‌کند:

ورودی: k ، تعداد خوشه‌ها است. پایگاه داده X شامل n شیء،

$$X = \{x_1, x_2, \dots, x_n\}, x_i \in R^m$$

خروجی: یک مجموعه از k خوشه که معیار مربع خطا را حداقل می‌کند.

الگوریتم:

- ۱- به صورت تصادفی k داده دلخواه را به عنوان مراکز خوشه‌های ابتدایی انتخاب می‌کنیم
- ۲- هر داده را با توجه به نزدیکی آن به مراکز خوشه‌ها، به خوشه‌ها تخصیص می‌دهیم. معیار نزدیکی را فاصله اقلیدوسی در نظر می‌گیریم (معادله (۳-۵۷)):

$$d(x_i, m_j) = \|x_i - m_j\| = \sqrt{\sum_{k=1}^m (x_{ik} - m_{jk})^2} \quad (۳-۵۷)$$

- ۳- که در آن x_i ، i امین داده ورودی و m_j مرکز (میانگین) خوشه j ام است. داده فوق در خوشه ای قرار می‌گیرد که کمترین فاصله را با مرکز آن خوشه داشته باشد. مرکز خوشه نیز میانگین حسابی داده‌های آن خوشه می‌باشد.

- ۴- مراکز خوشه‌ها را به روز می‌کنیم یعنی برای هر خوشه میانگین اعضای خوشه را به دست می‌آوریم.

۵- با توجه به مراکز جدید خوشه‌ها به مرحله ۲ برمی‌گردیم و فرآیند فوق را تا جایی ادامه می‌دهیم که هیچ تغییری در خوشه‌ها رخ ندهد (در این حالت الگوریتم پایان یافته است) [۷۳].

در واقع هدف کمینه کردن معادله (۵۸-۳) می‌باشد.

برای مراکز خوشه‌ها $\{\mu_j, j = 1, \dots, k\}$ داریم [۶۶]:

$$\sum_{j=1}^k S_j \quad (58-3)$$

که در آن S_j مجموع مربعات برای خوشه j به صورت معادله (۵۹-۳) تعریف می‌شود [۶۶]:

$$S_j = \sum_{i=1}^n z_{ji} |x_i - \mu_j|^2 \quad (59-3)$$

که در آن $z_{ji} = 1$ است، اگر در گروه j باشد و در غیر این صورت برابر صفر است. پارامتر μ_j برابر میانگین گروه j است (معادله (۶۰-۳)) [۶۶]:

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^n z_{ji} x_i \quad (60-3)$$

از مزایای این الگوریتم به سادگی آن می‌توان اشاره کرد، همچنین این الگوریتم همگرایی سریعی در رسیدن به جواب نهایی دارد و از معایب آن نیز می‌توان به این اشاره کرد که چون در ابتدا تعداد خوشه‌ها به الگوریتم داده می‌شود بهینه بودن جواب به دست آمده را نمی‌توان تضمین کرد، چون که تعداد خوشه‌ها به صورت دستی است و اشتباه در تعیین این عدد در جواب نهایی تأثیر زیادی دارد. همچنین کارایی الگوریتم بسیار وابسته به نحوه تعیین مراکز اولیه برای خوشه‌هاست [۷۳].

۳-۲-۳ خوشه‌بندی Fuzzy c-means یا FCM

یکی از مهم‌ترین و پرکاربردترین الگوریتم‌های خوشه‌بندی، الگوریتم c میانگین می‌باشد. در این الگوریتم نمونه‌ها به c خوشه تقسیم می‌شوند و تعداد c از قبل مشخص شده است. در نسخه فازی این الگوریتم نیز تعداد خوشه‌ها (c) از قبل مشخص شده است (در فرمول‌های زیر تعداد خوشه‌ها g در نظر گرفته شده است). الگوریتم FCM برای پارامترهای $y_{ji} (i = 1, \dots, n; j = 1, \dots, g)$ حلی به صورت معادله (۶۱-۳) دارد (تابع هدف) [۶۶]: که در آن مراکز خوشه‌ها و y_{ji} مقادیر عضویت خوشه‌ها است.

$$J_r = \sum_{i=1}^n \sum_{j=1}^g y_{ji}^r |x_i - m_j|^2 \quad (۶۱-۳)$$

که تحت شرایط زیر کمینه می‌شود (از روی y_{ji} می‌توان یک ماتریس y تعریف کرد که دارای n سطر و g ستون می‌باشد و مؤلفه‌های آن هر مقداری بین ۰ تا ۱ را می‌توانند اختیار کنند. اگر تمامی مؤلفه‌های ماتریس y به صورت ۰ و یا ۱ باشند الگوریتم مشابه c میانگین کلاسیک خواهد بود. با این که مؤلفه‌های ماتریس y می‌توانند هر مقداری بین ۰ تا ۱ را اختیار کنند اما مجموع مؤلفه‌های هر یک از ستون‌ها باید برابر ۱ باشد) (معادله (۶۲-۳)) [۶۶]:

$$\sum_{j=1}^g y_{ji} = 1 \quad 1 \leq i \leq n$$

$$y_{ji} \geq 0 \quad i = 1, \dots, n; \quad j = 1, \dots, g \quad (۶۲-۳)$$

معنای این شرط این است که مجموع تعلق هر نمونه به c خوشه باید برابر ۱ باشد. پارامتر y_{ji} درجه تابع عضویت آمین شیء با آمین خوشه را نشان می‌دهد. پارامتر r یک عدد حقیقی است که فازی بودن را کنترل می‌کند ($r \geq 1$). اگر $r=1$ قرار دهیم تابع هدف خوشه‌بندی c میانگین غیر فازی به

دست می‌آید. پارامتر x_i نمونه i ام و پارامتر m_j مرکز ژامین خوشه است و n تعداد نمونه‌ها (اشیاء) می‌باشد.

برای به دست آوردن فرمول‌های مربوط به y_{ji} و m_j باید تابع هدف تعریف شده را کمینه کنیم. با استفاده از شرط فوق و برابر صفر قرار دادن مشتق تابع هدف خواهیم داشت (معادلات (۶۳-۳) و (۶۴-۳)) [۶۶]:

$$y_{ji} = \frac{1}{\sum_{k=1}^g \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{r-1}}} \quad (۶۳-۳)$$

$$m_j = \frac{\sum_{i=1}^n y_{ji}^r x_i}{\sum_{i=1}^n y_{ji}^r} \quad (۶۴-۳)$$

با استفاده از دو معادله (۶۳-۳) و (۶۴-۳) الگوریتم خوشه‌بندی c میانگین فازی را می‌توان به صورت زیر بیان کرد:

مراحل الگوریتم:

۱- انتخاب r ($1 < r < \infty$) مقدار دهی اولیه برای y_{ji} ($i = 1, \dots, n; j = 1, \dots, g$)

۲- مراکز خوشه‌ها محاسبه شوند (محاسبه m_j ها).

۳- محاسبه d_{ij} که $d_{ij} = |x_i - m_j|$ $i = 1, \dots, n; j = 1, \dots, g$

۴- محاسبه تابع عضویت: اگر $d_{il} = 0$ باشد برای مقادیری از l آنگاه $y_{li} = 1$ است و

$y_{ji} = 0$ است برای همه $j, j \neq l$ در غیر این صورت با توجه به معادله (۶۳-۳) حل می‌شود.

۵- اگر همگرا نبود، به مرحله ۲ برگرد.

یکی از شرط‌های توقف الگوریتم زمانی است که تغییرات نسبی در مقادیر مرکز کوچک

می‌شود (معادله (۶۵-۳)) [۶۶]:

$$D_z = \left\{ \sum_{j=1}^g |m_j(k) - m_j(k-1)|^2 \right\}^{\frac{1}{2}} < \varepsilon \quad (65-3)$$

که $m_j(k)$ مقدار ژامین مرکز روی k ژامین تکرار و ε حد مشخص شده است. شرطهای توقف دیگر بر اساس مقادیر تابع عضویت، y_{ji} یا تابع هزینه (j_r) می باشد. شرط دیگر بر اساس بهینه سازی محلی تابع هزینه که توسط سلیم و اسماعیل^۱ (۱۹۸۶) مطرح شده اند، است [۷۴]. زمانی توقف داریم که (معادله (۶۶-۳) و (۶۷-۳)):

$$\max_{1 \leq i \leq n} \alpha_i < \varepsilon \quad (66-3)$$

که در آن

$$\alpha_i = \max_{1 \leq j \leq g} y_{ji}^{r-1} |x_i - m_j|^2 - \min_{1 \leq j \leq g} y_{ji}^{r-1} |x_i - m_j|^2 \quad (67-3)$$

تا این که در یک کمینه محلی $\alpha_i = 0, i = 1, \dots, n$.

۴-۲-۳ الگوریتم گوستافسون کسل

الگوریتم گوستافسون-کسل فازی، تعمیم الگوریتم میانگین فازی با تغییری از فاصله اقلیدوسی با معادله (۶۸-۳) است:

$$d_{GK}^2 = (x_k - v_i)^T A_i (x_k - v_i) \quad (68-3)$$

که A_i با استفاده از وارون ماتریس کوواریانس فازی برای هر خوشه محاسبه می شود. پارامتر T ترانهاده ماتریس می باشد.

الگوریتم گوستافسون کسل فازی با معیار معادله (۶۹-۳) کمینه می شود:

¹ Selim and Ismail

$$Q = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^m d_{GK}^2 = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^m (x_k - v_i)^T A_i (x_k - v_i) \quad (۶۹-۳)$$

بهینه‌سازی این معیار تابع محدودیت‌های استاندارد معادله (۷۰-۳) است:

$$\forall i, k \quad u_{ik} \in [0, 1], \quad \forall i \quad 0 < \sum_{k=1}^N u_{ik} < N, \text{ and } \forall k \quad \sum_{i=1}^c u_{ik} = 1. \quad (۷۰-۳)$$

ماتریس A_i بر اساس وارون ماتریس کوواریانس فازی C_i محاسبه می‌شود (معادله (۷۱-۳)):

$$A_i = (\rho_i | C_i |)^{1/d} C_i^{-1}, \quad C_i = \frac{\sum_{k=1}^N u_{ik}^m (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^N u_{ik}^m} \quad (۷۱-۳)$$

بدیهی است که C_i باید وارون‌پذیر باشد. در غیر این صورت الگوریتم هیچ جوابی را نمی‌دهد [۷۵]. با جایگزین کردن فاصله اقلیدسی با ماهولونوبیس^۱ (که توسط یک ماتریس متقارن، معین و مثبت ایجاد شده) در الگوریتم میانگین فازی می‌توان به شناسایی خوشه‌های بیضوی نیز دست یافت. این پیشنهاد توسط گوستافسون و کسل در سال ۱۹۷۹ ارائه شد [۷۶].

۳-۲-۵ مقایسه طبقه‌بندی و خوشه‌بندی

الف - طبقه بندی

طبقه بندی جزو روش‌های یادگیری با ناظر می‌باشد. در این روش سیستم یادگیر از طریق مجموعه داده‌های آموزشی سعی در پیش بینی داده‌های جدید دارد.

- یادگیری بر اساس داده‌های آموزشی انجام می‌شود و با داده‌های تستی سیستم ارزیابی می‌گردد

- روش‌هایی مانند SVM از این نوع می‌باشد

¹ Mahalanobis Distance

- از نوع یادگیری با ناظر است و مجموعه جواب گسسته می‌باشد
- تعداد کلاس‌ها معلوم است
- نمونه‌ها هم معلوم مربوط به کدام کلاس‌ها هستند (برچسب کلاس مشخص است)
- ویژگی‌ها یا مشخصه‌های نمونه‌ها را نیز می‌دانیم

ب-خوشه‌بندی

خوشه‌بندی دیتا تکنیک اصلی در روش یادگیری بدون ناظر می‌باشد، در این روش ما به دنبال یادگیری الگوها می‌باشیم.

- مجموعه آموزشی یا تست نداریم
 - از طریق ویژگی‌های مجموعه داده‌ها به دنبال الگوهای پنهان در دیتا می‌گردیم
 - خوشه‌بندی نسبت به طبقه‌بندی دو گام عقب‌تر است
 - تعداد کلاس‌ها مشخص نیست
 - نمی‌دانیم کدام نمونه‌ها مربوط به کدام کلاس‌ها هستند
 - تعدادی نمونه داریم و هر نمونه تعدادی ویژگی یا مشخصه دارد
- در واقع تفاوت و شباهت خوشه‌بندی و طبقه‌بندی (رده‌بندی) مانند تفاوت و شباهت بین ترکیب شیمیایی و ترکیب فیزیکی مواد می‌باشد. در ترکیب شیمیایی یک ماده جدید تولید می‌شود و ماده جدیدی با توجه به مواد اولیه به وجود می‌آید و در ترکیب فیزیکی فقط مواد از لحاظ فیزیکی ترکیب می‌شوند، ولی هر دوی این فرآیندها از فرآیندهای بنیادی علم شیمی می‌باشند. در خوشه‌بندی، فقط بر اساس شباهت‌های اشیاء (منظور رکوردهای داده‌ای می‌باشد) آن‌ها را به گروه‌های تقسیم می‌کنیم ولی در طبقه‌بندی (رده‌بندی) از رکوردهای موجود استفاده می‌نماییم تا بتوانیم مسئله دیگری را حل نماییم.

فصل چہارم

فرمولاسیون مسئلہ

۱-۴ ماتریس تشابه در تحلیل خوشه‌ای طیفی

شکل ۱-۴ الف دو مدل مصنوعی با توزیع‌های پارامتر فیزیکی مختلف در محیط‌های لایه‌ای را نشان می‌دهد. هر دو مدل ورودی در یک فضای پارامتری دو بعدی به کار برده می‌شود (شکل ۱-۴ ب). سپس تشابهات جفت جفت بین نمونه‌ها در فضای پارامتری دو بعدی تعیین خواهد شد. به طور کلی، تشابه، با فاصله یا مقدار اندازه‌گیری شده بین دو نمونه در فضای n بعدی نسبت عکس دارد. در اینجا از معیار تشابه گوسین^۱ $s_{i,j}$ استفاده شده است تا رابطه همسایگی محلی بین دو نمونه i و j را در فضای پارامتری دو بعدی به وسیله بردارهای موقعیت d_i و d_j بیان کند (معادله (۱-۴) [۲]).

$$s_{i,j} = \exp\left(\frac{-\|d_i - d_j\|^2}{2\sigma^2}\right) \quad (۱-۴)$$

که در آن $s_{i,j}$ ماتریس تشابه است. فاصله اقلیدوسی^۲ بین دو نمونه توسط پارامتر σ وزن دار می‌شود که این، کاهش تشابه را با افزایش فاصله نمونه در فضای پارامتری، کنترل می‌کند. پارامتر σ به صورت تجربی تعیین می‌شود. اگر بین ۱ تا ۱۰۰ انتخاب شود، یک اثر ضعیف روی نتایج خوشه‌بندی داده‌های نرمال شده دارد. زمانی که داده‌ها غیر نرمال هستند، انتخاب σ خیلی مشکل است. باید دقت کرد که انتخاب σ ، نتیجه یکتایی ماتریس تشابه S با همه $s_{i,j}$ ها نیست.

تشابهات جفت جفت بین نمونه‌ها در یک فضای n بعدی، معمولاً توسط گراف‌های متشابه [۲۶]

با استفاده از گره‌ها^۳ (رئوس) و لبه‌های^۴ (اضلاع) مرتبط با دو گره، مدل می‌شود. تعداد گره‌ها در یک گراف برابر تعداد نمونه‌های موجود یا سلول مدل‌ها در فضای پارامتری یا مدل‌های ورودی، می‌باشد. تعداد لبه‌ها ممکن است تغییر کند، اما حداکثر مقدار ممکن از لبه‌ها موقعی است که هر گره به همه

^۱ Gaussian similarity

^۲ Euclidian distance

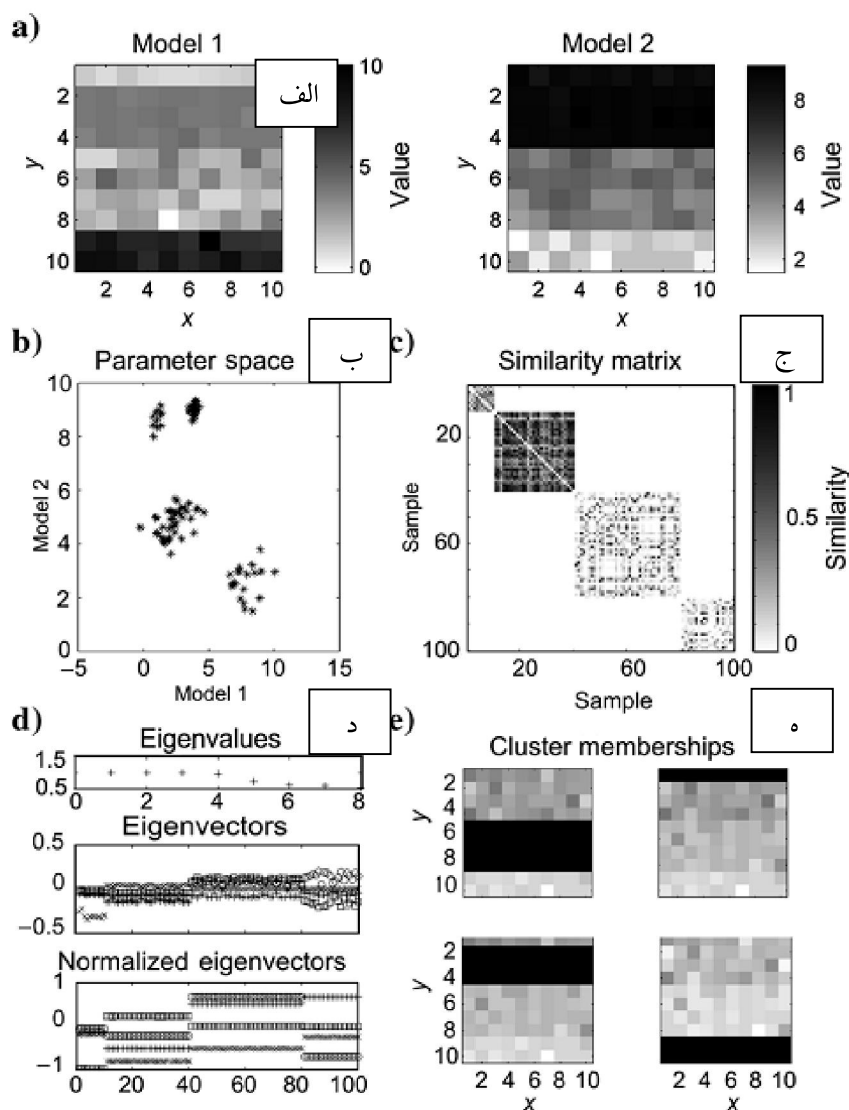
^۳ Nodes

^۴ Edges

گره‌های دیگر متصل باشد. در صورت نبود وابستگی جهتی در تشابهات جفت جفت، گراف غیر جهت‌دار می‌شود. تشابهات جفت جفت (معادله (۱-۴)) در گراف‌های غیر جهت‌دار، عناصری از ماتریس تشابه S هستند که در واقع وزن‌دار شده‌ی ماتریس مجاورتی گراف می‌باشند. این یک ماتریس مربعی متقارن است که بعد^۱ آن برابر تعداد گره‌های گراف است. سطر اول شامل تشابهات جفت نمونه d_1 با نمونه‌های $d_1, d_2, \dots, d_t$ می‌باشد که t تعداد نمونه‌های موجود می‌باشد (سلول‌های مدل). سطر دوم شامل تشابهات جفت جفت d_2 با همه نمونه‌های دیگر است. توجه این که وجود ۴ لایه غالب در مدل‌های ورودی، به وضوح در ماتریس تشابه مثال مورد نظر^۲، قابل مشاهده است (شکل ۱-۴ ج [۲]).

^۱ Dimension

^۲ Toy example



شکل ۴-۱- مثالی از تحلیل خوشه‌ای طیفی با استفاده از مثال مصنوعی. الف) دو مدل ورودی جدا از هم در مجاورت یکدیگر که توزیع‌های پارامتر فیزیکی مختلف و جدا از هم را نشان می‌دهد. ب) پارامتر فضایی دو بعدی به دست آمده از مدل‌های ورودی. ج) ماتریس تشابه، فواصل اقلیدوسی وزن‌دار شده را بین نقاط در فضای پارامتری دو بعدی نرمال شده، نشان می‌دهد. د) اولین هشت مقدار ویژه و چهار بردار ویژه از ماتریس به دست آمده در قسمت (ج). ه) توزیع عضویت فازی برای یک جواب چهار خوشه بر اساس بردار ویژه به دست آمده در قسمت (د) [۲].

به هر حال تولید ماتریس تشابه، یک مرحله حساس در خوشه‌بندی طیفی می‌باشد، زیرا این ماتریس، ناهمگنی ساختاری موجود در همه مدل‌های ورودی را شامل می‌شود. انتخاب یک اندازه تشابه مناسب و تعداد تشابهات جفت جفت، تحت تأثیر نتایج خوشه‌بندی نهایی است. برای مثال به جای استفاده از یک گراف کلی که از ماتریس تشابه کاملاً مقید شده به دست می‌آید، می‌توان از گراف‌های دیگر که بر

اساس یک انتخاب مقدماتی (پیش انتخاب) از تشابهات جفت جفت می‌باشد، استفاده کرد. مثلاً گراف نزدیک‌ترین همسایگی k و گراف نزدیک‌ترین همسایگی متقابل k . در اینجا تشابهات جفت فقط برای k نمونه که حداکثر تشابه و یا تشابه متقابل با d_i ها دارند، مورد ملاحظه قرار می‌گیرد. تشابه گره‌های ناپیوسته در گراف با صفر مشخص شده‌اند [۱۷]. بر اساس نتایج آزمایشی، به طور تجربی، مقدار k تقریباً معادل تعداد نمونه‌ها (سلول‌های مدل) تقسیم بر ۱۰ پیشنهاد شد [۲].

۲-۴ معادلات خوشه‌بندی طیفی

ابزار اصلی برای خوشه‌بندی طیفی، ماتریس لاپلاسین گراف است. در تئوری گراف طیفی، لاپلاسین‌های متفاوت گراف وجود دارد که در خوشه‌بندی طیفی استفاده می‌کنند. در اینجا از الگوریتم خوشه‌بندی طیفی نگ^۱ و همکاران (۲۰۰۲) استفاده شد و لاپلاسین نرمال شده L از گراف تشابه که از ماتریس تشابه به دست آمده است، محاسبه گردید (معادله (۲-۴)) [۲].

$$L = A \frac{-1}{2} S A \frac{-1}{2} \quad (2-4)$$

که S ماتریس تشابه و A درجه ماتریس گراف تشابه را نشان می‌دهد و یک ماتریس قطری با عناصر زیر می‌باشد (معادله (۳-۴)) [۲]:

$$a_{i,i} = \sum_{j=1}^t S_{i,j} \quad (3-4)$$

با توجه به الگوریتم نگ و همکاران (۲۰۰۲)، مقدار ویژه (طیف L ، شکل ۱-۴ د) و بردار ویژه متناظر L محاسبه گردید. توجه این که تعداد عناصر در هر بردار ویژه برابر تعداد سلول‌های موجود در

^۱ Ng

داده‌های ورودی می‌باشد. برای خوشه‌بندی داده‌ها در c زیر مجموعه (در مثال مصنوعی $c=4$)، c بردار ویژه متناظر با بزرگ‌ترین c مقدار ویژه انتخاب شد و یک ماتریس $t \times c$ (تعداد سطر ماتریس است) به اسم U تولید گردید که ستون‌های آن c بردار ویژه و عناصر آن $u_{i,j}$ می‌باشد. به صورت متوالی ماتریس B توسط نرمال کردن سطرهای ماتریس U به نرم واحد تشکیل گردید که عناصر آن به صورت زیر می‌باشد (معادله (۴-۴)) [۲]:

$$b_{i,j} = \frac{u_{i,j}}{(\sum_{j=1}^c u_{i,j}^2)^{1/2}} \quad (4-4)$$

برای $i=1, \dots, t$ ، اگر b_i برداری مطابق با i امین سطر از ماتریس B تعریف شود، موقعیت‌های نمونه‌ها در فضای c -بعدی نرمال شده، با اولین بردار ویژه c در ماتریس L تعریف می‌شود. سرانجام همه نمونه‌های $b_{i,j}$ تحت تحلیل خوشه‌ای تقسیم‌بندی کننده فازی قرار می‌گیرند. تحلیل خوشه‌ای به صورت مکرر انجام شده و مراکز خوشه را به صورت تصادفی مقدار دهی می‌کند تا عضویت‌های خوشه بهینه با حداقل انحراف پیدا شود. نتایج الگوریتم فازی در یک مجموعه از عضویت‌های خوشه در شکل ۴-۱ ه نشان داده شده است. عضویت‌های خوشه می‌تواند به عنوان ورودی برای خارج کردن از حالت فازی به کار برده شود (شکل ۱-۱ را ببینید) [۲].

الگوریتم‌های خوشه‌بندی طیفی زیادی وجود دارد. همه آن‌ها در این مشترک‌اند که گراف‌های تشابه با روابط همسایگی محلی مدل بین نمونه‌های تعریف شده توسط بردارهای موقعیت چندبعدی، ایجاد می‌شوند. بر اساس تجربه، در حال حاضر هیچ قاعده‌ای برای انتخاب مناسب اندازه تشابه وجود ندارد. به هر حال ویژگی محلی تحلیل تشابه، زمانی که روندها و آشفتگی‌های سیستماتیک در مدل مورد نظر وجود دارد، حائز اهمیت می‌باشد. علاوه بر استفاده از اندازه تشابه مناسب، تعداد لبه‌های مورد نظر (تشابهات جفت جفت) در گراف تشابه در نتایج اثر بخش می‌باشد. از طرف دیگر هیچ قاعده مشخصی برای انتخاب یک پیوستگی مناسب برای گراف‌ها موجود نیست. از این گذشته محاسبه

لاپلاسن گراف برای الگوریتم‌های خوشه‌ای طیفی متفاوت، متغیر است. اغلب در همه الگوریتم‌های خوشه‌بندی طیفی، مقدار ویژه طیف از لاپلاسن گراف محاسبه می‌شود و انتخاب بردار ویژه برای تحلیل خوشه‌ای متوالی، بر اساس آن می‌باشد. بر اساس تجربه، خوشه‌بندی طیفی بسیار انعطاف پذیر است و می‌تواند با هر نوع داده‌ای برای شناسایی الگوها و ساختارهای غالب موجود در داده‌های چند متغیره، سازگاری داشته باشد. به هر حال مطابق تحلیل خوشه‌ای تقسیم‌بندی کننده استاندارد، تصمیم‌های زیادی برای تنظیمات الگوریتمی توسط کاربر می‌تواند گرفته شود. بیشترین حساسیت را انتخاب اندازه تشابه مناسب یا لاپلاسن گراف دارد [۷۷]. در حال حاضر هیچ قاعده خاصی برای انتخاب تنظیمات مناسب وجود ندارد. این دلایل، تأثیرات دانش محوری را در نتایج خوشه‌بندی در تحلیل خوشه‌ای تقسیم‌بندی کننده، افزایش می‌دهد [۲].

۳-۴ روش شاخص دان

روش شاخص دان یکی از روش‌های دقیق کمی برای محاسبه تعداد خوشه‌های بهینه در الگوریتم‌های مختلف خوشه‌بندی است. فاصله بین خوشه‌های C_k و $C_{k'}$ با محاسبه فاصله بین نزدیک‌ترین نقاط آنها اندازه‌گیری می‌شوند (معادله (۴-۵)) [۷۸].

$$d_{kk'} = \min_{\substack{i \in I_k \\ j \in I_{k'}}} \|M_i^{\{k\}} - M_j^{\{k'\}}\| \quad (۵-۴)$$

و $d_{\min}$ کوچک‌ترین فاصله $d_{kk'}$ یا به عبارتی کمینه فاصله (معمولاً اقلیدوسی) بین نقاط در خوشه‌های مختلف است (معادله (۴-۶)):

$$d_{\min} = \min_{k \neq k'} d_{kk'} \quad (۶-۴)$$

برای هر خوشه C_k ، معادله D_k به عنوان بزرگ‌ترین فاصله نقاط مشخص در خوشه (قطر خوشه هم نامیده می‌شود)، تعریف می‌شود (معادله (۷-۴)):

$$D_k = \max_{\substack{i, j \in I_k \\ i \neq j}} \|M_i^{\{k\}} - M_j^{\{k\}}\| \quad (۷-۴)$$

و $d_{\max}$ بزرگ‌ترین فاصله D_k یا به عبارتی بیشینه فاصله (معمولاً اقلیدوسی) بین نقاط در خوشه‌های مختلف است (معادله (۸-۴)):

$$d_{\max} = \max_{1 \leq k \leq K} D_k \quad (۸-۴)$$

شاخص دان با کسر زیر تعریف می‌شود (معادله (۹-۴)):

$$c = \frac{d_{\min}}{d_{\max}} \quad (۹-۴)$$

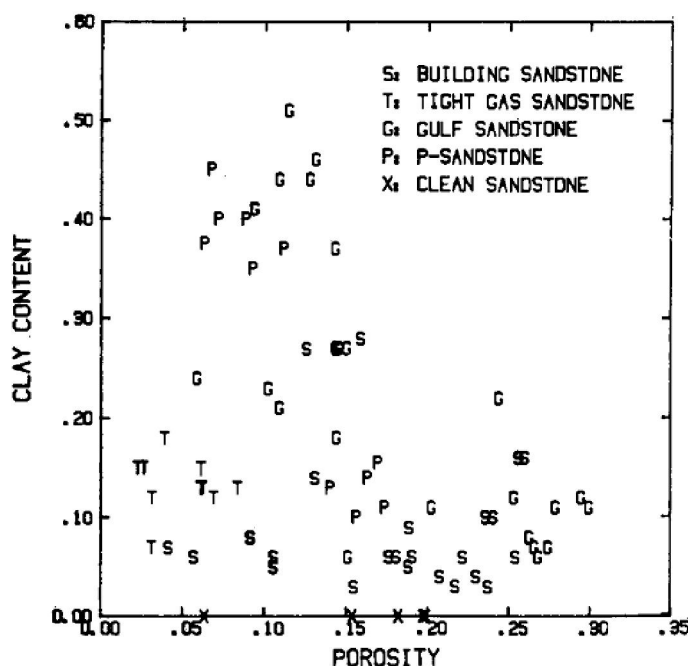
۴-۴ رابطه تجربی هان و همکاران برای محاسبه سرعت

لرزه‌ای

۱-۴-۴ توصیف نمونه

در این مطالعه ۷۵ نمونه ماسه سنگ که از مغزه‌های چاه‌ها یا کانسنگ‌ها استخراج شده، استفاده گردیده است. تخلخل نمونه‌ها بین ۲ تا ۳۰ درصد و محتوای رس از ۰ تا ۵۰ درصد تغییر می‌کند

(شکل ۲-۴). ۱۰ نمونه ماسه سنگ مخزن گازی متراکم^۱ (T) با تخلخل پایین هستند. ۲۴ نمونه از چاه‌های داخل دریا در خلیج مکزیک^۲ (G) هستند که بعضی از این نمونه‌ها تحکیم نیافته^۳ هستند. ۱۱ نمونه (P) از مغزه‌های چاه هستند که خوب تحکیم یافته^۴ هستند. ۳۰ نمونه سفت و محکم (S) از معدن شن و ماسه هستند که ۵ تا آن (X) ماسه سنگ تمیز با محتوای رس^۵ کمتر از یک درصد می‌باشد. شکل ۲-۴ محتوای رس را در برابر تخلخل برای همه نمونه‌ها نشان می‌دهد. قابل توجه است که نمونه‌هایی با بیشتر از ۲۰ درصد محتوای رس، تخلخل‌های پایینی در حد ۵ تا ۱۵ درصد دارند [۳۴].



شکل ۲-۴- تغییرات محتوای رس و تخلخل برای ۷۵ نمونه ماسه سنگ. تغییرات تخلخل از ۲ تا ۳۰ درصد و تغییرات محتوای رس از ۰ تا ۵۰ درصد می‌باشد. داده‌ها نشان می‌دهند که ماسه سنگ‌ها با محتوای رس بالا، تخلخل پایینی دارند [۳۴].

¹ Tight gas sandstones
² Gulf of Mexico
³ Poorly consolidated
⁴ Well-consolidated
⁵ Clay content

۲-۴-۴ اندازه‌گیری صوتی

سرعت‌های موج تراکمی یا V_p و برشی یا V_s و دامنه A_p اولین پیک ورودی موج تراکمی^۱، به عنوان توابع فشار و وضعیت اشباع از آب اندازه‌گیری شدند. فشار محبوس کردن (همه جانبه)^۲ P_c و فشار منفذی^۳ P_p به طور جدا از هم کنترل می‌شوند. با فشار تفاضلی^۴ P_d که تا ۵۰ مگا پاسکال و فشار تخلخل که تا یک مگا پاسکال محدود شده است، سیستم می‌تواند شرایط فشار را تا عمق بالای ۲۰۰۰ متر شبیه‌سازی کند [۳۴]. سرعت‌ها در فرکانس‌های مافوق صوت با استفاده از روش انتقال پالس^۵ اندازه‌گیری شدند [۷۹]. فرکانس‌های مرکزی مبدل‌ها برای امواج P و S به ترتیب ۱ و ۰/۶ مگاهرتز استفاده شد. طول موج برای موج‌های P و S بزرگ‌تر از ۵ میلی‌متر و حداقل ۵ برابر بزرگ‌تر از میانگین اندازه ذرات نمونه‌ها است. بر روی نمونه‌ها، ابتدا اندازه‌گیری‌ها در حالت خشک خلا (فشار خلا کمتر از ۰/۰۱ تور^۶ (۱/۳۳ پاسکال)) و سپس در اشباع شدگی کامل با آب انجام گرفت. در بعضی نمونه‌ها با محتوای رس بالا، آب شور به عنوان سیال داخل فضاها ی خالی برداشت می‌شود. تفاوت سرعت بین حالت اشباع با آب شور و اشباع با آب خالص وجود ندارد. نمونه‌ها دارای ۵ سانتی‌متر قطر، ۲ تا ۵ سانتی‌متر طول می‌باشند.

زمان ورود موج P در حدود ۰/۰۳ میکروثانیه می‌باشد که منجر به خطای مطلق کمتر از یک درصد در اندازه‌گیری V_p می‌شود. خطای اندازه‌گیری V_s کمتر از ۲ درصد است به جز برای ماسه سنگ تحکیم نیافته در فشار تفاضلی پایین ($P_d < 10 \text{ MPa}$) که خطا به علت سیگنال ضعیف ممکن است بالای ۳ درصد باشد [۳۴].

^۱ Compressional

^۲ Confining pressure

^۳ Pore pressure

^۴ Differential pressure

^۵ Pulse transmission

^۶ Torr

سرعت‌ها در طول بارگذاری و تخلیه^۱ سیکل فشار اندازه‌گیری می‌شوند. بعد از اولین سیکل فشار، سرعت‌ها در طول سیکل تخلیه اندازه‌گیری می‌شوند و با تعادل رطوبتی کمتر از ۱ درصد در نمونه‌های خوب تحکیم یافته و کمتر از ۲ درصد در نمونه‌های تحکیم نیافته تکرار پذیر هستند [۳۴].

۳-۴-۴ اثرات تخلخل و محتوای رس بر روی سرعت‌های V_s و V_p

بر اساس ترکیب داده‌های صوتی و پتروگرافی^۲، اثرات فشار همه جانبه P_c ، فشار منفذی P_p و اشباع سیال روی سرعت‌های موج P و S و دامنه مورد مطالعه قرار گرفت.

سرعت موج تراکمی V_p و سرعت موج برشی V_s در برابر تخلخل ϕ برای همه نمونه‌ها در شکل ۳-۴ الف و شکل ۳-۴ ب نشان داده شده است. با وجود پراکندگی مهم داده‌ها، روندها نشان می‌دهد که هر دو سرعت V_s و V_p با افزایش تخلخل کاهش می‌یابد. برای آزمایش اول، معادله بهبود یافته متوسط زمان $\frac{1}{V} = B_0 + B_1\phi$ بر روی داده‌ها برازش شد (پارامتر ϕ ، تخلخل و B_0 و B_1 ثابت‌های تجربی هستند). با اعمال رگرسیون کمترین مربعات^۳، نتایج برازش شده (

¹ Loading and unloading

² Acoustic and petrographic

³ Least-squares regression

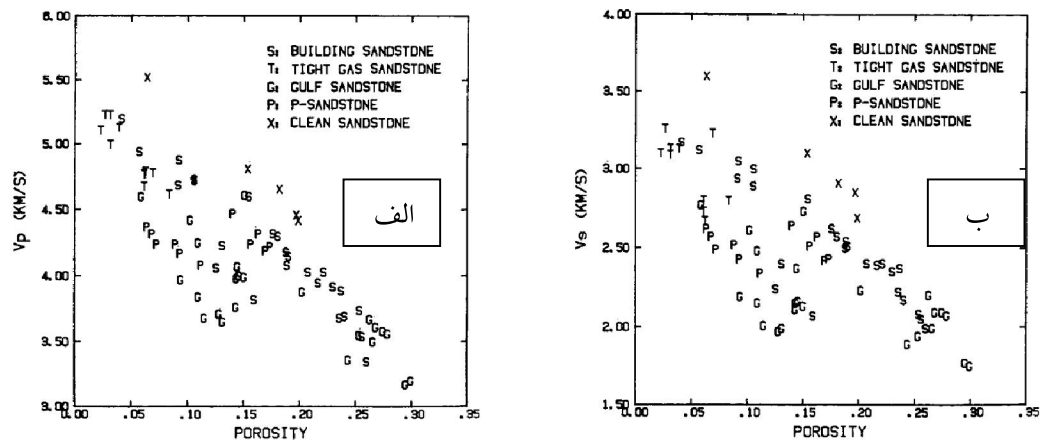
جدول ۴-۱) به صورت انحراف نسبی^۱ در برابر تخلخل در شکل ۴-۴ الف نشان داده شده است.

$$V_m = \frac{1}{B_0} = 5.15 \text{ km/s}$$

می باشد (B₀ ثابت تجربی است)، که خیلی کمتر از مقدار سرعت تراکمی کوارتز غیر متخلخل می باشد،

به هر حال این انحرافها به طور واضح وابسته به محتوای رس می باشند (شکل ۴-۴ ب) و نشان

می دهند که محتوای رس به طور سیستماتیک بر روی سرعت های V_p و V_s اثر می گذارد [۳۴].



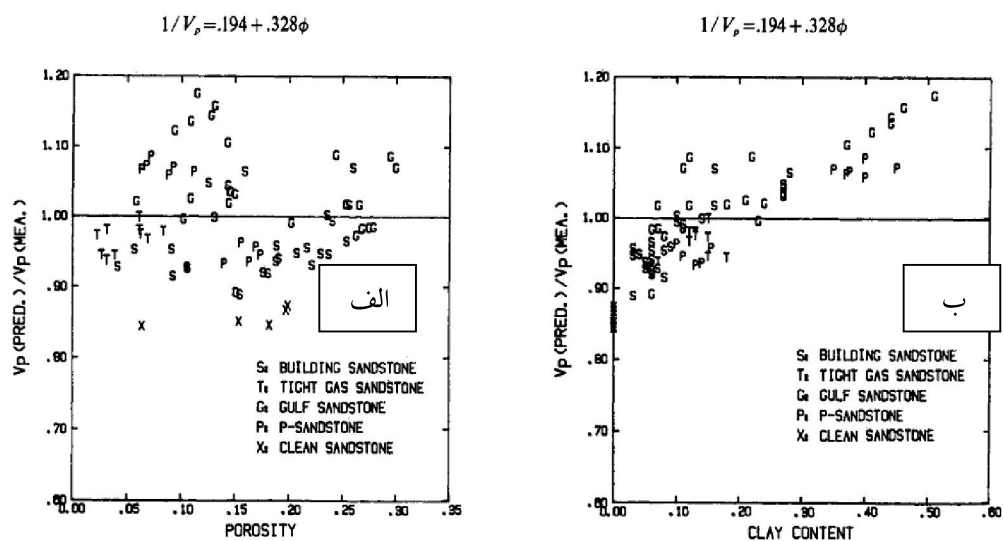
شکل ۴-۳ الف) موج تراکمی اندازه گیری شده و ب) سرعت موج برشی اندازه گیری شده در برابر تخلخل برای ۷۵

نمونه ماسه سنگ در فشار ۴۰ مگا پاسکال و ۱ مگا پاسکال [۳۴].

¹ Relative deviations

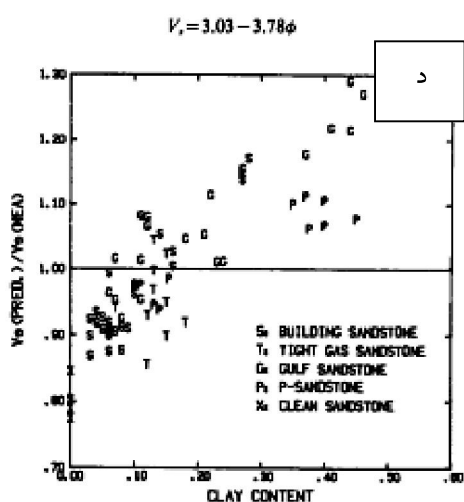
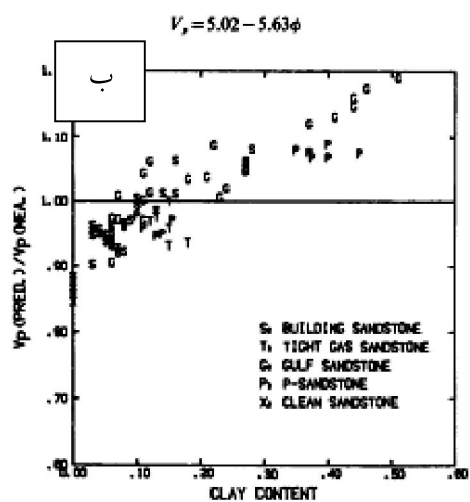
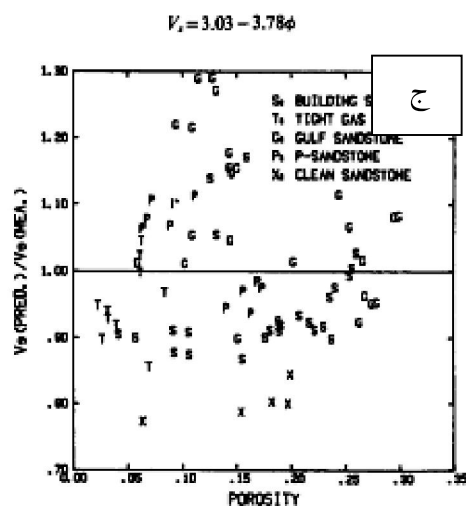
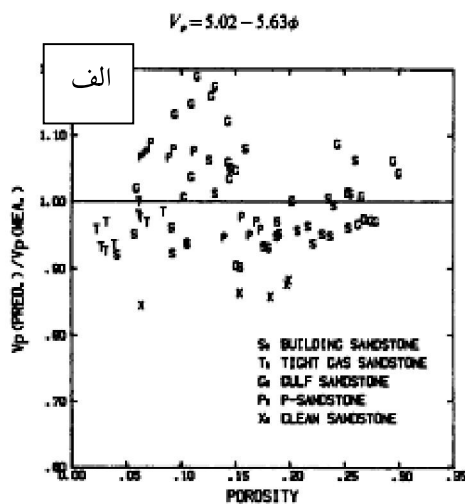
جدول ۴-۱- مدل‌های برازش شده در داده‌های سرعت تجربی [۳۴]

معادلات بدون محتوای رس		
$\frac{1}{V_p} = 0.194 + 0.328\phi$	R: 0.844	rms: 6.6%
$\frac{1}{V_s} = 0.322 + 0.628\phi$	R: 0.750	rms: 10.3%
$V_p = 5.02 - 5.63\phi$	R: 0.840	rms: 6.8%
$V_s = 3.03 - 3.78\phi$	R: 0.754	rms: 10.8%
معادلات با محتوای رس		
۱. معادلات خطی		
$V_p = 5.59 - 6.93\phi - 2.18C$	R: 0.985	rms: 2.1%
$V_s = 3.52 - 4.01\phi - 1.89C$	R: 0.959	rms: 4.3%
۲. معادلات میانگین زمان سیر		
$\frac{1}{V_p} = 0.163 + 0.399\phi + 0.119C$	R: 0.972	rms: 2.8%
$\frac{1}{V_s} = 0.242 + 0.812\phi + 0.307C$	R: 0.945	rms: 5.1%
R: درجه همبستگی		
rms: انحراف معیار نسبی در سطح اعتماد ۶۸.۳ درصد		



شکل ۴-۴- سرعت موج تراکمی V_p در فشار ۴۰ و ۱ مگا پاسکال که از معادله زمان- میانگین به دست آمده است. (الف) انحراف نسبی در برابر تخلخل که با پراکندگی بالا نشان داده شده است. (ب) انحراف نسبی یک همبستگی واضح را با محتوای رس نشان می‌دهد [۳۴].

همچنین از مدل تجربی خطی $V = A_0 - A_1\phi$ و A_1 و A_0 ثابت‌های تجربی هستند) برای برآزش کردن داده‌های V_p و V_s استفاده شد. انحرافات نسبی داده‌ها از مقدار پیش‌بینی‌شده توسط معادله، بسیار زیاد (شکل ۵-۴ الف و ج) و به وضوح وابسته به محتوای رس هستند (شکل ۵-۴ ب و د) [۳۴].



شکل ۴-۵- سرعت‌های موج تراکمی V_p و برشی V_s که با معادله $V = A_0 - A_1\phi$ برازش شده است. الف) انحراف نسبی V_p در برابر تخلخل، با پراکندگی بالا. ب) انحراف نسبی V_p در برابر محتوای رس. ج) انحراف نسبی V_s در برابر تخلخل، با پراکندگی بالا. د) انحراف نسبی V_s در برابر محتوای رس [۳۴].

با توجه به شکل ۴-۴ الف، شکل ۴-۵ الف و شکل ۴-۵ ج، سرعت‌های ماسه سنگ‌های تمیز به طور سیستماتیک از مقدار پیش‌بینی‌شده‌ی حاصل از برازش بالاتر است. به خاطر اهمیت اثر محتوای رس، داده‌های به کار رفته در برازش کردن مدل، ماسه سنگ‌های فاقد رس مستثنی شدند.

بر اساس نتایج بالا، نتیجه‌گیری می‌شود که هر مدل استفاده شده برای برازش کردن داده‌های

V_s و V_p در ماسه سنگ‌های شیلی باید محتوای رس آن‌ها محاسبه شود. دو معادله ساده که شامل

محتوای رس هستند و با رگرسیون کمترین مربعات توصیف می‌شوند به صورت معادله (۱۰-۴) بیان می‌شوند [۳۴]:

$$\begin{aligned} V &= A_0 - A_1\phi - A_2C, \\ 1/V &= B_0 - B_1\phi - B_2C \end{aligned} \quad (10-4)$$

معادله (۱۰-۴) یک معادله میانگین زمان^۱ می‌باشد که با یک عبارت خطی از محتوای رس بهبود یافته است. ضرایب B_0 ، B_1 و B_2 به شرح معادله (۱۱-۴) تعریف می‌شوند [۳۴]:

$$\begin{aligned} B_0 &= 1/V_m; \\ B_1 &= (1/V_f - 1/V_m); \\ B_2 &= (1/V_c - 1/V_m), \end{aligned} \quad (11-4)$$

که V_m سرعت ماتریکس، V_f سرعت سیال و V_c سرعت رس است.

نتایج برازش کردن این معادلات در شکل ۱-۴ آمده است. هر دو معادله، ضرایب همبستگی مشابه و انحرافات نسبی rms داده‌ها را نشان می‌دهند. به هر حال، معادله میانگین زمان برای موج برشی بی معنی است. علاوه بر این، سرعت‌های محاسبه شده از این معادله به طور سیستماتیک بالاتر از سرعت اندازه‌گیری شده در تغییرات پایین تخلخل ($\phi < 7\%$) و تغییرات بالای تخلخل ($\phi > 25\%$) در مقایسه با تخلخل در تغییرات متوسط می‌باشد. از طرف دیگر، انحرافات داده‌ها از رابطه خطی، به طور تصادفی توزیع شده‌اند. به علاوه معادله (۱۰-۴) داده‌های V_p و V_s را اندکی بهتر برازش می‌کند. بنابراین معادله (۱۰-۴) برای نشان دادن وابستگی سرعت به تخلخل و محتوای رس در ماسه‌سنگ، انتخاب شد. معادلات مشابه با روابط مطرح شده توسط توسا و نور^۲ (۱۹۸۲)، کوالیس^۳ و همکاران (۱۹۸۴) [۸۱]، کوستاگنا^۴ و همکاران (۱۹۸۵) [۸۲] می‌باشند. بهترین برازش برای داده‌ها در فشار همه

¹Time-Average Equation

²Tosaya and Nur

³Kowalis

⁴Costagna

جانبه ۴۰ مگا پاسکال و فشار منفذی ۱ مگا پاسکال با رگرسیون کمترین مربعات به دست می‌آید.

برای سرعت تراکمی (V_p) داریم (معادله (۴-۱۲)) [۳۴]:

$$V_p (km/s) = 5.59 - 6.93\phi - 2.18C, \quad (4-12)$$

و برای سرعت برشی (V_s) داریم (معادله (۴-۱۳)) [۳۴]:

$$V_s (km/s) = 3.52 - 4.91\phi - 1.89C. \quad (4-13)$$

ضریب همبستگی برای V_p برابر ۰/۹۸۵ و انحراف نسبی rms آن برابر ۰/۰۲۱ و برای V_s ضریب همبستگی و انحراف نسبی به ترتیب برابر ۰/۹۵۹ و ۰/۰۴۳ می‌باشند. این مقادیر نشان می‌دهد که برازش کردن به طور قابل ملاحظه‌ای با استفاده از تخلخل تنها، بهبود یافته است [۳۴].

۴-۵ رابطه تجربی نول برای محاسبه مقاومت ویژه

۴-۵-۱ توصیف نمونه‌ها و فرآیند آزمایش

نمونه‌های استفاده شده دارای دو جزء جامد بودند: ماسه سیلیسی^۱ و رس کائولینیتی^۲. ماسه استفاده شده از کوه لین^۳ در واشنگتن برداشت گردید. برای رس از کائولینیت استفاده شد. کائولینیت از ذخیره مواد معدنی در دانشگاه میسوری^۴ گرفته شده است [۳۳].

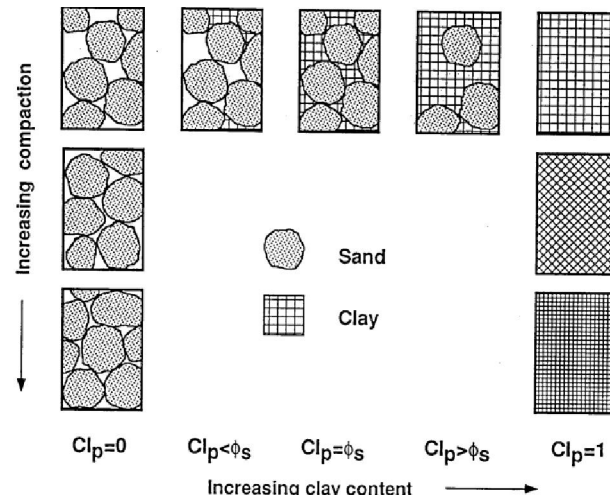
¹ Silica sand

² Kaolinite clay

³ Lane Mountain, Washington

⁴ University of Missouri

شکل ۴-۶ آرایش میکرو هندسی اجزاء در نظر گرفته شده در ترکیب‌های مختلف را نشان می‌دهد. قسمت بالایی شکل ترکیب ماسه-رس و قسمت‌های سمت چپ و راست به ترتیب ماسه خالص و کائولینیت خالص را نشان می‌دهد [۳۳].

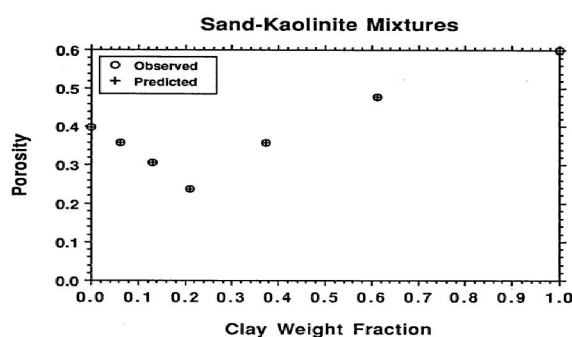


شکل ۴-۶- شکل مفهومی هندسه تراکم برای ماسه خالص، رس خالص و ترکیب رس - ماسه در شرایط پارامترهای تجربی: تخلخل (یا درجه فشردگی) و محتوای رس. پارامتر Cl_p محتوای رس و ϕ_s تخلخل ماسه خالص که در متن آمده است [۳۳].

شکل ۴-۷ روابط پیش‌بینی شده و مشاهده شده بین تخلخل و مقدار وزنی رس (نسبتی از ۱) برای ۷ نمونه ماسه - کائولینیت را نشان می‌دهد. همه مقادیر وزنی رس گزارش شده ترکیب خشک دارند. مقدار تخلخل برای ماسه خالص ۰/۳۹۹ و برای کائولینیت خالص برابر ۰/۵۹۹ می‌باشد. این مقدار تخلخل برای نمونه‌های پر شده در ظرف نمونه ولی فشرده نشده^۱ است (یعنی فشار اندکی بر روی نمونه اعمال می‌شود). دو نمونه ماسه- کائولینیت با مقدار تخلخل حدود ۰/۳۶ در آزمایش خشک استفاده شدند. همان طور که در شکل ۴-۷ نشان داده شده است، این نمونه‌ها در بخش تخلخل کمینه یا نقطه بحرانی که در مقدار وزنی رس در ۰/۲۰۹ اتفاق می‌افتد، قرار می‌گیرند [۳۳].

^۱ Uncompacted

اندازه‌گیری‌های جریان الکتریکی اول تحت شرایط اتاق خشک و سپس اشباع قرار گرفتند. برای اشباع کردن نمونه‌ها، آب مقطر با یک رسانندگی حدود 2×10^{-4} S/m (۱۰ هرتز تا ۱۰ مگاهرتز) از یک سوراخ تزریق شد. با استفاده از این فرآیند، مقدار اشباع شدگی بزرگ‌تر از ۰/۸ برای همه نمونه‌ها به دست آمد. مقادیر اشباع گزارش شده از نظر گراوی متری مشخص و برای آب نگه داشته شده در حفره و آب آغشته شده در سطح مشترک استوانه – الکتروود، تصحیح شدند. عدم قطعیت در مقدار S_w برابر ± 0.05 می‌باشد [۳۳].

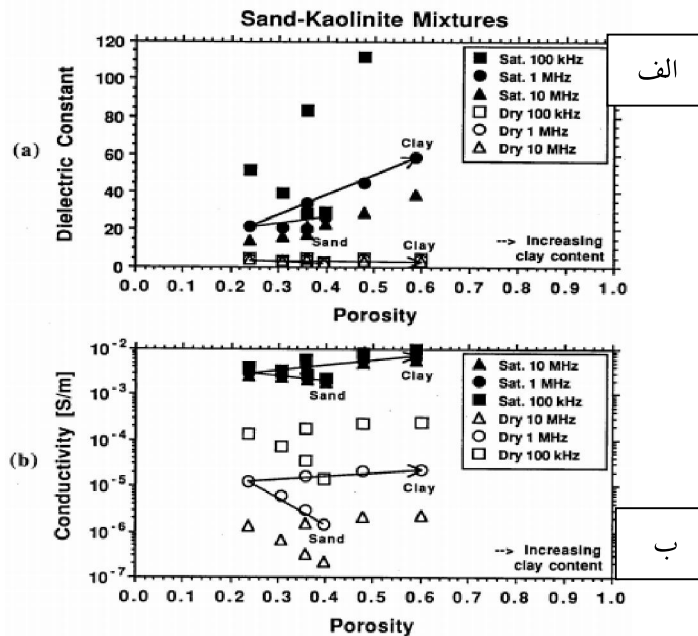


شکل ۴-۷- روابط پیش‌بینی شده و مشاهده شده بین تخلخل و درصد وزنی رس برای ترکیب ماسه – کائولینیت در فشار همه جانبه پایین. مقادیر پیش‌بینی شده بر اساس مدل میکرو هندسی برای ترکیب ماسه – رس توسط ماریون^۱ و همکاران (۱۹۹۲) بیان شده است [۳۳].

۴-۵-۲ نتایج تجربی

خواص الکتریکی بر حسب تخلخل و همچنین رسانندگی بر حسب تخلخل برای نمونه ماسه- کائولینیت در شکل ۴-۸ نشان داده شده است. تعداد ۷ نمونه در شرایط اتاق خشک و ۷ نمونه در شرایط اشباع اندازه‌گیری شدند [۳۳]. با محاسبه رابطه بین تخلخل و رسانندگی، رابطه بین تخلخل و مقاومت ویژه نیز بدست می‌آید (مقاومت ویژه برابر عکس رسانندگی است).

¹ Marion



شکل ۴-۸- الف) ثابت دی الکتریک بر حسب تخلخل، و ب) رسانندگی بر حسب تخلخل، برای نمونه‌های ماسه کائولینیت در سطوح مختلف از اشباع شدگی. علامت‌های تو خالی خواص الکتریکی اندازه‌گیری شده در شرایط اتاق خشک، علامت‌های تو پر خواص الکتریکی اندازه‌گیری شده بعد از این که نمونه‌ها با آب مقطر اشباع شدند را نشان می‌دهند ($\sigma_w \approx 4 \times 10^{-4} \text{ S/m}$). فرکانس‌های اندازه‌گیری شده ۱۰۰ کیلوهرتز، ۱ مگاهرتز و ۱۰ مگاهرتز هستند. فلش‌ها جهت افزایش محتوای رس را نشان می‌دهند [۳۳].

۴-۶ معادلات روش میانگین متحرک تبدیل فوریه سریع

روش میانگین متحرک (MA) برای شبیه‌سازی حوزه آماری گوسین یک‌بعدی توسط جورنل در سال ۱۹۷۴ ارائه شد [۸۳]. سپس در سال ۱۹۹۵ توسط الیور برای حالت‌های دو و سه بعدی به کار گرفته شد [۸۴]. این روش خیلی شبیه تجزیه چولسکی است و تنها تفاوت آن در روش تجزیه است. الگوریتم میانگین متحرک در تولید حوزه تصادفی گوسین با تابع کوواریانس ثابت، بسیار موثر و کارآمد است [۸۵]. روش میانگین متحرک تبدیل فوریه سریع^۱ یا FFTMA یک روش شبیه‌سازی طیفی در حوزه فرکانس بر اساس زمین‌آمار می‌باشد که می‌تواند ساختار فضایی و عناصر تصادفی را در حین شبیه‌سازی از هم جدا کند.

^۱ Fast Fourier Transform Moving Average (FFTMA)

نظریه بوچنر (۱۹۳۶) [۸۶] نشان می‌دهد که هر فرایند ثابت یک تابع کوواریانس C به فرم زیر

دارد:

$$C(x) = \int_{-\infty}^{+\infty} S(f) \exp(2i\pi f x) df \quad (۱۴-۴)$$

پارامتر S طیف توان یا تابع چگالی طیفی است و وابسته به فرکانس f است. انتگرال یک اختصار برای انتگرال n -fold است که n ، بعد فضا می‌باشد. کوواریانس و تابع طیف توان اطلاعات مشابهی را نشان می‌دهند، اما فضای متفاوتی دارند. تابع C معکوس تبدیل فوری S می‌باشد. تابع S به صورت زیر تعریف می‌شود:

$$S(x) = \int_{-\infty}^{+\infty} C(f) \exp(-2i\pi f x) dx \quad (۱۵-۴)$$

تبدیل فوری n بعدی از یک همبستگی n بعدی، تولید تبدیل فوری n بعدی تابع همبستگی شده می‌شود که به صورت زیر است:

$$F(g_1 * g_2) = F(g_1) \cdot F(g_2) \quad (۱۶-۴)$$

تابع کوواریانس و طیف توان، حقیقی و زوج هستند. در مجموع یک خاصیت مهم از فرآیندهای ثابت، تعامد مؤلفه‌های طیفی است [۸۷].

یک میانگین صفر حوزه تصادفی گوسین $z(x)$ با کوواریانس $C(h)$ یک فاصله برداری بین x_i و x_j است) به صورت زیر تولید می‌شود:

$$z(x) = g(h) * y(x) \quad (۱۷-۴)$$

که $*$ عملگر همبستگی است، $y(x)$ نویز سفید گوسین و $g(h)$ جذر همبستگی کوواریانس می‌باشد. یعنی $C(h) = g(h) * \hat{g}(h)$ که $\hat{g}(h) = g(-h)$ است. برای کوواریانس متقارن و کوواریانس متقابل انتخاب $\hat{g}(h) = g(-h)$ ممکن است، بنابراین جذر همبستگی متقارن وجود دارد (اگرچه جذر نامتقارن هم وجود دارد). در سال ۱۹۹۵، اولیور توابع کوواریانس و جذر اتو کوواریانس در دو و سه بعد را معرفی کرد [۸۴]. در سال ۲۰۰۰ لی راوالک و همکاران نشان دادند که همبستگی فضایی معادله

(۴-۱۴) محاسبات مؤثرتری در حوزه طیفی با استفاده از FFT دارد، از این رو روش FFTMA توسعه پیدا کرد [۸۵]. در سال ۲۰۰۳ اولیور استفاده از تجزیه چولسکی در شبیه‌سازی دو بعدی مدل‌های N-LMC در ترکیب با انواع مدل‌های متفاوت برای کوواریانس‌های مستقیم یا تغییرات متفاوت را مطرح کرد [۸۸].

تئوری هم‌میخت در مدل‌های پیوسته و یا گسسته نشان می‌دهد که تبدیل فوری هم‌میخت دو تابع، ضرب تبدیل فوریه دو تابع است.

$$F(g_1 * g_2) = F(g_1)F(g_2) \quad (۴-۱۸)$$

در مورد تک متغیره با کوواریانس متقارن $C=g*g$ ، که g ریشه دوم متقارن است. تابع چگالی طیفی $S = F(C) = F(g * g) = [F(g)]^2$ است و $F(g) = \sqrt{S}$ می‌باشد. بنابراین برای بدست آوردن جذر طیف در حالت تک متغیره ریشه دوم تابع چگالی طیفی کوواریانس را می‌گیرند. با داشتن جذر طیف و ضرب آن با طیف یک نویز سفید گوسین و سپس معکوس تبدیل فوریه، دامنه یا حوزه فضایی را محاسبه می‌کنند. نتیجه نهایی، کانولوشن ریشه کوواریانس با یک نویز سفید گوسین (یک میانگین متحرک) برای تکثیر صحیح کوواریانس مورد نظر است.

۴-۷ جمع بندی و نتیجه‌گیری

در این فصل فرمولاسیون معادلات خوشه‌بندی طیفی و روش تعیین تعداد خوشه بهینه و همچنین روابط تجربی بکار رفته برای محاسبات سرعت موج تراکمی لرزه‌ای و مقاومت‌ویژه الکتریکی، مطرح و مورد بررسی قرار گرفته است. برای تعیین تعداد خوشه بهینه از روش شاخص دان استفاده شده است. روابط تجربی به کار رفته در محاسبه سرعت موج تراکمی از معادله هان و همکاران (۱۹۸۶) و در محاسبه مقاومت‌ویژه الکتریکی از معادله نول و همکاران (۱۹۹۶)، بوده است.

فصل پنجم

وارون سازی داده های مصنوعی و

صحرائی داده های لرزه ای و مقاومت ویژه

۵-۱ مقدمه

داده‌های صحرایی حاصل از برداشت‌های توموگرافی لرزه انکساری و مقاومت‌ویژه بر روی یک سد در حال احداث، با هماهنگی شرکت توسعه منابع آب و نیروی ایران از شرکت سایبر تهیه شد. تعهد شد که در این رساله نام پروژه و نام منطقه مورد مطالعه که در آن یکی از پروژه‌های سدی ایران در حال اجراست، ذکر نشود.

گستره مورد مطالعه از دیدگاه تقسیم‌بندی‌های موجود زمین شناسی، در محدوده زون زاگرس رورانده یا زاگرس خرد شده قرار می‌گیرد. این گستره از شمال خاور توسط گسل معکوس اصلی زرآب و از جنوب باختر توسط ارتفاعات بختیاری و گسل زردکوه محدود می‌گردد. زاگرس مرتفع یا رورانده نواری باریک و کم عرض (۱۰ تا ۷۰ کیلومتر) می‌باشد که چون مرتفع‌ترین کوه‌های زاگرس در این بخش قرار دارند، تحت عنوان زاگرس مرتفع خوانده شده و از طرفی به دلیل خردشدگی و روراندگی‌های شدید موجود در آن تحت عنوان زاگرس رورانده نیز نام برده می‌شود. زاگرس مرتفع عمیق‌ترین قسمت فرورفتگی (دریا) زاگرس را طی مزوزوئیک و اوایل ترشیری تشکیل می‌داده است. در این فرورفتگی ضخامت لایه‌های دریایی تریاس تا ائوسن به ۵۳۰۰ متر می‌رسد که شامل مارن‌های گلوبوژی‌رین‌دار و رادیولاریت‌دار است. وجود سنگ‌های تخریبی (فلیش) در مجموعه رسوبات پلیوسن آن نشان از حرکات مهم در کرتاسه بالا- پلیوسن است که در محدوده زاگرس رورانده رخ داده و سبب بالا آمدن زمین، چین خوردگی و تخریب شده است [۸۹].

در گستره مورد مطالعه قسمت اعظم بارش به صورت برف است. وجود کوه‌های با رخنمون لایه‌های آهک و با ارتفاع بیش از ۲۵۰۰ متر، بارش برف و تکتونیک فعال در آن شرایطی را فراهم نموده که امکان تشکیل آبخوان‌های بزرگ را در آن فراهم آید. تغذیه این آبخوان‌ها توسط آب حاصل از ذوب برف و نفوذ از طریق شکستگی‌ها و پدیده‌های انحلالی صورت می‌پذیرد. آب آبخوان‌های فوق

توسط چشمه‌های فصلی متعدد در دامنه کوه‌های گستره طرح و چشمه‌های دائمی تخلیه می‌گردد [۸۹].

نهشته‌های رخنمون شده در گستره مورد مطالعه تماماً شامل سنگ‌ها و لایه‌های رسوبی هستند. مجموعه رسوبات برجا در این گستره شامل لایه‌های با سن کرتاسه بالایی، ترشیری و رسوبات دریاچه‌ای عهد حاضر و آبرفت‌های جوان است (Q^s). مشخصات کلی مجموعه لایه‌های رخنمون شده در محدوده به ترتیب از قدیم به جدید به عبارتند از:

الف - آهک‌های کرتاسه (K_2^{ml})

این آهک‌ها به صورت توده‌ای و به رنگ خاکستری تیره در بخش جنوب باختری کوه زرآب در محدوده ساختگاه سد رخنمون دارند. این آهک‌ها با سن کرتاسه در ستون چینه شناسی زاگرس معادل سازند سروک هستند. این آهک در محدوده ساختگاه از لایه بندی نازک تا ضخیم لایه در تغییر می‌باشند. که در بخش‌هایی به آهک‌های مارنی کاملاً نازک لایه تغییر می‌نمایند [۸۹].

ب - آهک‌های ائوسن (K_2^{II})

سنگ‌های آهکی خاکستری روشن با سن ائوسن در محدوده مخزن سد، سه رخنمون دارند که از مشخصات بارز آن‌ها در بیشتر منطقه وجود میکروفسیل‌های نومولیتی است. با توجه به وجود این فسیل‌ها سن سنگ‌های فوق ائوسن و معادل سازند شهبازان در نظر گرفته شده‌اند [۸۹].

ج - مارن سنگ و آهک‌های مارنی کرتاسه فوقانی (K_2^{ml})

رخنمون این سنگ‌ها در ابتدای مخزن سد توسط نهشته‌های واریزه‌ای عهد حاضر پوشیده شده است و بر مبنای مرحله دوم مطالعات ژئوتکنیک مشخص شده است که مرز این واحد با آهک‌های فوقانی گسلی می‌باشد. گسل مزبور حذف بخش عمده‌ای از واحد فوق را سبب شده است [۸۹].

د - مارن‌های میوسن (K_2^{I2})

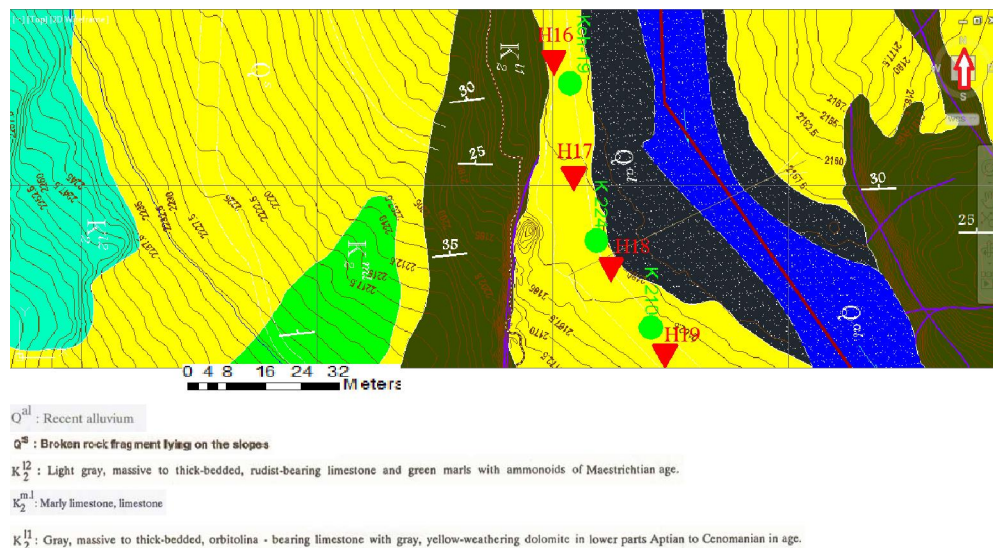
لایه‌های مارن، سیلتستون و شیل آهکی با سن میوسن در گستره مخزن سد رخنمون دارند.

ه - نهشته‌های دریاچه‌ای کواترنری (Q^{al})

این رسوبات شامل تناوب کنگلومرا، سیلتستون، گل سنگ، مارن‌های خاکستری و سبز است که در گستره ساختگاه و مخزن سد رخنمون دارند.

و - نهشته‌های آبرفتی و واریزه‌ای کواترنری (Q^{al})

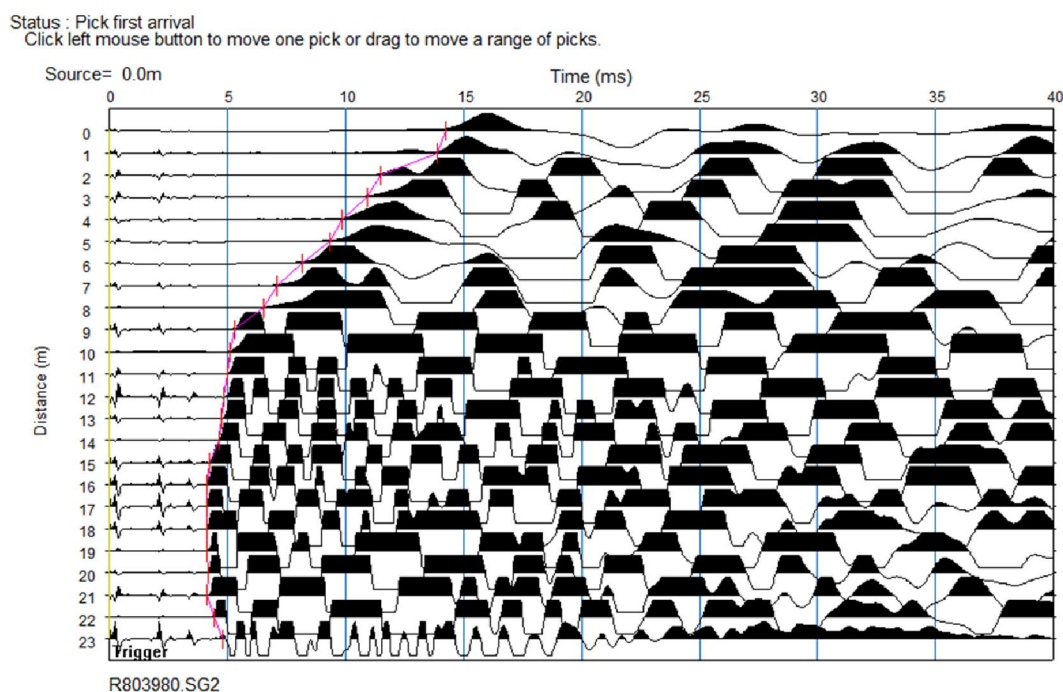
این نهشته‌ها حاصل رسوب‌گذاری در گستره رودخانه، پای دامنه‌ها و مخروط افکنه‌هاست. لازم به ذکر است که در تکیه گاه چپ مجموعه‌ای از قطعات خرد شده سنگ آهک که توسط سیمان آهکی یا رسی در کنار یکدیگر قرار گرفته‌اند در حجم بسیار زیاد قابل مشاهده است. این حجم از قطعات نابرجا و زاویه‌دار که اندازه آن‌ها از حد قلوه تا بلوک‌های چندمتری متغیر است، در اثر فعالیت گسله موجود در منطقه از ارتفاعات مشرف به محدوده تکیه گاه چپ جدا شده و در محل فعلی خود واقع و مجدداً سیمانی شده‌اند. با توجه به خطرات احتمالی ناشی از این توده برای بدنه سد، پایدار سازی و شناخت خصوصیات آن از اهمیت بالایی برخوردار است (شکل ۵-۱) [۸۹]. موقعیت گمانه‌های توموگرافی لرزه‌ای انکساری و سونداژهای برداشت مقاومت‌ویژه نیز در شکل ۵-۱ آمده است.



شکل ۵-۱- نقشه زمین‌شناسی و موقعیت گمانه‌های توموگرافی لرزه‌ای انکساری و سونداژهای برداشت مقاومت‌ویژه در محدوده مورد مطالعه [۸۹]

۲-۵ داده‌های توموگرافی لرزه انکساری

داده‌های توموگرافی لرزه انکساری در قالب فرمت ABM بودند که در ابتدا داده‌های خام به فرمت ورودی نرم‌افزار Pickwin (متعلق به شرکت OYO CORPORATION) یعنی SEG2 در آمده و مورد تحلیل قرار گرفتند. برای به دست آوردن دقیق اولین زمان رسید از نرم‌افزار IXSeg2Segy (متعلق به Golden Colorado USA) نیز استفاده گردید. بعد از محاسبه اولین زمان رسید برای تک تک رکوردها (شکل ۲-۵) و آرون‌سازی داده‌ها با استفاده از نرم‌افزار GeoTomCG (متعلق به آقای دکتر داریل تویتون^۱) انجام شد.



شکل ۲-۵- تعیین اولین زمان رسید رکوردهای توموگرافی لرزه‌ای انکساری

در شکل ۳-۵ نتایج و آرون‌سازی توموگرافی لرزه انکساری بین دو گمانه k224 و kch-19 آمده است. همان‌طور که مشاهده می‌شود، روش توموگرافی لرزه‌ای، به علت چگالی بالای پرتوها، قدرت تفکیک و

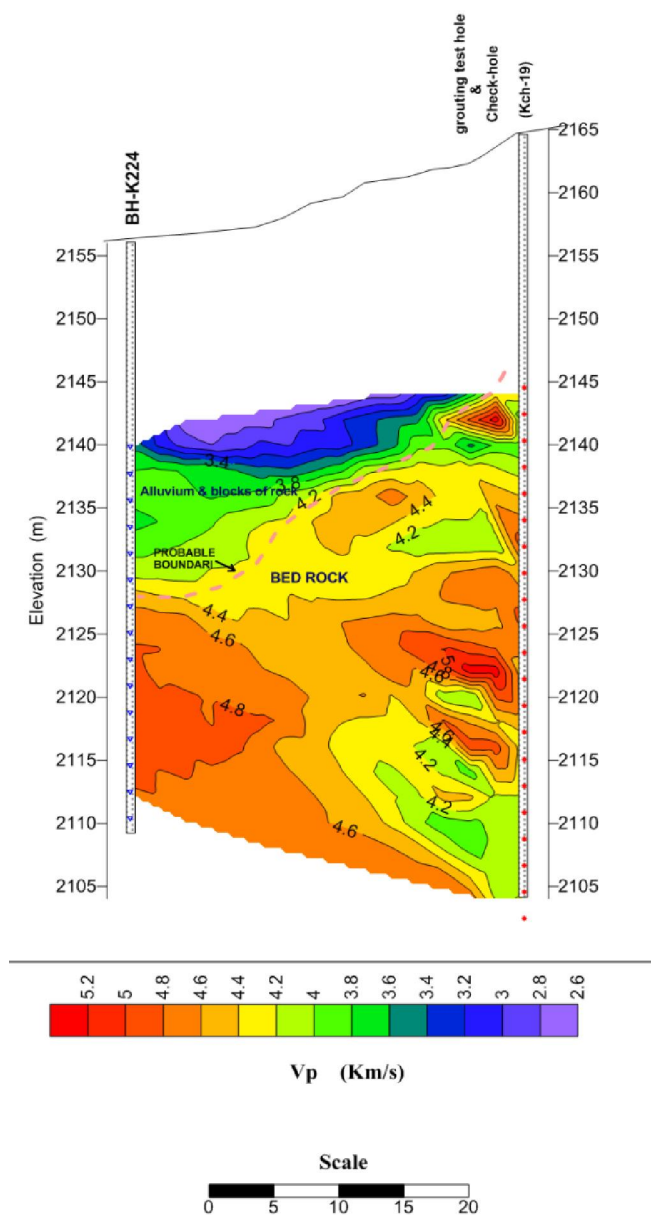
^۱ Daryl Tweeton

دقت بالایی دارد. با محاسبه وارون سازی داده ها و رسم منحنی های هم سرعت که نشان دهنده نحوه توزیع سرعت امواج در حد فاصل فرستنده ها و گیرنده هستند، کیفیت سنگ های آبرفتی و سنگ بستر مورد ارزیابی قرار گرفت. در مناطق سطحی یا آبرفت، سرعت لرزه ای پایین و در مناطق عمیق تر یا سنگ بستر، سرعت لرزه ای بالاتر می باشد.

در این مقطع (شکل ۵-۳)، گیرنده های لرزه ای در گمانه K-224 و چشمه های انفجاری برای تولید موج در گمانه Kch-19 قرار دارد. گمانه K-224 در حاشیه رودخانه در ساحل راست و به عنوان پیژومتر می باشد، گمانه Kch-19 نیز در ساحل راست در حاشیه جاده دسترسی به عنوان گمانه تزریق آزمایشی مورد استفاده بوده است که می بایست تأثیر احتمالی عملیات تزریق را بر محیط اطراف تا شعاع مؤثری در اطراف گمانه مزبور و در نتایج حاصله در نظر گرفت. باید اشاره نمود این گمانه ها به صورت خاص برای آزمایش ژئوفیزیک، حفاری و لوله گذاری نگردیده اند. با این توضیح که در گمانه Kch-19 با عمق حدود ۶۰ متر با اجرای هر چشمه انفجاری به دلیل لوله گذاری PVC نامناسب برای این شرایط چند متر از طول گمانه از دست می رفته است. سطح آب زیر زمینی در این گمانه در زمان آزمایش حدود ۱۴ متر بوده است.

مقطع توموگرافی حاصل، در زیر تراز آب زیرزمینی، از تراز حدود ۲۱۴۰ به سمت انتهای گمانه ها و اعماق بیشتر پوشش داده شده است. همان طور که مشاهده می شود یک زون آبرفتی با سرعت سیر موج تا حدود ۳۰۰۰ متر بر ثانیه از سطح تا تراز ۲۱۲۸ گمانه K-224 وجود دارد. پرسرعت ترین قسمت های توده سنگ با سرعت های عمده بین ۴۰۰۰ تا بیش از ۴۵۰۰ متر بر ثانیه نیز از حدود تراز ۲۱۴۰ در گمانه Kch-19 به حالت شیب دار تا تراز حدود ۲۱۲۸ در گمانه K-224 ادامه دارد. در این محدوده بر سرعت بر حسب مورد افزایش و کاهش های نسبی نیز مشاهده می شود. در حریم گمانه K-224 از تراز حدود ۲۱۲۸ به سمت بالا یک زون سرعتی مابین ۳۵۰۰ تا ۴۰۰۰ متر بر ثانیه مشاهده می شود که نسبت به نواحی عمیق آنومالی کم سرعت ولی نسبت به آبرفت با دانه بندی معمولی پرسرعت تر محسوب می شود که می تواند محدوده استقرار تخته سنگ های بزرگ مجزا با

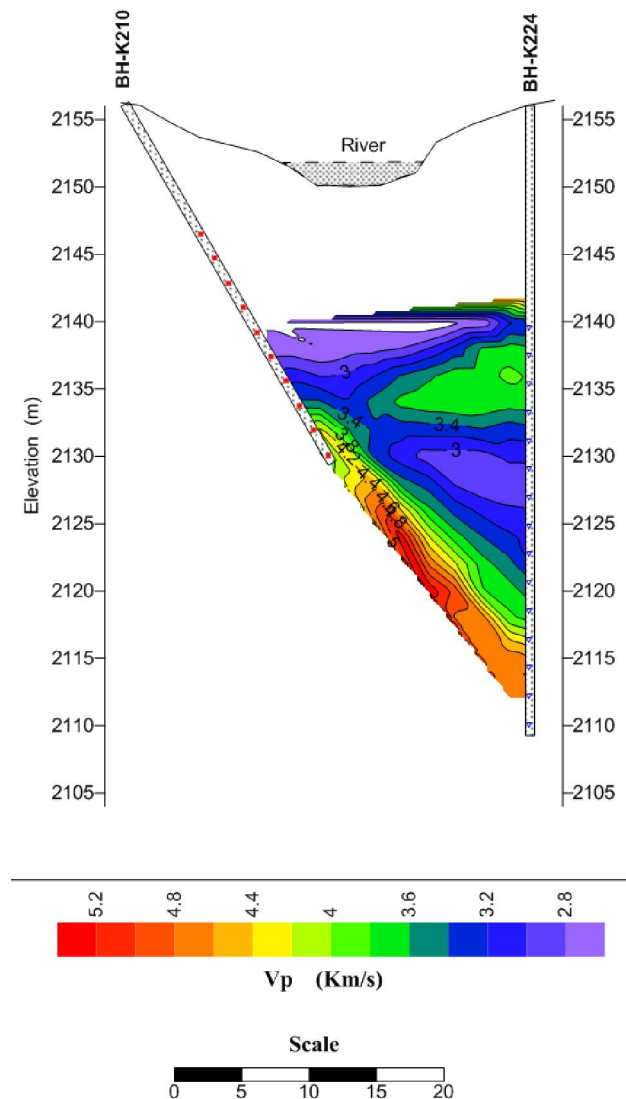
پرشدگی کامل فضاهای خالی بین آن‌ها توسط رسوبات آبرفتی در شرایط اشباع و یا تا حدودی متأثر از عملیات تزریق آزمایشی در گمانه kch19 نیز باشد. به طور کلی این مقطع وجود توده سنگ با کیفیت خوب و به حالت شیب‌دار و از سمت گمانه Kch-19 در تراز حدود ۲۱۴۰ متر به سوی اعماق گمانه K-224 تا حدود تراز ۲۱۲۸ متر را نشان می‌دهد. اگر چه بعضی زون‌های با سرعت نسبی کمتر نیز در این پهنه مشاهده می‌شود.



شکل ۵-۳- نتایج وارون‌سازی توموگرافی لرزه انکساری بین دو گمانه k224 و kch-19

در شکل ۴-۵ نتایج وارون‌سازی توموگرافی لرزه انکساری بین دو گمانه k224 و k210 آمده است. در این مقطع نیز گیرنده همچنان مانند مقطع قبل در گمانه K-224 و چشمه‌های انفجاری در گمانه K210 قرار داشته‌اند. گمانه اخیر در ساحل چپ رودخانه قرار داشته و با زاویه ۳۰ درجه به سمت ساحل راست و به سوی گمانه K-224 حفاری گردیده است این گمانه با لوله فلزی لوله گذاری شده و اگرچه ظاهراً تا عمق حدود ۶۰-۵۷ متر ادامه داشته ولی در زمان آزمایش با خارج کردن دهانه آن از زیر واریزه‌ها و بازگشایی آن، تنها تا اعماق حدود ۳۰ متر باز و به سمت اعماق بیشتر مسدود بوده است.

مقطع حاصله با توجه به شرایط و موقعیت گیرنده‌ها و چشمه‌های لرزه‌ای در شکل ۴-۵ نشان داده شده است. همان‌طور که مقادیر سرعت موج و نحوه پراکندگی پهنه‌ها نشان می‌دهد، یک توده مخلوط ظاهراً آبرفتی- ریزشی که شامل تخته سنگ‌های بزرگ تا فضاهای پر شده آبرفتی می‌باشد در محدوده مقطع که پهنه زیر بستر رودخانه را شامل می‌شود استقرار یافته است. یک توده پرسرعت نیز به حالت شیب‌دار در اعماق مقطع مشاهده می‌شود. این زون سرعتی با توجه به هندسه آزمایش ممکن است از دقت مطلوبی برخوردار نباشد ولی به هر حال می‌تواند نشانه وجود توده سنگ یا تخته سنگ‌های منفصل ولی بزرگ در اعماق باشد.



شکل ۴-۵- مسیر پرتوهای عبوری و نتایج وارون سازی توموگرافی لرزه انکساری بین دو گمانه k210 و k224

۳-۵ داده‌های مقاومت ویژه الکتریکی

داده‌های مقاومت ویژه الکتریکی که با آرایه شلومبرژه با طول پروفیل ۳۰۰ متر، برداشت شده بودند؛ بعد از به فرمت درآمدن، وارد نرم افزار RES2DINV شده و وارون سازی داده‌ها با استفاده از این نرم افزار انجام شد. با انجام وارون سازی بر روی داده‌های برداشت شده در طول یک پروفیل، خطای RMS وارون سازی بعد از ۱۰ بار تکرار برابر ۴/۷ درصد بدست آمد. وجود آبرفت‌های خشک و دارای

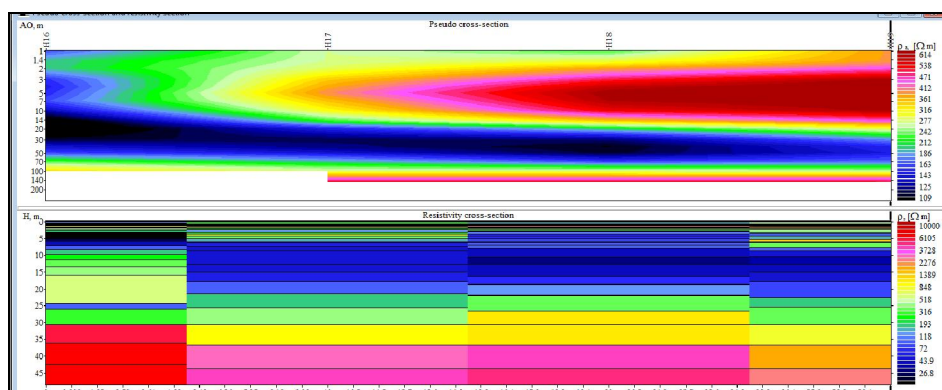
شن و ماسه باعث افزایش مقاومت ویژه الکتریکی شده و در زون آبدار، کاهش مقاومت ویژه را شاهد هستیم. در سنگ بستر نیز مقاومت ویژه الکتریکی افزایش پیدا می کند (شکل ۵-۶).

در محدوده ساختگاه در امتداد رودخانه و در ساحل راست یک راه خاکی جهت دسترسی به محدوده وجود دارد که حاشیه مسیر این راه جهت اجرای سونداژهای الکتریکی قائم (VES) از بالادست محور تا پایین دست محور انتخاب گردید.

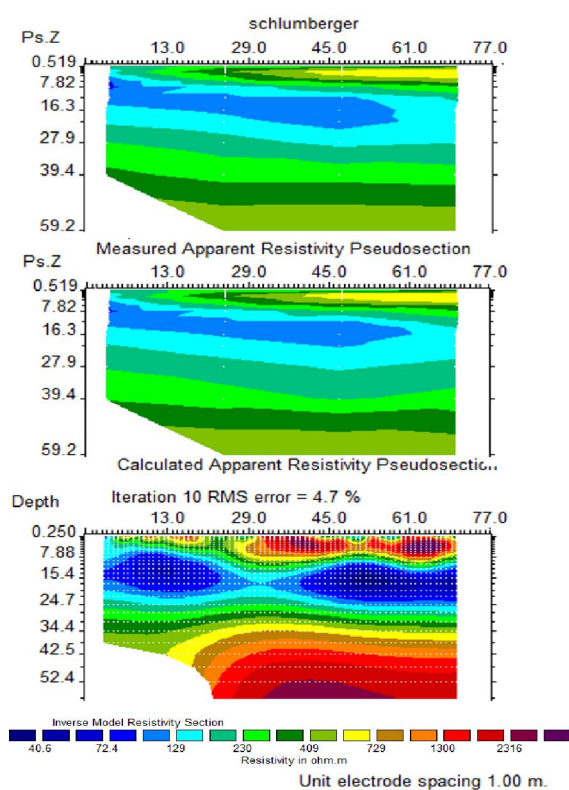
هدف از این برداشت ها بررسی وضعیت روباره آبرفتی و نیز تغییرات ضخامت آن و همچنین بررسی شرایط سنگ بستر در مسیر رودخانه در ساحل راست از بالا دست به سمت پایین دست و در مسیر به طول حدود ۶۵ متر بوده است.

برای انطباق مقطع مقاومت ویژه و توموگرافی لرزه انکساری، تنها سونداژهای H16، H17، H18 و H19 مورد تحلیل و ارزیابی قرار گرفتند (شکل ۵-۵). همان طور که مشاهده می شود تصویر مقطع نشان داده شده در شکل ۵-۶ بیانگر پوشیده شدن سنگ بستر آهکی توسط لایه روباره آبرفتی می باشد؛ که بر حسب موقعیت ضخامت این لایه آبرفتی از حدود ۱۴ متر تا ظاهراً حدود نزدیک به ۳۵ متر تغییر می کند. مقاومت ویژه الکتریکی این روباره های آبرفتی در زون خشک یعنی لایه بالایی بین حدود ۵۴ تا ۴۵۰۰ اهم متر و در زون اشباع و یا همراه با رطوبت بالای ناشی از رودخانه و آب زیر زمینی یعنی در لایه دوم آبرفتی عمدتاً بین حدود ۳۵ تا ۲۳۰ اهم متر متغیر می باشد. سنگ بستر آهکی زیرین نیز دارای سطح یکنواخت ثابتی نبوده و طبیعتاً ناهمواری های توپوگرافی را به اشکال مختلف و بر حسب موقعیت نشان می دهد. از دیدگاه مقاومت ویژه الکتریکی سنگ بستر در این مسیر مقاومت ویژه الکتریکی از حدود ۲۳۰ تا ۲۵۰۰ اهم متر را نشان می دهد. با توجه به اشباع بودن منطقه و واقع شدن در زیر تراز آب زیر زمینی قطعاً باید مناطق با مقاومت ویژه الکتریکی بالا شامل سنگ های آهکی یا دولومیتی با کیفیت و تراکم خوب باشند و گر نه این حدود از مقاومت ویژه الکتریکی قاعدتاً نباید برای سنگ های با درز و شکاف فراوان و یا حفره دار و آن هم در زیر سطح آب زیرزمینی مشاهده گردد. از سوی دیگر محدوده هایی که افت مقاومت ویژه الکتریکی را نشان می دهند را می توان به دو

گونه تفسیر نمود که این افت نسبی یا ناشی از تغییرات ذاتی سنگ شناسی لایه‌هاست (مثلاً تا حدودی ماری شدن آهک‌ها) و یا در صورت وجود شرایط سنگ شناسی یکسان در تمامی مسیر آزمایش، این افت نسبی مقاومت‌ویژه الکتریکی در بعضی قسمت‌ها را می‌توان به بیشتر درز و شکافدار بودن و یا همراه بودن با برخی حفرات انحلالی تفسیر نمود.

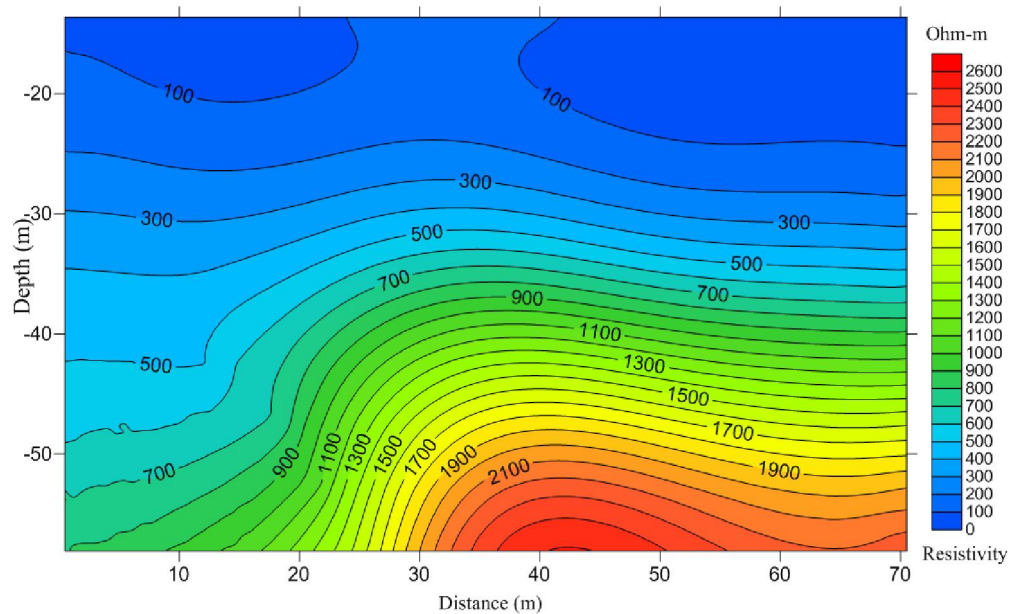


شکل ۵-۵- تحلیل یک بعدی سونداژهای مقاومت ویژه H16، H17، H18، H19



شکل ۵-۶- نتایج حاصل از وارون‌سازی داده‌های مقاومت‌ویژه الکتریکی برداشت شده در طول پروفیل ۲۰۰ متر با استفاده از نرم‌افزار RES2DINV با خطای RMS برابر ۴/۷ درصد

برداشت توموگرافی لرزه انکساری از عمق ۱۴ متر به پایین و زیر سطح ایستایی انجام شد. بنابراین برای انطباق نتایج وارون سازی توموگرافی لرزه انکساری و مقاومت ویژه الکتریکی، مقطع مقاومت ویژه الکتریکی نیز از عمق ۱۴ متری مورد ارزیابی قرار گرفت (شکل ۷-۵).



شکل ۷-۵- مقطع مقاومت ویژه الکتریکی در عمق پایین تر از ۱۴ متر (این عمق در زیر سطح ایستایی قرار دارد)

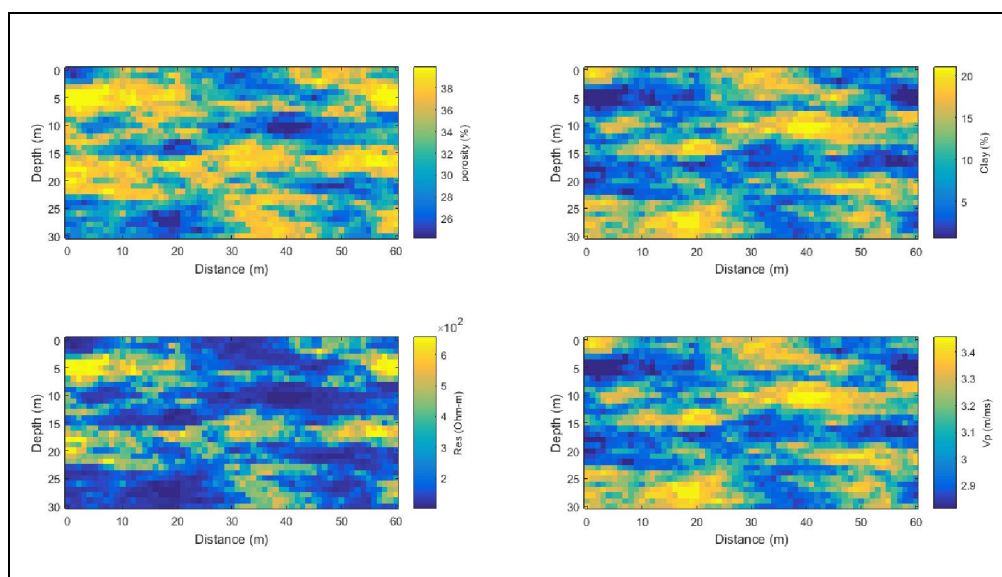
۴-۵ مطالعه مدل مصنوعی

۴-۵-۱ تولید و وارون سازی داده های مصنوعی با استفاده از روابط

تجربی

برای تولید و وارون سازی داده های مصنوعی با استفاده از روابط تجربی از رابطه تجربی هان و همکاران (۱۹۸۶) برای محاسبه سرعت لرزه ای و نیز از رابطه تجربی نول (۱۹۹۶) برای محاسبه مقاومت ویژه الکتریکی استفاده گردید. برای این کار، از روش میانگین متحرک تبدیل فوریه سریع (FFTMA) برای

تولید پردازش‌های آماری گوسین جهت مدل‌سازی سرعت لرزه انکساری و مقاومت ویژه الکتریکی با استفاده از توزیع تخلخل و محتوای رس دوبعدی استفاده شده است. روش FFTMA یک روش شبیه‌سازی طیفی در حوزه فرکانس بر اساس زمین آمار می‌باشد که می‌تواند ساختار فضایی و عناصر تصادفی را در حین شبیه‌سازی از هم جدا کند. این مدل در بسیاری از موارد در توصیف ساختارهای ناهمسان‌گرد و ناهمگن از قبیل آبخوان‌های رسوبی، استفاده شده است. مدل تخلخل و محتوای رس دارای یک تابع چگالی احتمال یا PDF گوسی^۱ است. طول همبستگی افقی و عمودی به ترتیب ۶۰ متر و ۳۰ متر است (شکل ۵-۸).



شکل ۵-۸- محاسبه سرعت لرزه‌ای و مقاومت ویژه الکتریکی با استفاده از تخلخل و محتوای رس (روابط تجربی هان و همکاران (۱۹۸۶) و نول و همکاران (۱۹۹۶))

تخلخل و محتوای رس به دست آمده از روش FFTMA به ترتیب در محدوده عددی بین ۲۳/۹ تا ۳۹/۹ درصد و ۰ تا ۲۰ درصد قرار دارد. با توجه به شکل ۵-۸ مشاهده می‌شود که تخلخل و محتوای رس رابطه عکس با هم دارند. البته با توجه به شکل ۴-۷ تا محتوای رس ۲۳/۹ درصد بین این دو کمیت رابطه عکس و از مقدار محتوای رس ۲۳/۹ درصد به بالا، بین این دو کمیت رابطه مستقیم و

^۱ Gaussian Probability Density Function

لگاریتمی وجود دارد. برای تولید اعداد مصنوعی از بازه رابطه عکس محتوای رس و تخلخل، استفاده شده است (به دلیل اینکه تغییرات تخلخل و محتوای رس در این بازه در نظر گرفته شده است).

۵-۵ جمع بندی و نتیجه گیری

در این فصل تولید داده‌های مصنوعی با استفاده از کدهای محاسباتی در نرم افزار MATLAB و همچنین وارون سازی داده‌های مصنوعی و صحرایی توموگرافی لرزه‌انکساری و مقاومت ویژه با استفاده از نرم افزارهای RES2DINV، Dc2DinvRes و GeoTom CG، مورد بحث قرار گرفته است. به علت پوشش ضعیف سیگنال‌های ژئوفیزیکی در مناطق عمیق تر مدل، قدرت تفکیک و تباین در مناطق عمقی کاهش چشمگیری داشته و بازسازی نامشخص تری را نشان می‌دهد.

فصل هشتم

تلفیق نتایج حاصل از وارون سازی

داده های تو موکرانی لرزه ای و مقاومت ویژه و

اعتبار سنجی آن

۱-۶ مقدمه

وارون سازی داده های مقاومت ویژه الکتریکی با روش کمترین مربعات گوس- نیوتن و با استفاده از نرم افزار RES2DINV صورت گرفته و سپس اولین زمان رسیدن با استفاده از نرم افزار PickWin محاسبه و وارون سازی داده های توموگرافی لرزه ای انکساری با استفاده از نرم افزار Geotom CG انجام شد. بعد از محاسبه ماتریس شباهت، مقادیر ویژه و بردارهای ویژه به دست آمدند و سپس داده ها با روش خوشه بندی طیفی مورد ارزیابی قرار گرفتند. با استفاده شاخص دان برای بهینه سازی تعداد خوشه ها عدد ۱۲ به دست آمد؛ که با توجه به نقشه های به دست آمده (کلاسه بندی ها) برای مقاومت ویژه الکتریکی و توموگرافی لرزه ای انکساری، تعداد خوشه مناسبی است. نتایج الگوریتم TSVM بر اساس خوشه بندی طیفی (TSVMSC)، همبستگی و تطابق خوبی با نتایج حاصل از لاگ گمانه حفاری شده در کف سد را نشان می دهد؛ که در ادامه مورد بررسی قرار می گیرد.

۲-۶ تلفیق داده های توموگرافی لرزه انکساری و مقاومت ویژه

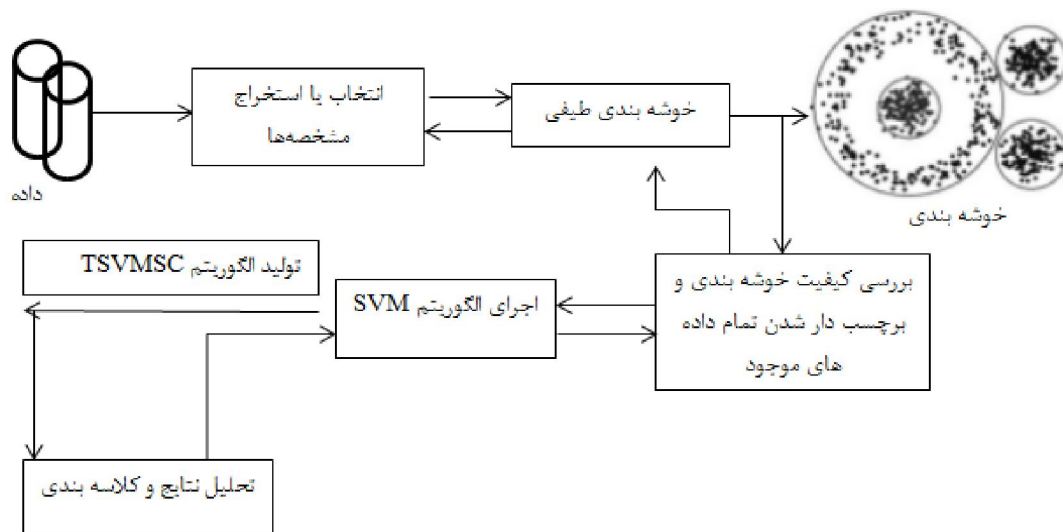
الکتریکی حاصل از روابط تجربی هان (۱۹۸۶) و نول

(۱۹۹۶) و استفاده از الگوریتم TSVMSC در مدل

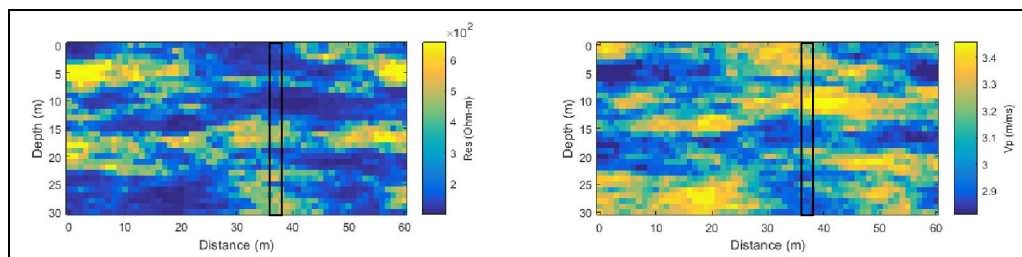
مصنوعی

برای بررسی الگوریتم TSVMSC در یک وضعیت کنترل شده، نتایج حاصل از داده های مصنوعی توموگرافی لرزه انکساری و مقاومت ویژه الکتریکی، تلفیق شده و مورد تحلیل قرار گرفت (فلوچارت این الگوریتم در شکل ۱-۶ آمده است). با توجه به شکل ۲-۶ و شکل ۳-۶، زون هایی با سرعت تراکمی و مقاومت ویژه یکسان در کلاس های مشابه و زون هایی با سرعت تراکمی و مقاومت ویژه الکتریکی

متفاوت در کلاس‌های متمایز قرار گرفته‌اند. برای بهتر نشان دادن این موضوع گمانه‌ای فرضی در موقعیت $x=35$ در نظر گرفته شده است. با استفاده از شاخص دان^۱ تعداد خوشه بهینه برابر با عدد ۱۲ بدست آمده است (محاسبات این شاخص در شکل ۶-۶ آورده شده است). با توجه به این گمانه، تلفیق تغییرات سرعت تراکمی و مقاومت‌ویژه الکتریکی با استفاده از روش TSVMSC با دقت زیاد به دست آمده است. برای نمونه، موقعیت $x=35$ و $y=-3$ در کلاس ۷ و در محدوده سرعتی ۳۳۵۰ متر بر ثانیه و محدوده مقاومت‌ویژه ۱۰۰ اهم متر قرار دارد.

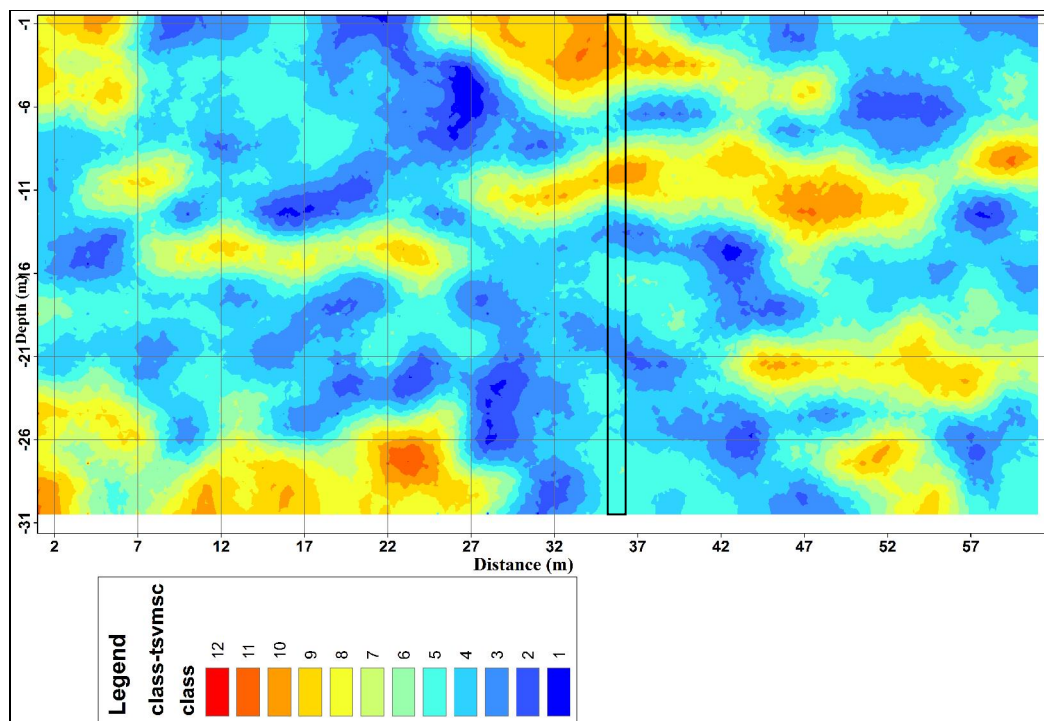


شکل ۶-۱- فلوچارت الگوریتم TSVMSC



شکل ۶-۲- نتایج حاصل از وارون‌سازی داده‌های مصنوعی توموگرافی لرزه‌انکساری (سمت راست) و مقاومت‌ویژه الکتریکی (سمت چپ) (استفاده از روابط تجربی هان (۱۹۸۶) و نول (۱۹۹۶))

^۱ Dunn index



شکل ۳-۶- تلفیق نتایج حاصل از وارون سازی داده های مصنوعی توموگرافی لرزه انکساری و مقاومت ویژه الکتریکی با استفاده از الگوریتم TSVMSC

۳-۶ تحلیل داده های صحرایی

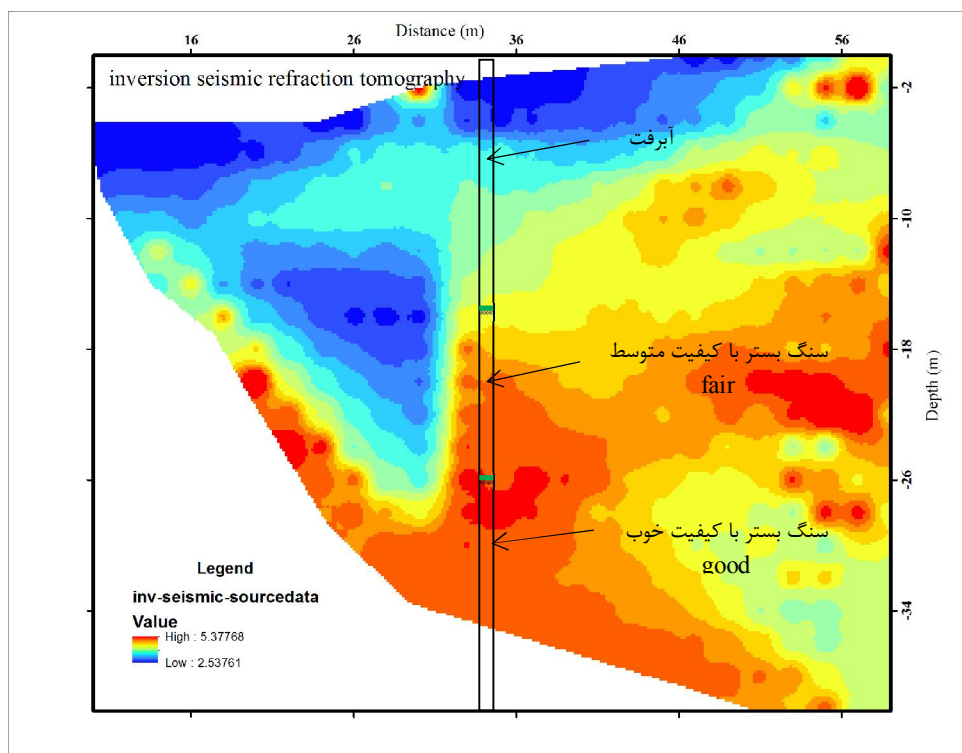
همانطور که در فصل قبل گفته شد، داده های توموگرافی لرزه ای و مقاومت ویژه الکتریکی از ساختگاه یکی از پروژه های سد در حال اجرا، در ایران برداشت شده است که به دلیل تقاضای شرکت کارفرما (شرکت سابیر) مبنی بر عدم ذکر منطقه و سد موردنظر، از ذکر نام منطقه و سد خودداری می شود.

۶-۳-۱ داده‌های توموگرافی لرزه‌ای

جهت تحلیل روش لرزه درون چاهی^۱ یکی از سدهای در حال احداث ایران برای ارزیابی کیفیت توده سنگ زیر سطحی و تفکیک آبرفت سنگ بستر، از ۲۱ عدد فرستنده لرزه‌ای در یک گمانه و ۱۵ عدد گیرنده ژئوفون به فاصله ۲ متر از یکدیگر در گمانه دیگر استفاده شده است. تولید انرژی لرزه‌ای با استفاده از انفجار انجام شده است.

روش توموگرافی لرزه‌ای انکساری از توانایی قابل توجهی در تشخیص ویژگی‌های لرزه‌شناسی لایه‌های زمین برخوردار است. ابتدا با استفاده از نرم‌افزار PickWin اولین زمان رسیده‌ها به دست آمده و بعد با استفاده از نرم‌افزار GeotTom CG، وارون‌سازی داده‌های توموگرافی لرزه‌ای انکساری انجام شد با توجه به نتایج وارون‌سازی داده‌های توموگرافی لرزه انکساری، لایه آبرفتی و هوازده سرعتی معادل ۳ کیلومتر بر ثانیه و سنگ بستر سرعتی بالای ۴ کیلومتر بر ثانیه دارد. با کیفیت‌ترین قسمت‌های توده سنگ دارای سرعت‌های عمده بین ۴ تا بیش از ۴/۵ کیلومتر بر ثانیه هستند. قسمت‌های با سرعت کم، قسمت‌هایی با کیفیت پایین توده سنگ و یا از نوع آبرفتی هستند (شکل ۶-۴).

^۱ Downhole Seismic

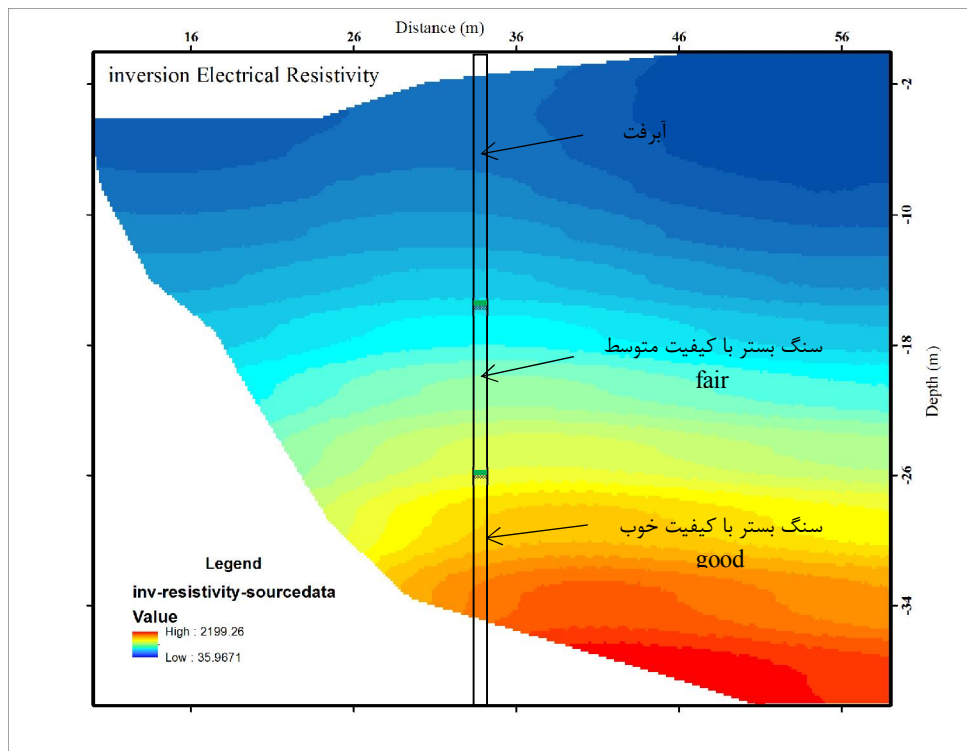


شکل ۴-۶- نتایج وارون سازی توموگرافی لرزه‌ای انکساری برای اعماق پایین‌تر از ۱۴ متری سطح زمین.

۲-۳-۶ داده‌های مقاومت‌ویژه الکتریکی

برای برداشت داده‌های توموگرافی مقاومت‌ویژه الکتریکی، از آرایه شلومبرژه با فاصله الکترودی (AB/2) حداقل یک و حداکثر ۱۵۰ متر استفاده شده است. مدل‌سازی وارون داده‌ها با استفاده از نرم‌افزار RES2DINV انجام شد.

با توجه به نتایج حاصل از وارون‌سازی داده‌های مقاومت‌ویژه الکتریکی لایه‌های زیر سطحی مطابق شکل ۵-۶ به دست آمد. وارون‌سازی با روش کمترین مربعات گوس- نیوتن و با استفاده از نرم‌افزار RES2DINV محاسبه شده است. برای تلفیق دو نقشه حاصل از وارون‌سازی داده‌های مقاومت‌ویژه و توموگرافی لرزه انکساری، داده‌های وارون‌سازی شده از عمق ۱۴ متر به پایین و زیر سطح ایستابی مورد استفاده قرار گرفتند. عمق صفر در نقشه برابر عمق ۱۴ متر در واقعیت می‌باشد. لایه‌های به دست آمده نشان می‌دهد که مقاومت‌ویژه به طرف عمق به تدریج افزایش می‌یابد.



شکل ۵-۶- نتایج وارون سازی داده‌های مقاومت ویژه الکتریکی برای اعماق پایین‌تر از ۱۴ متری سطح زمین.

۳-۳-۶ اعتبارسنجی نتایج حاصل از وارون سازی داده‌ها

برای اعتبار سنجی مدل‌های به دست آمده، از اطلاعات یا نتایج نگار حفاری برداشت شده از گمانه k224 استفاده می‌شود (جدول ۱-۶ و جدول ۲-۶). مقدار شاخص کیفیت سنگ^۱ یا RQD به ناپیوستگی موجود در توده سنگ وابسته است. به منظور محاسبه این شاخص، برای طول معینی از حفاری، طول مغزه‌های سنگ بزرگ‌تر از ۱۰ سانتی‌متر را اندازه‌گیری کرده و مجموع آن‌ها را به طول حفاری معین شده تقسیم نموده و شاخص RQD را بر حسب درصد محاسبه می‌کنند و سپس توده سنگ رده‌بندی می‌شود. شاخص دیگری که هم زمان با محاسبه درصد RQD در محل گمانه در حال حفاری محاسبه می‌گردد؛ درصد مغزه بازیافتی^۲ می‌باشد. به این منظور، مجموع طول مغزه‌های حاصل

^۱ Rock Quality Designation

^۲ Core Recovery

از حفاری را به طول معین حفاری تقسیم می‌کنند و درصد آن را به دست می‌آورند. شاخص تعداد شکستگی‌ها در هر متر^۱، نیز به تعداد درزه و شکاف در یک متر اشاره دارد.

گمانه K224 در کف سد (در فاصله ۳۳ متری مقطع به دست آمده شکل ۴-۶) حفاری شده و اطلاعات مغزه های آن مورد بررسی قرار گرفت. نتایج نشان می‌دهند که روش‌های ژئوفیزیکی مرز لایه آبرفت و سنگ بستر را به خوبی تفکیک کرده‌اند و تغییر خصوصیات در سنگ‌ها مؤید این مطلب است. با تلفیق نگار حفاری شده و شکل ۵-۶، مناطق آبرفت، سنگ بستر با کیفیت متوسط و سنگ بستر با کیفیت خوب ارزیابی شده است (بر اساس اطلاعات نگار حفاری، ضخامت آبرفت ۳۰ متر است و با توجه به این که مقطع ژئوفیزیکی (شکل ۴-۶ و شکل ۵-۶)، از عمق ۱۴ متری به پایین می‌باشد، بنابراین، در عمق ۱۶ متری مرز آبرفت مشخص شده است، سنگ بستر با کیفیت متوسط^۲ یا fair تا عمق ۲۶ متری و سنگ بستر با کیفیت خوب^۳ یا good نیز از عمق ۲۶ متر شروع و تا ۴۱ متر ادامه دارد).

جدول ۱-۶- نگار گمانه حفاری شده در کف سد (راهنمای نگار گمانه در

جدول ۲-۶ آمده است)

Drillhole	KD3	K224
Elevation	2156	2156
Depth [m]	5	
10		
15		
20		
25		
30		
35	Rock	Rock
40		
45		
50		
55		

¹ Fracture Per M

² Fair

³ Good

جدول ۲-۶- راهنمای نگار گمانه حفاری شده

تعداد شکستگی‌ها در یک متر	شاخص کیفیت سنگ (%)	بازیابی مغزه (%)	دسته‌بندی مغزه
0-1	90-100	90-100	Excellent
2-4	75-89	75-89	good
5-9	50-74	50-74	fair
> 10	25-49	25-49	Poor
منطقه خرد شده	0-24	0-24	Very poor

۴-۶ تلفیق نتایج حاصل از وارون سازی داده‌های واقعی

توموگرافی لرزه‌ای و مقاومت‌ویژه الکتریکی

با استفاده از روش شاخص دان برای بهینه سازی تعداد خوشه‌ها عدد ۱۲ به دست آمد (شکل ۶-۶) که با توجه به نقشه‌های به دست آمده برای مقاومت‌ویژه الکتریکی و توموگرافی لرزه‌ای انکساری (شکل ۴-۶ و شکل ۵-۶)، تعداد خوشه مناسبی از لحاظ تفکیک مناطق آبرفتی و سنگ بستر است. روش‌های مختلفی برای بهینه سازی تعداد خوشه‌ها، نظیر: شاخص دان، دیویس-بولدین^۱، بال-هال^۲، کالینسکی-هاراباسز^۳، وجود دارد [۷۸]. شاخص دان بر پایه ایده‌ای که درون خوشه‌ها به هم نزدیک‌تر و خوشه‌ها از هم جدا باشند، بیان شده است. بنابراین این شاخص، انتخاب و به کار برده شد. با به کار بردن فرمول‌های هر یک از الگوریتم‌های خوشه‌بندی (روش میانگین k ، روش میانگین فازی و روش گوستافسون کسل) در محیط MATLAB، کد محاسباتی برای آن‌ها نوشته شده و نقشه مربوطه ترسیم گردید (شکل ۷-۶ و شکل ۸-۶ و شکل ۹-۶). در مقایسه با میانگین فازی و میانگین k در روش

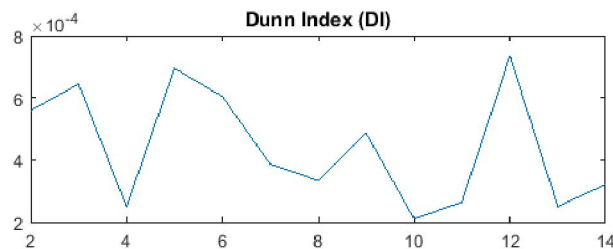
¹ Davies Bouldin

² Ball Hall

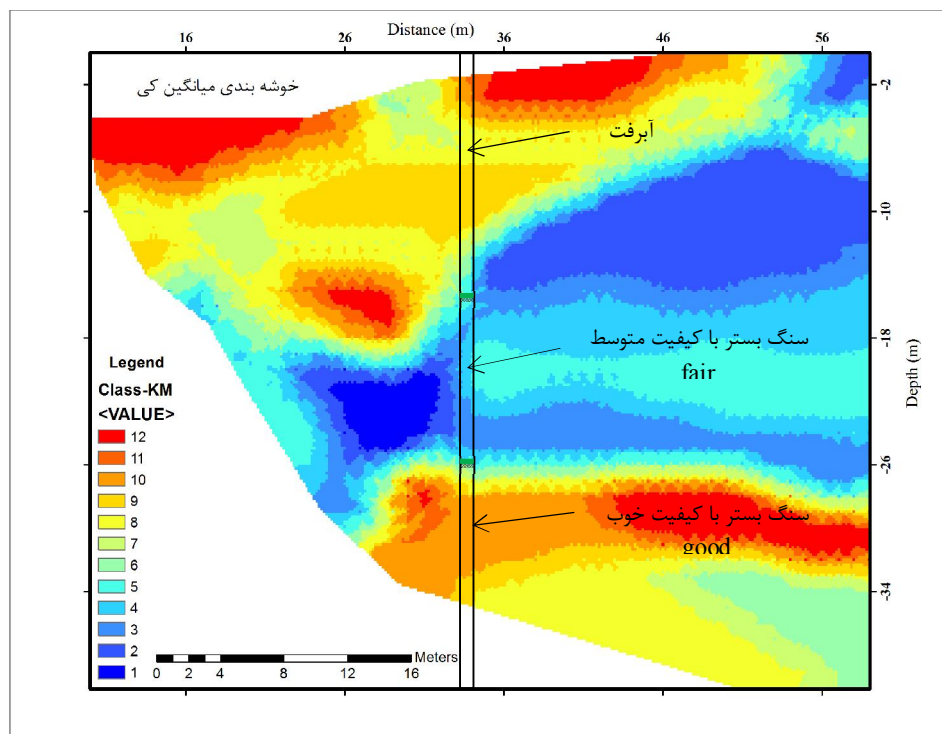
³ Calinski Harabasz

گوستافسون- کسل هر خوشه علاوه بر مرکز خوشه، توسط یک ماتریس متقارن، معین و مثبت مشخص می‌شود. این ماتریس برای هر خوشه یک مقیاس ایجاد می‌کند. باید این موضوع را هم در نظر گرفت که با انتخاب دلخواه و اختیاری ماتریس‌ها، فاصله‌ها می‌توانند به طور دلخواه کوچک شوند. برای اجتناب از حداقل سازی تابع هدف با ماتریس‌های با ورودی‌های تقریباً صفر، نیاز به مقداری ثابت برای خوشه‌ها با ماتریسی با دترمینان یک داریم. که در این صورت، شکل خوشه‌ها متغیر است و اندازه‌هایشان ثابت است. اما گوستافسون و کسل، اشکال متفاوت برای خوشه‌ها را نیز مقدور ساختند. بدین ترتیب که یک مقدار ثابت e برای هر ماتریس A معرفی کردند و در حالت کلی باید $\det(A)=e$ باشد [۷۶]. اگرچه انتخاب ثابت‌ها نیز نیاز به یک دانش پیشین در مورد خوشه‌ها دارد. اشکال بیضوی حاصل از روش خوشه‌بندی گوستافسون- کسل در مورد مثال‌های یکسان با روش میانگین فازی که در نتیجه آن اشکال کروی حاصل می‌شود، خوشه‌بندی بهتری ارائه می‌دهد. البته موضوع تجمع نقاط در مراکز خوشه‌ها در مورد خوشه‌بندی امکانی با روش گوستافسون- کسل نیز به صورت یک مسئله باقی می‌ماند. اگر داده‌ها با رویکرد امکان خوشه‌بندی شوند، فاکتورهای توسعه برای خوشه‌ها می‌توانند برای تشخیص اشکال در تصاویر مورد استفاده قرار گیرند. موقعیت و جهت را از مرکز خوشه و ماتریس می‌توان به دست آورد. کیفیت نتیجه حاصل از این روش به شدت وابسته به داده‌های در دسترس است. در مقایسه با میانگین فازی هزینه‌های محاسباتی روش خوشه‌بندی گوستافسون کسل بیشتر است. یک روش پیشنهادی مؤثر در کاهش گام‌های تکرار و افزایش سرعت همگرایی، آغاز الگوریتم گوستافسون کسل با نتایج حاصل از یک اجرای روش میانگین فازی می‌باشد که با پیاده کردن آن نتایج مطلوبی حاصل شد. با توجه به نتایج حفاری و تحلیل مغزه‌ها، سه قسمت آبرفت، سنگ بستر با کیفیت متوسط و سنگ بستر با کیفیت خوب مشخص گردید. با تطابق این نتایج با خوشه‌بندی‌های مختلف، این نتیجه حاصل گردید که روش خوشه‌بندی گوستافسون- کسل فازی، نتیجه بهتری را نسبت به روش‌های میانگین k و میانگین فازی، نشان داده و لایه‌های مختلف را به خوبی از هم تفکیک نموده است (شکل ۶-۹). در شکل ۶-۷ که نتایج خوشه‌بندی میانگین k را نشان

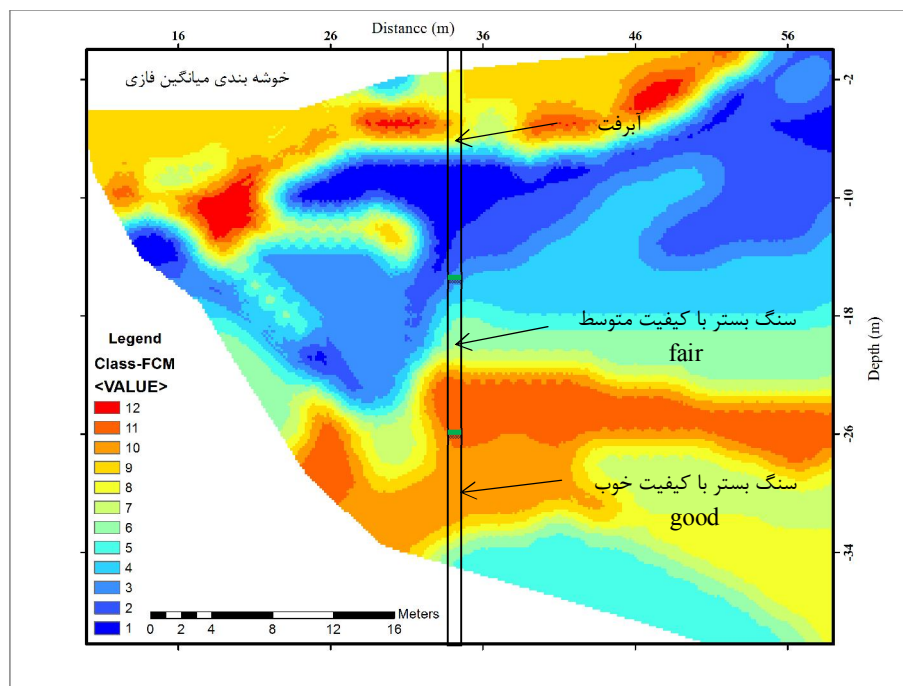
می‌دهد، خوشه‌ها لایه‌های مختلف زمین را به خوبی از هم تمایز نمی‌دهد به طوری که شروع و انتهای لایه، انطباق مناسبی با مرز خوشه‌بندی ایجاد نکرده است. در شکل ۶-۸ که نتایج خوشه‌بندی میانگین فازی را نشان می‌دهد نیز لایه‌ها با مرز خوشه‌بندی هماهنگی مناسبی ندارد.



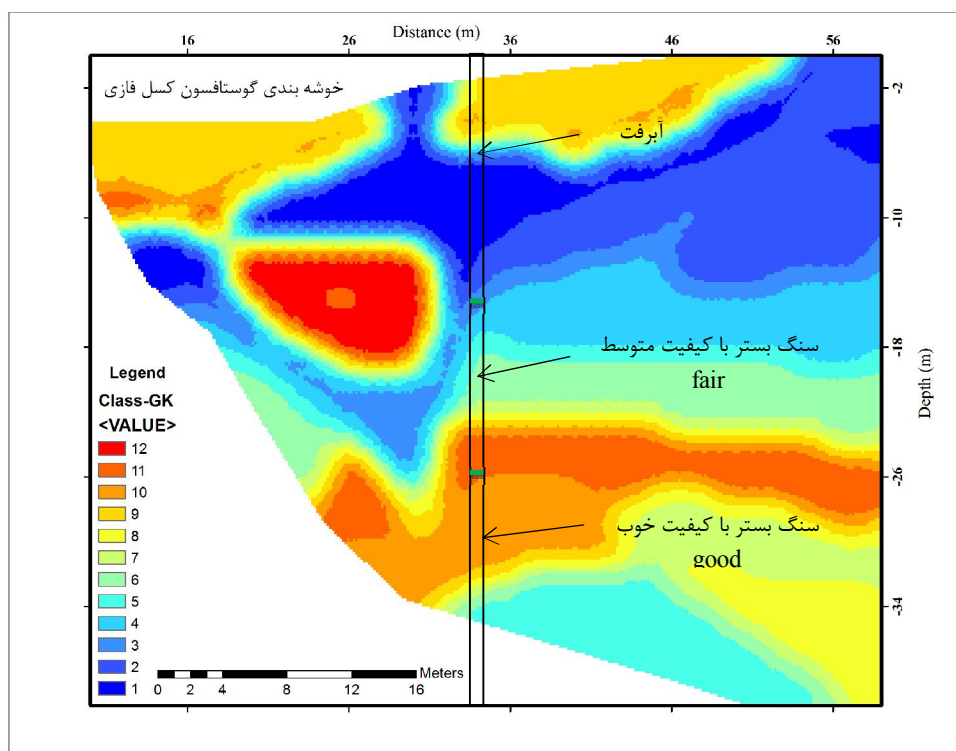
شکل ۶-۶- شاخص دان که برای محاسبه تعداد خوشه‌های بهینه مورد استفاده قرار گرفته است. در این شاخص حداکثر مقدار به دست آمده برای خوشه‌ها، به عنوان تعداد خوشه بهینه در نظر گرفته می‌شود. محور افقی تعداد خوشه‌ها و محور عمودی مقدار شاخص را نشان می‌دهد.



شکل ۶-۷- مدل منطقه‌بندی به دست آمده از اعمال الگوریتم میانگین k بر روی مدل‌های وارون‌سازی شده. زون‌ها (طبقه‌ها) با رنگ مشخص شده‌اند.



شکل ۶-۸- مدل منطقه‌بندی به دست آمده از اعمال الگوریتم میانگین فازی بر روی مدل‌های وارون‌سازی شده. زون‌ها (طبقه‌ها) با رنگ مشخص شده‌اند.



شکل ۶-۹- مدل منطقه‌بندی به دست آمده از اعمال الگوریتم گوستافسون-کسل فازی بر روی مدل‌های وارون‌سازی شده. زون‌ها (طبقه‌ها) با رنگ مشخص شده‌اند.

خوشه‌بندی فازی از تلفیق رویکرد فازی در بحث خوشه‌بندی برای کاربردی‌تر نمودن آن و انطباق بیشتر با دنیای واقعی حاصل شده است. با توجه به نتایج مشاهده می‌شود که استفاده هم‌زمان از چندین مجموعه داده نتایج دقیق‌تری را در تفسیر ارائه می‌دهد. با وارون‌سازی داده‌ها، کیفیت مواد آبرفتی و سنگ بستر مورد ارزیابی قرار گرفته و روش‌های خوشه‌بندی نیز این مرزها را مورد تحلیل قرار داد. با توجه به نتایج حفاری و تحلیل مغزه‌ها و انطباق این نتایج با خوشه‌بندی‌های میانگین k ، میانگین فازی و گوستافسون-کسل فازی، این نتیجه حاصل گردید که روش خوشه‌بندی گوستافسون-کسل فازی نتیجه بهتری را نشان داده و لایه‌های مختلف را به خوبی از هم تفکیک کرده است. با اجرای الگوریتم گوستافسون-کسل با نتایج حاصل از روش میانگین فازی، گام‌های تکرار، کاهش و سرعت همگرایی افزایش یافت. با توجه به شکل ۶-۹، لایه‌بندی آبرفت، سنگ بستر با کیفیت متوسط و سنگ بستر با کیفیت خوب با مرز خوشه‌بندی گوستافسون-کسل، سازگاری خوبی را نشان می‌دهد. در دو الگوریتم دیگر لایه‌ها با دقت پایین‌تری از هم تفکیک شده‌اند.

۵-۶ اعمال الگوریتم TSVM بر اساس خوشه‌بندی طیفی

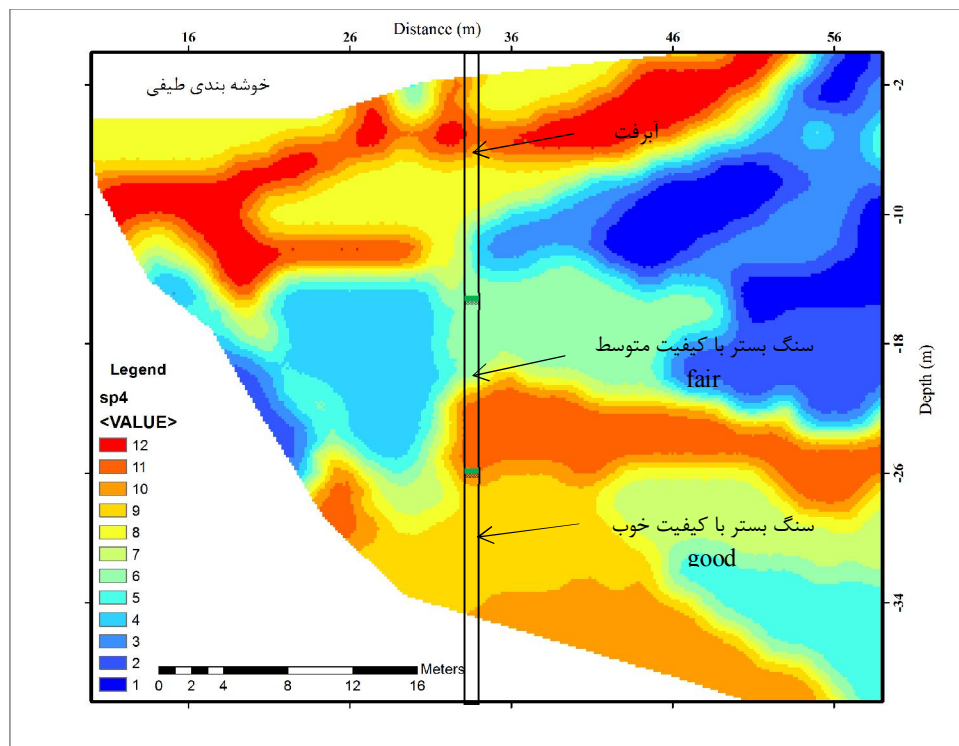
(TSVMSC) بر روی نتایج حاص از وارون‌سازی داده‌های

ژئوفیزیکی

با به کار بردن فرمول‌های الگوریتم خوشه‌بندی طیفی و شاخص دان در محیط MATLAB، کد محاسباتی برای آن‌ها نوشته شده و نقشه مربوطه ترسیم شد (شکل ۶-۱۰).

مدل‌های ارائه شده در شکل ۶-۴ و شکل ۶-۵ نرمال شده و تحت تحلیل خوشه‌ای طیفی قرار گرفتند. با محاسبه تشابهات جفت جفت گوسین با $\sigma = 2$ ، (پارامتر σ در فرمول ۱۰ معرفی شده

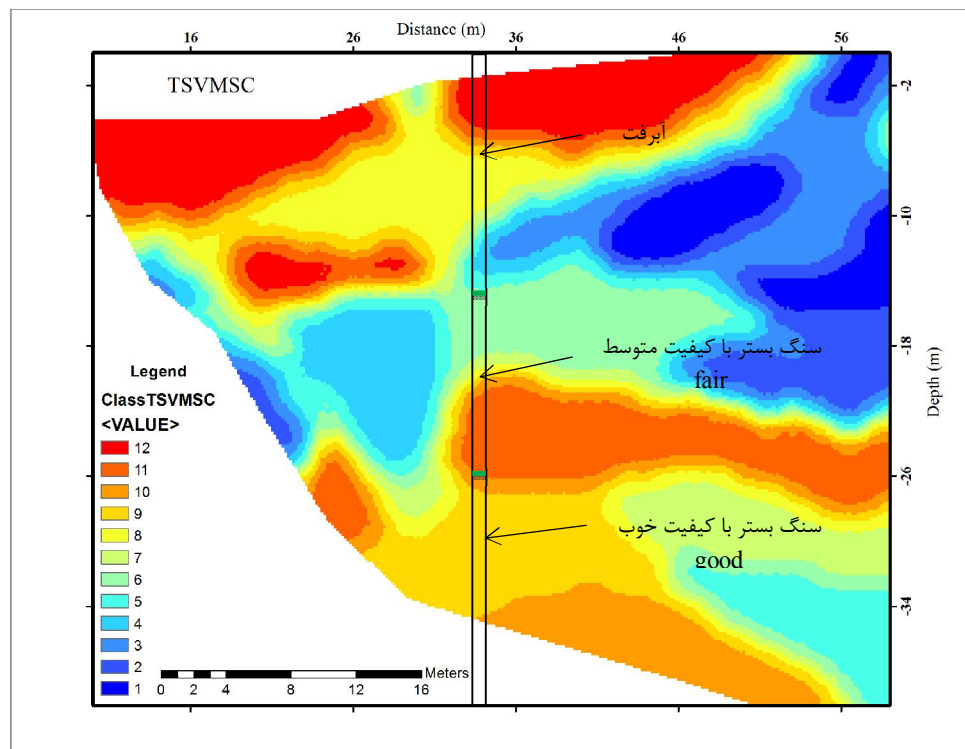
است) ماتریس شباهت به دست آمد. توجه این که محاسبه تشابهات جفت جفت نیازمند همه مدل‌های مورد نظر است که باید به طور کامل در کنار هم قرار گرفته و به صورت یکسان از هم تفکیک شوند. بردارهای ویژه مطابق با بزرگ‌ترین مقادیر ویژه لاپلاسین گراف مورد مطالعه قرار گرفتند (به دلیل بزرگ بودن ماتریس (۱۸۹۱ سطر در ۱۸۹۱ ستون)، از آوردن مقادیر و بردارهای ویژه خودداری می‌شود). مدل منطقه‌بندی با استفاده از خوشه‌بندی طیفی در شکل ۶-۱۰ آمده است. بر اساس نتایج حاصل از اعمال روش TSVMSC بر روی نتایج حاصل از وارون سازی داده‌های ژئوفیزیکی، سه منطقه آبرفتی و ۹ منطقه سنگ بستر بدست آمده و تفکیک بهتری نسبت به خوشه‌بندی‌های انجام شده توسط روش‌های میانگین k ، میانگین فازی و گوستافسون-کسل فازی، نشان داده است. منطقه‌بندی با استفاده از روش TSVMSC، با توجه به دقت بالا و عملکرد پایدار آن نسبت به دیگر الگوریتم‌ها، تلفیق بهبود یافته‌تری را نشان می‌دهد (شکل ۶-۱۱).



شکل ۶-۱۰- مدل منطقه‌بندی به دست آمده از اعمال خوشه‌بندی طیفی بر روی مدل‌های وارون. زون‌ها (طبقه‌ها) با رنگ‌های آبی تا قرمز مشخص شده‌اند.

با استفاده از خوشه‌بندی طیفی در ماشین بردار پشتیبان، الگوریتم TSVM بر اساس خوشه‌بندی طیفی حاصل گردید. به گونه ای که ابتدا از خوشه‌بندی طیفی برای خوشه‌بندی نمونه‌های مشابه بدون برچسب، استفاده می‌شود. سپس همه نمونه‌ها را در همان خوشه یا طبقه قرار می‌دهد و سرانجام کلاسیفایر از روش یادگیری ترانس داکتیو برای طبقه‌بندی استفاده می‌کند. چون این روش از خوشه‌بندی طیفی در توزیع مؤثر اطلاعات نمونه‌های بدون برچسب، استفاده می‌کند، بنابراین می‌تواند به طور مؤثری عملکرد الگوریتم TSVM را بهبود بخشد. شکل ۶-۱۱ مدل منطقه‌بندی شده نهایی را با یک جواب ۱۲ خوشه‌ای نشان می‌دهد. ساختارهای اصلی در مدل‌های لرزه‌ای و مقاومت‌ویژه (شکل ۶-۴ و شکل ۶-۵) به عنوان اطلاعات ترکیب شده توسط خوشه‌ها، نشان داده شده‌اند. هر زون می‌تواند با مقادیر سرعت موج P و مقاومت‌ویژه الکتریکی توصیف شود. در منطقه آبرفتی سرعت موج P به دلیل از هم گسستگی ذرات و مقاومت‌ویژه به دلیل رسانش بالاتر، پایین است. هر چه از سطح به عمق می‌رویم، سرعت موج P و مقاومت‌ویژه به طور نسبی افزایش نشان می‌دهند. نقشه ارائه شده در شکل ۶-۱۱ که حاصل از اعمال روش TSVMSC بر روی نتایج وارون سازی است. سه منطقه آبرفتی با کیفیت پایین توده سنگ و ۹ منطقه سنگ بستر با کیفیت بالای توده سنگ را به صورت زون‌بندی نمایش می‌دهد و نیز با توجه به نتایج اعتبارسنجی الگوریتم، در تفکیک مناطق به خوبی عمل کرده و نسبت به روش‌های دیگر خوشه‌بندی دقیق‌تر است. الگوریتم TSVMSC در فرآیند تلفیق مدل، ناهمگنی فضایی موجود در همه مدل‌های توموگرافی را در برمی‌گیرد. در خوشه‌های ۴ و ۶ که در شکل ۶-۱۱ نشان داده شده است، سرعت موج P با توجه به نقشه‌های شکل ۶-۴ و شکل ۶-۵ نسبت به لایه‌های بالاتر کاهش و مقاومت‌ویژه افزایش را نشان می‌دهد که این مناطق در خوشه‌بندی به خوبی تفکیک شده‌اند. از لحاظ زمین شناسی، این قسمت سنگ آهک متخلخل با تخلخل مؤثر پایین است. با تلفیق نتایج وارون‌سازی داده‌های توموگرافی لرزه‌ای انکساری و مقاومت‌ویژه الکتریکی با استفاده از این الگوریتم کیفیت توده سنگ از ضعیف به خوب به ترتیب از ۱ تا ۱۲ در شکل ۶-۱۱ ارزیابی شده است. مناطق آبرفتی (خوشه‌های ۳، ۷ و ۱۲) به دلیل از هم گسستگی ذرات، دارای

سرعت لرزه‌ای پایین هستند و نیز به خاطر رسانش بالاتر در اثر وجود رس و تخلخل، مقاومت ویژه الکتریکی پایین می‌باشد. در محدوده سنگ بستر به طور نسبی به طرف عمق (به جزء خوشه‌های ۴ و ۶)، سرعت لرزه‌ای و مقاومت ویژه الکتریکی افزایش می‌یابند. با توجه به نتایج حفاری و تحلیل مغزه‌ها و تطابق آن با الگوریتم، این نتیجه حاصل شد که الگوریتم TSVMSC، لایه‌های مختلف زمین را به خوبی از هم تفکیک نموده و در خوشه‌های ۱ تا ۱۲ جای داده است. با توجه به نتایج به دست آمده از ۱۲ کلاس مورد بررسی (که با استفاده از روش شاخص دان محاسبه شد)، سه کلاس در منطقه آبرفتی با کیفیت پایین توده سنگ و ۹ کلاس در منطقه سنگ بستر با کیفیت بالای توده سنگ واقع شده‌اند.



شکل ۶-۱۱- مدل منطقه‌بندی به دست آمده از الگوریتم TSVMSC بر روی مدل‌های وارون. زون‌ها (طبقه‌ها) با رنگ‌های آبی تا قرمز مشخص شده‌اند.

۶-۶ اعتبار سنجی نتایج وارون سازی توמוگرافی لرزه‌ای

انکساری

به منظور تعیین خصوصیات فیزیکی و مکانیکی توده سنگ‌ها و نیز طبقه بندی آن‌ها، آزمایش‌هایی بر روی نمونه‌های اخذ شده از گمانه‌ها، مطابق با روش‌های پیشنهادی انجمن بین‌المللی مکانیک سنگ (ISRM) و استاندارد (ASTM) انجام گردید. نتایج آزمایش‌ها در جدول ۶-۳ آمده است. این آزمایش شامل اندازه‌گیری چگالی (در حالت طبیعی، اشباع و خشک)، رطوبت، تخلخل، جذب آب، سرعت تراکمی (در حالت طبیعی و اشباع) و سرعت برشی (در حالت طبیعی و اشباع) در متراژ مشخصی از نمونه صورت گرفته است. از هر گمانه حفر شده، چند نمونه برای اندازه‌گیری استفاده شده است. در مقطع استفاده شده در این مطالعه، تنها یک نمونه جهت اندازه‌گیری‌های سرعت تراکمی و برشی استفاده گردیده است.

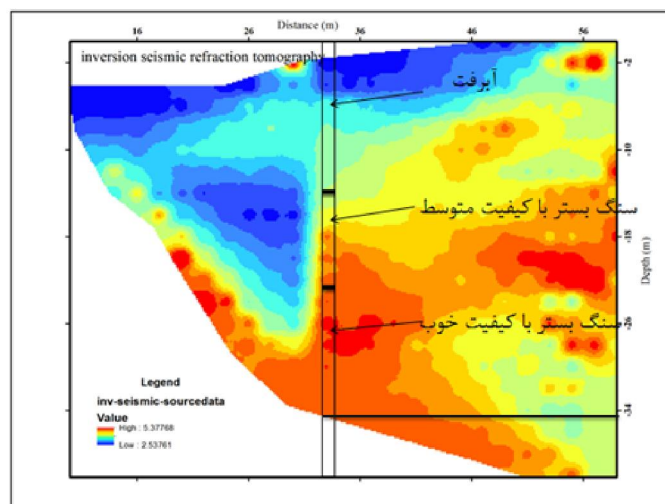
جدول ۶-۳- اندازه‌گیری سرعت‌های تراکمی (V_p) و برشی (V_s) به همراه دیگر پارامترها

B.H No	عمق (m)	چگالی (gr/cm ³)			رطوبت (%)	تخلخل (%)	جذب آب (%)	سرعت تراکمی (m/s)		سرعت برشی (m/s)	
		طبیعی	اشباع	خشک				طبیعی	اشباع	طبیعی	اشباع
K224	۴۸.۴۱ ۴۸.۸۵	۲.۷۵	۲.۷۵	۲.۷۴	۰.۲	۱.۱	۰.۴	۶۸۴۵	۶۸۴۵	۳۷۹۴	۳۷۹۴

در عمق ۴۸/۴۱ تا ۴۸/۸۵ متری (عمق ۳۴/۴۱ تا ۳۴/۸۵ مقطع) و فاصله افقی ۳۳ متری، سرعت حاصل از وارون‌سازی برابر با ۴/۷۳ کیلومتر بر ثانیه و سرعت آزمایشگاهی برابر با ۶/۸۵ کیلومتر بر ثانیه است (شکل ۶-۱۲). بنابراین خطای تخمین سرعت معادل ۳۰/۹۵ و دقت آن معادل ۶۹/۰۵ درصد می‌باشد (معادله ۶-۱).

$$\frac{4.73}{6.85} \times 100 = 69.05\%$$

(۶-۱)

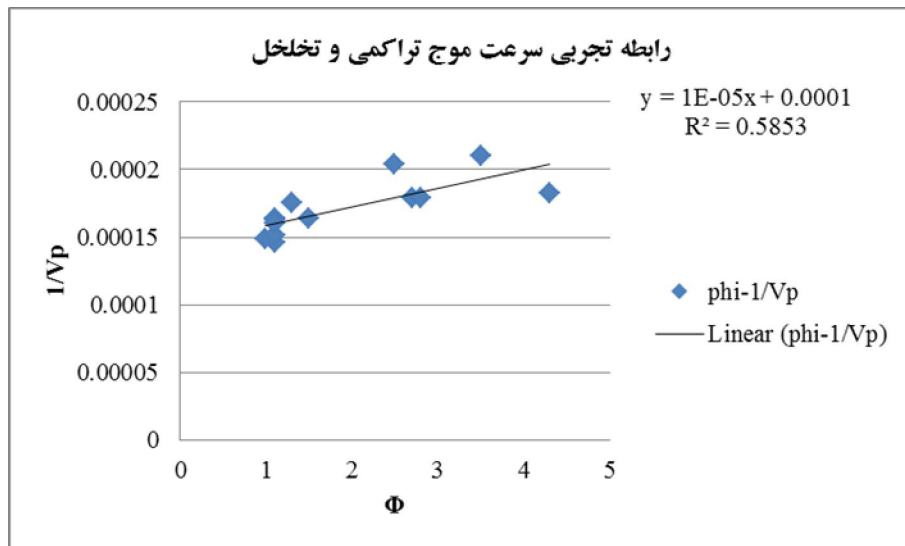


شکل ۶-۱۲- نمونه آزمایشگاهی برای تعیین سرعت موج لرزه‌ای که از عمق ۴۸/۴۱ تا ۴۸/۸۵ متری گرفته شد.

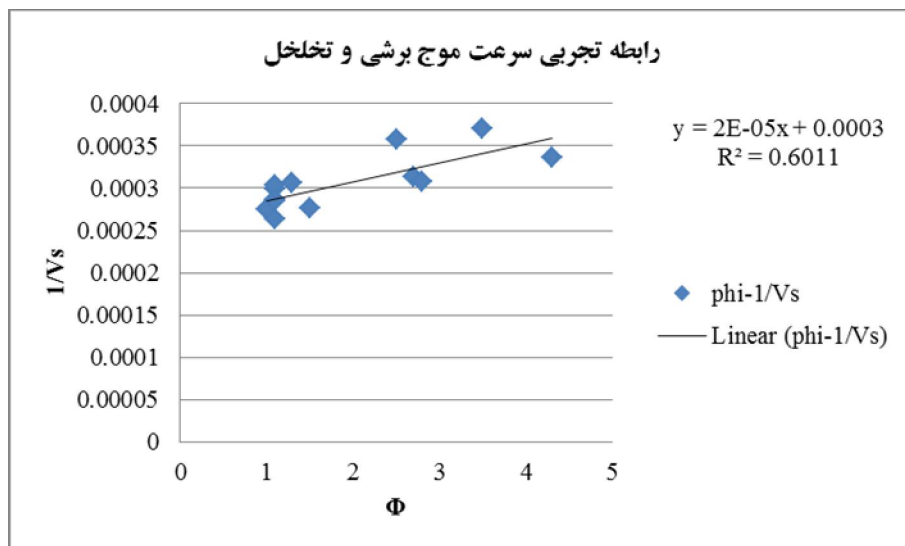
بر اساس نتایج حاصل از آزمایش‌ها برای سرعت تراکمی، سرعت برشی و تخلخل (جدول ۴-۶) روابط تجربی برای محاسبه سرعت‌های تراکمی و برشی، در این محدوده، به ترتیب با درصد همبستگی ۷۶/۵ و ۷۵/۵ به دست آمد (معادله ۶-۲). تعداد نمونه‌های استفاده شده برای این محاسبه برابر ۱۳ نمونه می‌باشد. سرعت تراکمی یا V_p (شکل ۶-۱۳) و سرعت برشی یا V_s (شکل ۶-۱۴) و تخلخل با Φ نشان داده شده است.

جدول ۴-۶- سرعت موج تراکمی (V_p) و برشی (V_s) و تخلخل متناظر نمونه‌ها

Φ	$1/V_p$	$1/V_s$
۱.۱	۰.۰۰۰۱۴۶	۰.۰۰۰۲۶۴
۴.۳	۰.۰۰۰۱۸۳	۰.۰۰۰۳۳۷
۱	۰.۰۰۰۱۴۸	۰.۰۰۰۲۷۵
۲.۵	۰.۰۰۰۲۰۴	۰.۰۰۰۳۵۷
۱.۵	۰.۰۰۰۱۶۴	۰.۰۰۰۲۷۶
۲.۷	۰.۰۰۰۱۷۹	۰.۰۰۰۳۱۳
۱.۱	۰.۰۰۰۱۶۴	۰.۰۰۰۲۹۹
۱.۱	۰.۰۰۰۱۵۱	۰.۰۰۰۲۸۶
۱.۳	۰.۰۰۰۱۷۶	۰.۰۰۰۳۰۶
۱.۱	۰.۰۰۰۱۶۳	۰.۰۰۰۳۰۴
۳.۵	۰.۰۰۰۲۱۱	۰.۰۰۰۳۷۱
۲.۸	۰.۰۰۰۱۷۹	۰.۰۰۰۳۰۸
۱.۱	۰.۰۰۰۱۶	۰.۰۰۰۲۸۵



شکل ۱۳-۶- نمودار تعیین سرعت موج تراکمی و برشی با استفاده از تخلخل موجود



شکل ۱۴-۶- نمودار تعیین سرعت موج تراکمی و برشی با استفاده از تخلخل موجود

با توجه به همبستگی بالای روابط بین سرعت‌های موج تراکمی و برشی با تخلخل، این نتیجه حاصل می‌شود که وجود خلل و فرج نقش تعیین‌کننده‌ای در تفسیر سرعت‌ها ایفا می‌کند. همان‌طور که از روی شکل ۱۳-۶ و شکل ۱۴-۶ دیده می‌شود، روابط بین سرعت‌های موج تراکمی (V_p) و موج برشی

(V_s) با تخلخل به صورت معادله (۲-۶) است و ضریب همبستگی آن‌ها به ترتیب ۷۶/۵ و ۷۵/۵ می‌باشد.

$$\begin{aligned} \frac{1}{V_p} &= 1E - 05\phi + 0.0001 & R &= 76.5 \\ \frac{1}{V_s} &= 2E - 05\phi + 0.0003 & R &= 77.5 \end{aligned} \quad (۲-۶)$$

ضریب تعیین روابط سرعت موج تراکمی و برشی با تخلخل، به ترتیب ۵۸/۵۳ و ۶۰/۱۱ درصد به دست آمده است.

۶-۷ جمع بندی و نتیجه گیری

در این فصل تلفیق داده‌های مصنوعی و صحرایی توموگرافی لرزه‌انکساری و مقاومت‌ویژه الکتریکی با استفاده از الگوریتم TSVMSC مورد تحلیل قرار گرفت. اعتبار سنجی نتایج حاصل از وارون سازی و تلفیق نیز بررسی شد. الگوریتم TSVMSC، به دلیل اینکه ناهمگنی فضایی در همه مدل‌ها را در بر گرفته و عملکرد و دقت بالایی را نشان داده است، نسبت به دیگر روش‌های خوشه‌بندی برتری بالایی دارد.

فصل، منقسم

نتیجہ گیری و پیشہدات

۱-۷ مقدمه

همان طور که اشاره شد، رهیافت‌های چند روشی به طور گسترده‌ای در کاهش عدم قطعیت تحلیل‌ها و تفسیرهای ژئوفیزیک سطحی، به کار برده می‌شوند. تفسیر توأم مدل‌های چندگانه، به طور ذاتی مبتنی بر تجربه مفسر بوده و دانش محور است. نتایج حاصل از این پژوهش در ذیل مورد بررسی قرار می‌گیرد:

۲-۷ نتیجه‌گیری

- ۱- با توجه به نتایج وارون‌سازی داده‌های توموگرافی لرزه انکساری، سرعت تراکمی (V_p) در دامنه ۱۹۸۰ تا ۴۶۰۰ متر بر ثانیه قرار گرفته است.
- ۲- در ناحیه آبرفتی سرعت موج P به دلیل از هم گسستگی ذرات و مقاومت‌ویژه به دلیل رسانندگی الکتریکی بالاتر، پایین است. هر چه از سطح به عمق می‌رویم، سرعت موج P و مقاومت‌ویژه به طور نسبی افزایش نشان می‌دهند.
- ۳- با استفاده شاخص دان برای بهینه سازی تعداد خوشه‌ها، عدد ۱۲ به دست آمد که با توجه به نقشه‌های به دست آمده برای مقاومت‌ویژه الکتریکی و توموگرافی لرزه‌ای انکساری، تعداد خوشه مناسبی از لحاظ تفکیک مناطق آبرفتی و سنگ بستر است.
- ۵- همان‌طور که در این تحقیق نشان داده شد، الگوریتم TSVMSC، ناهمگنی فضایی در همه مدل‌ها را در بر گرفته و نیز انعطاف‌پذیر بوده و توانسته به شناسایی الگوها و ساختارهای غالب موجود در داده‌های چند متغیره (داده‌های مقاومت ویژه الکتریکی و توموگرافی لرزه انکساری) کمک شایانی بنماید.

۶- با توجه به نتایج وارون سازی داده های توموگرافی لرزه انکساری، لایه آبرفتی و هوازده سرعتی معادل ۳۰۰۰ متر بر ثانیه و سنگ بستر سرعتی بالای ۴۰۰۰ متر بر ثانیه دارد. با کیفیت ترین قسمت های توده سنگ با سرعت های عمده بین ۴۰۰۰ تا بیش از ۴۵۰۰ متر بر ثانیه هستند. قسمت های با سرعت پایین، نواحی با کیفیت کم و یا آبرفتی هستند

۷- همانطور که در تحقیقات مشابه نشان داده شده، هزینه های محاسباتی روش خوشه بندی گوستافسون- کسل در مقایسه با روش میانگین فازی بیشتر است. یک روش پیشنهادی مؤثر در کاهش گام های تکرار و افزایش سرعت همگرایی، آغاز الگوریتم گوستافسون- کسل با نتایج حاصل از یک اجرای روش میانگین فازی می باشد؛ که با پیاده کردن آن نتایج مطلوبی در این تحقیق حاصل شد.

۸- با توجه به نتایج خوشه بندی میانگین k و خوشه بندی میانگین فازی، خوشه ها لایه های مختلف زمین را به خوبی از هم تمایز نمی دهد به طوری که شروع و انتهای لایه، انطباق مناسبی با مرز خوشه بندی ایجاد نکرده است.

۹- با توجه به نتایج حفاری و تحلیل مغزه ها و انطباق این نتایج با خوشه بندی های میانگین k ، میانگین فازی و گوستافسون کسل فازی، این نتیجه حاصل شد که روش خوشه بندی گوستافسون- کسل فازی نتیجه بهتری را نشان می دهد و لایه های مختلف را به خوبی از هم تفکیک می کند.

۱۱- با توجه به نتایج، لایه بندی آبرفت، سنگ بستر با کیفیت متوسط و سنگ بستر با کیفیت خوب با مرز خوشه بندی گوستافسون کسل، سازگاری خوبی را نشان می دهد و در دو الگوریتم میانگین k و میانگین فازی، لایه ها با دقت پایین تری از هم تفکیک شده اند.

۱۲- نقشه حاصل از الگوریتم TSVMSC، سه منطقه آبرفتی با کیفیت پایین توده سنگ و ۹ منطقه سنگ بستر با کیفیت بالای توده سنگ را به صورت زون بندی نمایش می دهد و نیز با توجه به نتایج اعتبارسنجی، الگوریتم در تفکیک مناطق به خوبی عمل کرده و نسبت به روش های میانگین k ، میانگین فازی و گوستافسون کسل فازی، دقیق تر است.

- ۱۳- با اعتبار سنجی نتایج وارون سازی توموگرافی لرزه‌ای انکساری، در عمق ۴۸/۴۱ تا ۴۸/۸۵ متری (عمق ۳۴/۴۱ تا ۳۴/۸۵ مقطع) و فاصله افقی ۳۳ متری، سرعت حاصل از وارون‌سازی برابر با ۴۷۳۰ متر بر ثانیه و سرعت آزمایشگاهی برابر با ۶۸۵۰ متر بر ثانیه است. بنابراین خطای تخمین سرعت معادل ۳۰/۹۵ درصد می‌باشد.
- ۱۴- بر اساس نتایج حاصل از آزمایش‌ها برای سرعت تراکمی، سرعت برشی و تخلخل، روابط تجربی برای محاسبه سرعت‌های تراکمی و برشی، در این محدوده، به ترتیب با درصد همبستگی ۷۶/۵ و ۷۷/۵ به دست آمد.
- ۱۵- با توجه به همبستگی بالای روابط بین سرعت‌های موج تراکمی و برشی با تخلخل، این نتیجه حاصل می‌شود که وجود خلل و فرج نقش تعیین کننده ای در تفسیر سرعت‌ها ایفا می‌کند.
- ۱۶- ضریب تعیین روابط سرعت موج تراکمی و برشی با تخلخل، به ترتیب ۵۸/۵۳ و ۶۰/۱۱ درصد به دست آمده است.

۷-۳ پیشنهادات

- ۱- پیشنهاد می‌شود از روش‌های خوشه‌بندی میانگین امکان، پوسته فازی و الگوریتم‌های مبتنی بر هسته و روش‌های ترکیبی برای تلفیق نتایج وارون‌سازی مقاومت‌ویژه و توموگرافی لرزه انکساری نیز استفاده گردد و نتایج این روش‌ها با یکدیگر و نیز با نتایج حاصل از این تحقیق مقایسه شود.
- ۲- برای رسیدن به یک فرمول تجربی مشخص (مانند فرمول‌های تجربی هان و نول) در محاسبه مقادیر مقاومت‌ویژه الکتریکی و توموگرافی لرزه انکساری، در یک لیتولوژی خاص، پیشنهاد می‌شود، نمونه‌هایی از محل سدهای مختلف (دارای همان لیتولوژی خاص) گرفته شده و بعد از اندازه‌گیری‌ها، مورد ارزیابی قرار گیرند.

۳- برای دستیابی به نقشه‌های دقیق‌تر، پیشنهاد می‌شود که از لایه‌های مختلف زمین‌شناسی،

ژئوفیزیکی و مکانیک سنگی برای تلفیق توسط الگوریتم‌های ترکیبی، استفاده گردد.

مراجع

- [1] Butler, D. K. (2005), "Near-surface geophysics", SEG.
- [2] Hachmoler B. and Paasche H. (2013) "Integration of surface-based tomographic models for zonation and multimodel guided extrapolation of sparsely known petrophysical parameters" *Geophysics.*, 78, 4, pp 43-53.
- [3] Gallardo, L. A. (2007) "Multiple cross-gradient joint inversion for geospectral imaging" *Geophysical Research Letters.*, 34, L19301, pp 1-5.
- [4] Rumpel H. Binot F. Gabriel G. Hinsby K. Siemon B. Steuer A. Wiederhold H. (2006) "Groundwater resources in buried valleys— A challenge to geosciences" BurVal Working Group, Leibniz Institute for Applied Geosciences.
- [5] Höppner F. Klawonn F. Kruse R. Runkler T. (1999) "Fuzzy cluster analysis: Methods for classification, data analysis and image recognition" John Wiley and Sons.
- [6] Kaufmann L. Rousseeuw P.J. (1990) "Finding groups in data: An introduction to cluster analysis" John Wiley and Sons.
- [7] Paasche H. Tronicke J. Holliger K. Green A.G. Maurer H. (2006) "Integration of diverse physical-property models: Subsurface zonation and petrophysical parameter estimation based on fuzzy c-means cluster analyses" *Geophysics.*, 71, 3, pp H33–H44.
- [8] Schön J. H. (1998) "Physical properties of rocks: Fundamentals and principles of petrophysics" Pergamon Press.
- [9] Eppstein M.J. Dougherty D.E. (1996) "Simultaneous estimation of transmissivity values and zonation" *Water Resources Research.*, 32, pp 3321–3336.

- [10] Fechner T. Dietrich P. (1997) “Lithological inversion of tomographic data” Proceedings of the 3rd Conference on Environmental and Engineering Geophysics., EEGS, Expanded Abstracts, pp 355–358.
- [11] Dietrich P. Fechner T. Whittacker J. Teutsch G. (1998) “An integrated hydrogeophysical approach to subsurface characterization” Herbert M. and Kovar K. eds. “Groundwater quality: Remediation and protection” International association of hydrological sciences., 250, pp 513–519.
- [12] Tronicke J. Dietrich P. Wahlig U. Appel E. (2002) “Integrating surface georadar and crosshole radar tomography: A validation experiment in braided stream deposits” Geophysics., 67, pp 1516–1523.
- [13] Tronicke J. Holliger K. Barrash W. Knoll M.D. (2004) “Multivariate analysis of crosshole georadar velocity and attenuation tomograms for aquifer zonation” Water Resources Research., 40, W01519, pp 1-14.
- [14] Brauchler R. Doetsch J. Dietrich P. Sauter M. (2012) “Derivation of site-specific relationships between hydraulic parameters and P-wave velocities based on hydraulic and seismic tomography” Water Resources Research., 48, W03531, pp 1-42.
- [15] Günther T. Bentley L. R. Hirsch M. (2006) “A new joint inversion algorithm applied to the interpretation of DC resistivity and refraction data” XVI International Conference on Computational Methods in Water Resources., Expanded Abstracts, pp 1–8.
- [16] Paasche H. Günther T. Tronicke J. Green A.G. Maurer H.R. Holliger K. (2007) “Integrating multi-scale geophysical data for the 3D characterization of an alluvial aquifer” EAGE Near Surface 2007 — 13th Conference on Environmental and Engineering Geophysics., Expanded Abstracts, pp 1-11.
- [17] von Luxburg U. (2007) “A tutorial on spectral clustering” Statistics and Computing., 17, pp 395–416.
- [18] Gouveia W. Scales J.A. (1998) “Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis” Journal of Geophysical Research., 103, pp 2759–2779.
- [19] Tikhonov A.N. Arsenin V.Y. (1977) “Solutions of Ill-posed Problems” Winston, John Wiley and Sons., New York, pp 1-258.

- [20] Oldenburg D.W. Li Y. (1999) "Estimating depth of investigation in DC resistivity and IP surveys" *Geophysics.*, 64, pp 403–416.
- [21] Ng A.Y. Jordan M.I. Weiss Y. (2002) "On spectral clustering: Analysis and an algorithm" Dietterich T. Becker S. and Ghahramani Z. eds. *Advances in neural information processing systems 14*: MIT Press., pp 849–856.
- [22] Meila M. Shi J. (2001) "Learning segmentation by random walks" Leen T.K. Dietterich T.G. and Tresp V. eds. *Advances in neural information processing systems 13*: MIT Press., pp 873–879.
- [23] Shi J. Malik J. (2000) "Normalized cuts and image segmentation" *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 22, pp 888–905.
- [24] Zhang X. Jiao L. Liu F. Bo L. Gong M. (2008) "Spectral clustering ensemble applied to SAR image segmentation" *IEEE Transactions on Geoscience and Remote Sensing.*, 46, pp 2126–2136.
- [25] Zelnik-Manor L. Perona P. (2005) "Self-tuning spectral clustering" *Advances in neural information processing systems 17*: MIT Press., pp 1601–1608.
- [26] West D. (2007) "Introduction to graph theory" Prentice Hall.
- [27] Tronicke J. Holliger K. (2005) "Quantitative integration of hydrogeophysical data: Conditional geostatistical simulation for characterizing heterogeneous alluvial aquifers" *Geophysics.*, 70, 3, pp H1–H10.
- [28] Goff J. A. Jordan T. H. (1988) "Stochastic modeling of seafloor-morphology: Inversion of sea beam data for second-order statistics" *Journal of Geophysical Research.*, 93, pp 13589–13608.
- [29] Hardy H.H. Beier R.A. (1994) "Fractals in reservoir engineering" World Scientific.
- [30] Desbarats A.J. Bachu S. (1994) "Geostatistical analysis of aquifer heterogeneity from the core to the basin scale: A case study" *Water Resources Research.*, 30, pp 673–684.
- [31] Gelhar L.W. (1993) "Stochastic subsurface hydrology" Prentice Hall.

- [32] Mylopoulos Y.A. Theodosiou N. Mylopoulos N.A. (1999) “A stochastic optimization approach in the design of an aquifer remediation under hydrogeologic uncertainty” *Water Resources Management.*, 13, pp 335–351.
- [33] Knoll M. D. (1996), Ph.D. thesis, “A petrophysical basis for ground penetrating radar and very early time electromagnetics: Electrical properties of sand-clay mixtures”, University of British Columbia.
- [34] Han D. Nur A. Morgan D. (1986) “Effects of porosity and clay content on wave velocities in sandstones” *Geophysics.*, 51, pp 2093–2107.
- [35] Knödel K. Krummel H. Lange G. (1997) “Handbuch zur erkundung des untergrundes von deponien und altlasten band 3” *Geophysik.*, Springer.
- [36] Paasche H. Wendrich A. Tronicke J. Trela C. (2008) “Detecting voids in masonry by cooperatively inverting P-wave and georadar traveltimes” *Journal of Geophysics and Engineering.*, 5, pp 256–267.
- [37] Friedel S. (2003) “Resolution, stability and efficiency of resistivity tomography estimated from a generalized inverse approach” *Geophysical Journal International.*, 153, pp 305–316.
- [38] Colton D. Kress R. (1983) “Integral equation methods in scattering theory” Wiley and Sons.
- [39] Paasche H. Eberle D. (2011) “Automated compilation of pseudo-lithology maps from geophysical data sets: A comparison of Gustafson-Kessel and fuzzy c-means cluster analysis” *Exploration Geophysics.*, 42, pp 275–285.
- [40] Paasche H. Tronicke J. Dietrich P. (2010) “Automated integration of partially colocated models: Subsurface zonation using a modified fuzzy c-means cluster analysis algorithm” *Geophysics.*, 75, 3, pp P11–P22.
- [41] Vapnik VN. (1995) “The Nature of Statistical Learning Theory” New York, Springer-Verlag.
- [42] Cortes C. Vapnik VN. (1995) “Support vector networks” *Machine Learning.*, 20, pp 273-297.
- [43] Gammerman A. Vapnik VN. Vowk V. (1998) “Learning by transduction” *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence.*, Wisconsin., pp 148-156.
- [44] Yu X. Yang J. Zhang J. (2012) “A Transductive Support Vector Machine Algorithm Based on Spectral Clustering” *AASRI Procedia.*, PP 384-388.

- [45] Joachims T. (1999) "Transductive inference for text classification using support vector machines" Proceedings of the 16th International Conference on Machine Learning (ICML)., San Francisco, Morgan Kaufmann, pp 200-209.
- [46] Joachims T. (2001) "SVMlight" http://www-ai.cs.uni-ortmund.de/SOFTWARE/SVM_LIGHT/svm_light.eng.html.
- [47] Burges CJC. (1998) "A tutorial on support vector machine for pattern recognition" Data Mining and Knowledge Discovery., 2, pp 121-167.
- [48] Stewart R.R. (1992) "Exploration Seismic Tomography: Fundamentals" SEG Course Notes Series., University of Calgary, Calgary, Alberta, Canada, 3, pp 1-39.
- [49] Leucci G. Margiotta S. Negri N. (2004) "Geophysical and Geological Investigations in a Karstic Environment (SaliceSalentino, Lecce, Italy)" Journal of Environmental and Engineering Geophysics., 09, pp 25-34.
- [50] Dahlin T. Zhou B. (2001) "A numerical comparison or 2D resistivity imaging with 10 electrode arrays" Geophysical Prospecting., 52, pp 379-398
- [51] Griffiths D.H. Barker R.D. (1993) "Two-dimensional resistivity imaging and modelling in areas of complex geology" Journal of Applied Geophysics., 29, pp 211-226
- [52] Renolds J.K. (1997) "An introduction to Applied end Environmental Geophysics" John Wiley and Sons., pp 1-796.
- [53] Loke M.H. Barker R.D. (1996a) "Rapid least-square inverson of apparent resistivity pseudosections using a quasi-Newton method" Geophysical Prospecting., 44, pp 131-152.
- [54] Loke M.H. (2004) "2-D and 3-D electrical imaging surveys" Tutorial.
- [55] Carcione J.M. Gérard C. Herman G.C. Kroode A.P.E. (2002) "Seismic modeling" Geophysics., Luxembourg, 67, 4, pp 1304-1325.
- [56] Aki K. Richards P.G. (1980) "Quantative seismology, Theory and methods" W.H.Freeman and Co.
- [57] Gordon R. Bender R. Herman GT. (1970) "Algebraic Reconstruction Techniques (ART) for three dimensional electron microscopy and x-ray photography" J. Theor, Biol., 29, pp 471-481.

- [58] Gilbert P. (1972) "Iterative methods for the three-dimensional reconstruction of an object from projections" J. Theor., Biol. 36, pp 105-117.
- [59] Peterson J.E. Paulsson N.P. McEvelly T.V. (1985) "Applications of algebraic reconstruction techniques to crosshole seismic data" Geophysics., 50, 10, pp 1566-1580.
- [60] Lehmann B. (2007) "Seismic travel time tomography for engineering and exploration applications" EAGE Publications bv., pp 1-14.
- [61] Kaczmarz S. (1937) "Angenahert auflosung von systemen linearen gleichungen" Bull. Acad. Polon. Sci. Lett. A., pp 1-355.
- [62] Dines K.A. Lytle R.J. (1979) "Computerized geophysical tomography" Proc. IEEE., 67, pp 1065-1073.
- [63] Zhou H. (1997) "Determination of velocities and interfaces by multi-scale tomography" 67th Ann. Internat. Mtg: Society of Exploration Geophysicists., Expanded Abstracts, pp 1877-1880.
- [64] Zhou H. (2003) "Multi-scale travelttime tomography" Geophysics., 68, pp 1639-1649.
- [65] Shuweihdi F. (2009), PHD. Thesis, "Clustering and Classification with Shape Examples" Submitted in accordance with the requirements for the degree of Doctor of Philosophy, The University of Leeds, Department of Statistics.
- [66] Andrew R. Webb. QinetiQ Ltd. Malvern UK. (2002) "Statistical Pattern Recognition" Second Edition, John Wiley & Sons Ltd., The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England.
- [67] Fletcher R. (1988) "Practical Methods of Optimization" Wiley., New York.
- [68] Cristianini N. Shawe-Taylor J. (2000) "An Introduction to Support Vector Machines" Cambridge University Press., Cambridge.
- [69] Vapnik V.N. (1998) "Statistical Learning Theory" Wiley., New York.
- [70] Courant R. Hilbert D. (1959) "Methods of Mathematical Physics" Wiley., New York.

[۷۱] نودهی ثابت، الف، (۱۳۹۲)، پایان نامه کارشناسی ارشد: "خوشه‌بندی و نشانندن گراف روی

درخت"، علوم کامپیوتر، دانشگاه صنعتی شریف، دانشکده علوم ریاضی.

[۷۲] شریعت رضوی، ب، (۱۳۹۰)، پایان نامه کارشناسی ارشد: "روش‌های بخش‌بندی گراف‌ها و

کاربرد آن‌ها در تقطیع تصویر"، دانشگاه صنعتی شریف، دانشکده علوم ریاضی.

[۷۳] هداوندی، الف، (۱۳۸۸)، پایان نامه کارشناسی ارشد: "طراحی یک سیستم خبره فازی بر پایه

یک مدل ترکیبی هوش مصنوعی برای مدل‌سازی پیش‌بینی فروش"، دانشگاه صنعتی شریف،

دانشکده مهندسی صنایع.

[74] Selim S.Z. Ismail M.A. (1986) "On the local optimality of the fuzzy isodata clustering algorithm" IEEE Transactions on Pattern Analysis and Machine Intelligence., 8, 2, pp 284–288.

[75] Graves D. Pedrycz W. (2007) "Fuzzy C-Means, Gustafson-Kessel FCM, and Kernel-Based FCM: a comparative stud" Analysis and Design of Intelligent Systems Using Soft Computing Techniques., 41, pp 140–149.

[76] Gustafson E. E. Kessel W. C. (1979) "Fuzzy Clustering With A Fuzzy covariance Matrix" Proceedings of The IEEE Conference on Decision and Control, San Diego., California, pp 761–766.

[77] Verma D. Meila M. (2003) "A comparison of spectral cluster algorithms" University of Washington, Technical report UW CSE, 03-05-01.

[78] Desgraupes B. (2013) "Clustering Indices" University Paris Ouest., Lab Modal'X.

[79] Birch F. (1960) "The velocity of compressional waves in rocks to 10 kilobars" I: J. Geophys. Res., 65, pp 1083-1102.

[80] Robert S.C. (1982) "handbook of physical properties of rocks" CRC press., 2, pp 1-356.

[81] Kowallis B. Jones L. E. A. Wang H.F. (1984) "Velocityporosity-clay content; systematics of poorly consolidated sandstones" J. Geophys. Res., 89, pp 10355-10364.

[82] Costagna J. P. Batzle M. L. Eastwood R. L. (1985) "Relationship between compressional wave and shear wave velocities in clastic silicate rocks" Geophysics., SO, pp 551-570.

[83] Journel A.G. (1974) "Geostatistics for conditional simulation of ore bodies" Econ, Geol., 69, 5, pp 673–687.

- [84] Oliver D.S. (1995) "Moving averages for Gaussian simulation in two and three dimensions" Math. Geol., 27, 8, pp 939–960.
- [85] Le Ravalec-Dupin M. Noetinger B. Hu L. (2000) "The FFT moving average generator: an efficient numerical method for generating and conditioning Gaussian simulations" Math. Geol., 32, pp 701–722.
- [86] Bochner S. (1936) "Lectures on Fourier analysis" Princeton University Press, Princeton., NJ.
- [87] Priestley M.B. (1981) "Spectral analysis and time series" Academic Press., London, GB.
- [88] Oliver D.S. (2003) "Gaussian cosimulation: modelling of the cross-covariance" Math. Geol., 35, pp 681–698.

[۸۹] شرکت سهامی آب منطقه ای اصفهان، (۱۳۸۷)، "گزارش مکانیک سنگی شرکت سهامی آب

منطقه ای اصفهان"، خدمات مهندسی مرحله دوم، شماره گزارش، ۱۵۴۸۲، ۸۱.

Abstract

Transductive support vector machine spectral clustering (TSVMSC) inference algorithm considers spatial heterogeneity in all models, and is flexible and can be used with any type of data to identify the dominant patterns and structures in the multi-variable data. This algorithm has a stable performance and a high level of accuracy. The purpose of this thesis is to improve the integration of the results of inversion of refraction seismic tomography and resistivity data using the TSVMSC algorithm as a knowledge-based approach. For this purpose, the seismic refraction tomography and electrical resistivity data acquired from one of dams in Iran have been analyzed. To verify the modified model obtained as a result of integration of the data by the TSVMSC algorithm in a controlled implementation, we have used a synthetic example of a heterogeneous unconsolidated aquifer. After the inversion of synthetic and field data, for the integration of the results, k-means, FCM, Gustaffson-Kessel and TSVMSC algorithm have been used. Based on the results, the separation of alluvial regions, submerged bedrock and high quality bedrock has been made, and in this way, the results have been validated. Moreover, the results have shown that the TSVMSC algorithm has been more accurate and more effective in integrating heterogeneity and complexity of the two models. The Gustaffson-Kessel fuzzy cluster has also been more successful than other clustering methods used in this research. Based on the validation of the seismic refraction tomography data inversion results, the estimation error was calculated to be 30.95%. Based on the results of experiments, the experimental relationships for compressive and shear velocity were obtained, and consequently, correlation coefficients of 76.5 and 77.5 percent were obtained for compressive velocity and shear velocity with porosity, respectively.

Keywords: TSVMSC, Seismic refraction tomography, Electrical resistivity, Clustering.



Shahrood University of
Technology

Faculty of Mining, Petroleum and Geophysics Engineering
Ph.D. Thesis in Mineral Exploration

Using TSVMSC algorithm to Improve the Integration of Inversion Results of Refraction Seismic Tomography and Electrical Resistivity Data

By:

Bahman Mohammadi

Supervisor:

Dr. Abolghasem Kamkar Rouhani

September, 2018