

بهینه‌سازی تعیین رده رخساره‌های پتروفیزیکی با تحلیل سریع مؤلفه‌های مستقل و تبدیل کسینوسی گسسته پایه K نزدیک‌ترین همسایه در میدان مارون مخزن آسماری

سیدعلی هاشمی‌نژاد^۱ و کامیار احمدی^{۲*}

^۱ کارشناسی ارشد، گروه مهندسی نفت، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران

^۲ دکتر، گروه مهندسی نفت، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران

تاریخ دریافت: ۱۳۹۵/۰۹/۰۹ تاریخ پذیرش: ۱۳۹۶/۰۳/۰۲

چکیده

هدف از انجام این پژوهش، تعیین بهینه رخساره‌های پتروفیزیکی از روی داده‌های نموداری چاه است. با استفاده از روش دسته‌بندی خودکار K نزدیک‌ترین همسایه می‌توان رخساره‌های پتروفیزیکی را تعیین کرد؛ اما نه به صورت بهینه. برای تعیین بهینه رخساره‌ها، روش K نزدیک‌ترین همسایه با روش‌های تحلیل سریع مؤلفه‌های مستقل و تبدیل کسینوسی گسسته تلفیق می‌شود که این امر موجب افزایش نرخ موفقیت روش K نزدیک‌ترین همسایه و تعیین بهینه رخساره‌های پتروفیزیکی می‌شود که به دنبال آن می‌توان مدل‌سازی و توصیف مخازن هیدروکربنی را انجام داد. پژوهش در دو مسیر متفاوت انجام می‌شود. در مسیر اول، روش تحلیل سریع مؤلفه‌های مستقل روی داده اعمال و سپس توسط روش K نزدیک‌ترین همسایه دسته‌بندی می‌شود و در مسیر دوم روش‌های تحلیل سریع مؤلفه‌های مستقل و تبدیل کسینوسی گسسته روی داده‌ها اعمال و سپس توسط روش K نزدیک‌ترین همسایه دسته‌بندی می‌شود. در پایان نرخ موفقیت دسته‌بندی توسط K نزدیک‌ترین همسایه در هر دو مسیر به منظور تعیین بهینه رخساره‌های پتروفیزیکی مورد ارزیابی قرار می‌گیرد. ارزیابی انجام شده نشان می‌دهد که اعمال روش مسیر دوم روی داده‌ها به‌طور معنی‌داری نرخ موفقیت دسته‌بندی توسط روش K نزدیک‌ترین همسایه را افزایش می‌دهد که این امر موجب تعیین بهینه رخساره‌های پتروفیزیکی می‌شود که همان هدف از انجام این پژوهش است. داده‌های مورد استفاده، نمودارهای صوتی (Sonic Log)، پرتوی گاما (Gamma Ray)، چگالی (FDC یا RHOB)، تخلخل نوترون (CNL یا NPHI) و نگارهای القایی ژرف (ILD) از میدان نفتی مارون در جنوب کشور ایران هستند.

کلیدواژه‌ها: تبدیل کسینوسی گسسته، تحلیل سریع مؤلفه‌های مستقل، تعیین رخساره‌های پتروفیزیکی، K نزدیک‌ترین همسایه، هموارسازی.

* نویسنده مسئول: کامیار احمدی

E-mail: petroyaran@gmail.com

۱- پیش‌نوشتار

– Sanchett et al. (2013) با هموارسازی داده‌های نموداری چاه توسط روش‌های تحلیل سریع مؤلفه‌های مستقل و تبدیل کسینوسی گسسته و به کارگیری روش طبقه‌بندی خودکار K نزدیک‌ترین همسایه، موفق به تعیین بهتر رخساره‌ها در میدان نفتی Campos Basin در کشور برزیل شدند.

همه انواع داده‌ها می‌تواند به زیرمجموعه‌های بسیاری تفکیک شود که هر المان داده‌ای موجود در یک زیرمجموعه تصادفی، دارای اطلاعاتی در ارتباط با المان‌های دیگر موجود در آن زیرمجموعه خواهد بود. به عبارت دیگر دسته‌بندی همه المان‌ها بر پایه ویژگی‌های مشترک شناسایی شده در یک مجموعه داده مشخص، صورت می‌گیرد. این موضوع پایه دسته‌بندی خودکار است (MacQueen, 1967). روش‌های گوناگون برای دسته‌بندی خودکار به‌طور کلی به دو گروه اصلی تقسیم می‌شود:

۱-۱ روش‌های نظارت یافته

این روش‌ها نیازمند مشخص بودن برجسب‌ها یا اهداف برای دسته‌بندی است؛ یعنی مشخص است که یک نمونه ورودی متناظر با کدام برجسب مشخص است. در این مقاله این برجسب‌ها، داده‌های مغزه است. نمونه‌ای از این روش‌ها به قرار زیر هستند:

– روش k تعداد از نزدیک‌ترین همسایه‌ها: در انجام این پژوهش از این روش استفاده شده است (Toussaint, 2005). روش K نزدیک‌ترین همسایه یک روش تشخیص الگوست که المان‌های موجود در یک مجموعه داده را بر پایه توزیع مکانی یک مجموعه آموزش دهنده، دسته‌بندی می‌کند.

– شبکه‌های عصبی مصنوعی: روش‌های نظارت یافته هنگامی مؤثر خواهند بود که: الف) بتوان فرض کرد که محیط ناهمگن یا هتروژن به‌صورت متناوب است (برای ساده‌سازی ریاضیاتی) و ب) محاسبات مسائل محلی به جای یک سلول منفرد، باید بر پایه یک R.E.V بزرگ (حجم اولیه معرف) انجام شود؛ زیرا روند تصادفی نتایج متفاوتی نسبت به روند تناوبی دارد (Amaziane et al., 2006).

داده‌های نگار چاه به منظور ارزیابی‌های زمین‌شناسی و ژئوفیزیکی مانند توصیف‌سازی مخزن مورد استفاده قرار می‌گیرند (Dubrule, 1994; Avseth et al., 2005). همچنین از داده‌های چاه‌نگاری (Doyen, 2007; Coconi-Morales et al., 2010). برای پیش‌بینی متغیرهای لرزه‌ای وابسته به داده‌های دامنه در برابر دورافت (AVO) همچون V_p ، V_s و نسبت پواسون (σ) استفاده می‌شود که به درک مخزن کمک می‌کنند (Rutheford and Williams, 1989). داده‌های چاه‌نگاری در ایجاد نقشه ساختمانی و چینه‌ای به‌طور گسترده مورد استفاده قرار می‌گیرند.

همیشه‌سازی داده‌های چاه‌نگاری با رخساره‌های سنگی در نمونه مغزه به منظور ارتباط دادن این داده‌ها با دیگر اطلاعات زمین‌شناسی و ژئوفیزیکی اهمیت بسزایی دارد. اما نمونه‌های مغزه بیشتر به دلیل محدودیت‌های مالی در کل طول چاه در اختیار قرار نمی‌گیرند. بنابراین روش‌های تشخیص الگو به منظور پیشگویی و دسته‌بندی اعمال می‌شود. نمونه‌های محدودی از کاربرد روش‌های تشخیص الگو در داده‌های نمودار چاه به قرار زیر است:

– Grana et al. (2012) روش آماری کاملی برای دست‌یابی به اطلاعات پتروفیزیکی در محل چاه و دسته‌بندی رخساره‌های متناظر با آن را ارائه کردند.

– Messina and Langer (2011) الگوریتم‌های نظارت یافته بر پایه نقشه‌های خودسازمان‌دهی و تحلیل خوشه را برای تحلیل و تفسیر داده‌های لرزه‌آشفشانی به اجرا گذاشتند.

– Turlapaty et al. (2010) روشی مبتنی بر استخراج ویژگی بر پایه موجک و ماشین‌های برداری را با ساپورت یک کلاسه برای تحلیل داده‌های دریافت از دور ماهواره مورد استفاده در نقشه‌برداری رطوبت خاک و زندگی گیاهی به اجرا گذاشتند.

– Rosati and Cardarelli (1997) برای دسته‌بندی الگوهای دیده شده از نقشه‌های بی‌هنجاری مغناطیسی، ویژگی‌های بافت مبتنی بر ماتریس‌های وابسته مکان با تُن خاکستری را اعمال کردند.

الگوریتم نقطه ثابت بر پایه روش نیوتن است (Hyvärinen, 1999) که تحت قالب رایج تحلیل مؤلفه‌های مستقل به صورت زیر فراخوانده می‌شود:

$$S^+ = S - \frac{f(X)}{f'(X)} \quad (1)$$

s بردار حدس اولیه تصادفی غیر گوسی است که با یک روند تکرار و از راه تابع هدف f(x) به s⁺ همگرا می‌شود. f(x) تابع هدف است که به طور معمول با معادلات نکتروپی تکمیل می‌شود. نکتروپی که ضد آنتروپی است؛ در واقع اندازه‌ای است که با دورتر شدن توزیع داده‌ها از توزیع گوسی، مقدار آن بیشینه خواهد شد (Common, 1994). در زمینه تحلیل سریع مؤلفه‌های مستقل، نکتروپی با در نظر گرفتن شرایط ایستایی (رابطه زیر) مورد استفاده قرار می‌گیرد:

$$\nabla f(X^*) + \sum_{i=1}^m \mu_i \nabla g_i(X^*) + \int_{j=1}^l \lambda_j \nabla h_j(X^*) = 0 \quad (2)$$

x* یک مینیمم محلی است. μ_i و λ_j ضرایب ثابت KKT نامیده می‌شوند. g توابع نامساوی قیدی، h_j توابع تساوی قیدی و ∇ علامت گرادیان است. این توابع با رابطه زیر بهینه می‌شوند (Hyvärinen, 1999):

$$E\{Xg(S^T X)\} - \beta S = 0 \quad (3)$$

$$\beta = E\{S_0^T Xg(S_0^T X)\} \quad (4)$$

که مقدار بهینه از این فرایند است.

روش نیوتن نیاز به مشتق نکتروپی دارد؛ زیرا s یک بردار است بنابراین لازم است تا ژاکوبین نکتروپی به دست آید. اگر معادله نکتروپی با F نشان داده شود؛ ژاکوبین آن به صورت زیر است:

$$JF(S) = E\{XX^T g'(S^T X)\} - \beta I \quad (5)$$

$E\{XX^T\} \approx I$ ؛ زیرا در تحلیل آماری اعمال فرایند پردازش سفیدکننده در بردارها از راه تجزیه مقدار ویژه رایج است (Franklin, 1968). در نتیجه رابطه 5 به صورت زیر خواهد شد:

$$JF(S) = E\{g'(S^T X)\} - \beta I \quad (6)$$

با تقریب β برای S به جای مقادیر S_0 ، می‌توان روش نقطه ثابت را بر حسب شکل کامل آن به صورت زیر بازنویسی کرد:

$$S^+ = \frac{E\{Xg(S^T X)\} - \beta S}{E\{g'(S^T X)\} - \beta} \quad (7)$$

$S^* = S^+ / \|S^+\|$ ؛ بهنجارسازی این رابطه با $E\{g'(S^T X)\} - \beta$ سبب می‌شود تا در نهایت رابطه الگوریتم نقطه ثابت به صورت زیر تعریف می‌شود:

$$S^+ = E\{Xg(S^T X)\} - E\{g'(S^T X)\}S \quad (8)$$

پاسخ بالا یک مؤلفه مستقل را جستجو می‌کند و به سادگی با اجرای مورد لزوم همه مؤلفه‌های مستقل ممکن یافت می‌شود.

این روش تنها نیاز به دی‌کروله‌سازی مؤلفه‌های مستقل دارد (برای اجرای فرایند دی‌کروله‌سازی لازم است تا n+1 مؤلفه مستقل به دست آید). زیرا هر یک پروژکشن، از مؤلفه‌های پیشین یافته شده کم می‌شود.

طرح انقباضی مبتنی بر دی‌کروله‌سازی Gram-Schmidt به این تفریق‌ها اعمال می‌شود (Farina and Studer, 1984):

$$S_{n+1} = S_{n+1} - \sum_{i=1}^n S_{n+1}^T S_i S_i \quad (9)$$

با در نظر گرفتن اینکه داده‌ها با فرایند سفیدکنندگی، پیش پردازش می‌شوند؛ نرمال‌سازی دوباره پاسخ کافی خواهد بود تا پاسخ $\|S_{n+1}\|$ به دست آید.

۳- تبدیل کسینوسی گسسته

تبدیل کسینوسی گسسته روشی است که بر پایه آن یک تابع بر حسب جمع وزنی کسینوس‌هایی بیان می‌شود که در بسامدهای مختلف نوسان می‌کنند. این روش به تبدیل فوریه گسسته بسیار وابسته است (Oppenheim et al., 2009).

برای یک توالی از N سیگنال $x_0, x_1, \dots, x_{(N-1)}$ مختلط توالی تبدیل گسسته فوریه‌های مربوطه آنها یعنی $F_0, F_1, \dots, F_{(N-1)}$ به صورت زیر تعریف شود (Blinn, 1993):

۲- روش‌های نظارت نیافته (Mitchell, 1997; Schuerman, 1996; Hart and Duda, 1973)

این روش‌ها، نمونه‌ها را بر پایه روش‌های آماری یا ریاضی به زیرمجموعه‌های مختلف خوشه‌بندی می‌کنند مانند:

- روش K-Means

- روش Self-Organizing Maps

متغیرهای زیادی روی کیفیت پیشگویی تأثیرگذار است مانند: انتخاب متغیرها، کیفیت داده‌های اندازه‌گیری شده، عدم قطعیت داده‌ها.

برای دسته‌بندی کننده فضایی مانند K نزدیک‌ترین همسایه، انتخاب مناسب، مجموعه‌ای از متغیرهای متعامد است که بدون هر گونه اطلاعات اضافی و نابه‌جا باشد. اما بیشتر در عمل به همبستگی زیاد میان نیمرخ‌های مختلف نگار چاه توجه می‌شود که اگر این همبستگی‌ها زیاد باشد به معنای وجود اطلاعات اضافی است که ممکن است اجرای روش K نزدیک‌ترین همسایه را به خطر اندازد (Carrasquilla and Leite, 2009).

برای کم کردن این مشکل و آماده‌سازی داده‌ها برای دسته‌بندی خودکار، یک تحلیل چندمتغیره به نام تحلیل مؤلفه مستقل اعمال می‌شود (Comon, 1994). روش تحلیل مؤلفه مستقل یک روش تفکیک منشأ کور است که سیگنال ورودی را به مجموعه جدیدی از سیگنال‌های مستقل تبدیل می‌کند. از لحاظ آماری مستقل بودن به معنای عدم ارتباط رخداد یک سیگنال با رخداد سیگنال دیگر موجود در مجموعه داده‌هاست (Russell and Norvig, 2002). استقلال سیگنال‌ها مجموعه متعامدی از متغیرهای جدید را ایجاد می‌کند که این متغیرها می‌توانند به عنوان رویدادهای منحصر به فرد در نظر گرفته شوند.

به منظور کاهش مقدار محاسبات، یک الگوریتم تکرار مبتنی بر روش نیوتن توسط Hyvärinen (1999) گسترش یافته است که تحلیل سریع مؤلفه‌های مستقل یا تحلیل سریع مؤلفه مستقل نامیده می‌شود. متغیرهای مستقل لزوماً متعامد هستند؛ بنابراین روش تحلیل سریع مؤلفه‌های مستقل خاصیت تعامدی را در دسته‌بندی مهیا می‌سازد. همچنین به منظور ایجاد داده‌های هموار، تبدیل کسینوسی گسسته اعمال می‌شود (Rao and Yip, 1990).

۲- تحلیل سریع مؤلفه مستقل

روش تحلیل مؤلفه‌های مستقل روشی برای یافتن عوامل یا مؤلفه‌های زیربنایی از داده‌های چندمتغیره‌ای وابسته به تفکیک سیگنال است (Hyvärinen et al., 2001). این روش محدودیت‌هایی مانند بیشینه‌سازی توابع تباین، هزینه محاسباتی، انتخاب نرخ فراگیری دارد.

الگوریتم‌های متفاوتی برای روش تحلیل مؤلفه‌های مستقل وجود دارد که یکی از پر استفاده‌ترین آنها در نرم‌افزارهای صنعتی الگوریتم تحلیل سریع مؤلفه‌های مستقل است، این الگوریتم توسط Hyvärinen (1999) در دانشگاه تکنولوژی هلسینکی ابداع شده است.

روش تحلیل سریع مؤلفه‌های مستقل در واقع یک روش بهینه‌سازی شده از روش تحلیل مؤلفه‌های مستقل بوده و برای حل این محدودیت‌ها ارائه شده است. همگرا شدن به نتایج در این روش نیز، سریع‌تر از روش تحلیل مؤلفه‌های مستقل است (Hyvärine, 1999).

روش تحلیل سریع مؤلفه‌های مستقل مبتنی بر یک الگوریتم نقطه ثابت است که سرعت عملکرد شناخته شده‌ای دارد. چنین سرعتی قابل توجه خواهد بود؛ با دانستن اینکه تعداد زیادی از نقاط داده در یک گام از این الگوریتم به کار برده می‌شوند. این الگوریتم در مقایسه با الگوریتم‌های رایج نقطه ثابت، اصلاحاتی یافته است (Burden and Faires, 1985). همچنین این الگوریتم مشابه با برخی الگوریتم‌های عصبی، موازی و از دید محاسباتی ساده است و نیاز به حافظه محاسباتی کمتری دارد (Hyvärinen et al., 2001).

زیادی مؤثر است اما حافظه زیاد نیاز دارد و با پیچیدگی محاسباتی نیز همراه است (Bhatia and Vandana, 2010).

ورودی‌های روش K نزدیک‌ترین همسایه یک مجموعه آموزشی و یک مجموعه آزمون است.

مجموعه آموزشی از n زوج $(x_{1i}, \theta_i), (x_{2i}, \theta_i), \dots, (x_{ni}, \theta_i)$ تشکیل یافته است که x_{ni} مقادیر یک متغیر برحسب یک فضای متریک و θ_i دسته‌های مختلف برحسب‌ها برای هر جفت هستند. هر θ_i در مجموعه $\{1, 2, 3, \dots, M\}$ جای می‌گیرد که M تعداد کل برحسب‌ها را نشان می‌دهد (Cover and Hart, 1967).

مجموعه آزمون n' المان $x'_{1i}, x'_{2i}, \dots, x'_{ni}$ دارد و هدف دسته‌بندی هر یک از المان‌ها و به دست آوردن n زوج جدید به صورت $(x'_{1i}, \theta_i), (x'_{2i}, \theta_i), \dots, (x'_{ni}, \theta_i)$ هست.

در اینجا باید مقدار K که معرف تعداد نزدیک‌ترین همسایه‌هاست؛ انتخاب شود. این انتخاب روی تعداد نمونه‌هایی که برای برحسب‌گذاری مجموعه آزمون در نظر گرفته می‌شود، تأثیر می‌گذارد. در این پژوهش انتخاب k مبتنی بر بهترین نرخ موفقیت یافته شده در آزمون مقدماتی است که می‌توان در مرحله ۳ آن را دید.

می‌توان روش K نزدیک‌ترین همسایه را در چند گام خلاصه کرد؛ (۱) سازمان‌دهی یک مجموعه آموزشی با مقادیر متغیر x_j و برحسب‌های مربوطه θ_j ؛ (۲) سازمان‌دهی مجموعه آزمون با مقادیر متغیر x'_j و بدون هیچ‌گونه برحسبی؛ (۳) در نظر گرفتن یک مقدار برای k؛ (۴) تکرار این مراحل برای هر x'_j (الف) از یک تابع متریک

d برای تعیین نزدیک‌ترین نمونه از x'_j در مجموعه آموزشی استفاده شود؛ یعنی

$d = \min(x'_j, x_i) = \min(d(x'_j, x_i))$ برای این پژوهش به صورت خاص به منظور ساده‌سازی فاصله اقلیدسی

$d_E(x'_j, x_i) = \sqrt{X_j^2 - (X'_j)^2}$ انتخاب شد؛ (ب) x_j را از مجموعه آموزشی حذف و این مرحله تا حذف k نمونه از مجموعه آموزشی تکرار شود؛ (پ) برحسب‌های θ_i از

نمونه‌های حذف شده تحلیل و برحسب‌های مکرر بیشتری به x'_j تخصیص داده شود؛ (۵) n' زوج، $(x'_{1i}, \theta_i), (x'_{2i}, \theta_i), \dots, (x'_{ni}, \theta_i)$ را بر پایه روند بالا ایجاد شود.

۵- روش‌شناسی

داده‌های مورد استفاده در پژوهش شامل ۱۹۷۴ نمونه مغزه (جدول ۱) به همراه نگارهای آنهاست که از چهار حلقه چاه در میدان نفتی مارون در جنوب ایران مورد تحلیل قرار گرفت که توسط شرکت ملی مناطق نفت خیز جنوب تهیه شده و اسامی این نگارها به این قرار است: (۱) صوتی (DT)؛ (۲) پرتوگاما (GR)؛ (۳) مقاومت (ILD)؛ (۴) چگالی (RHOB)؛ (۵) تخلخل نوترون (NPHI).

روند انجام پژوهش در دو مسیر متفاوت به شکل زیر صورت گرفت و در پایان نتایج به دست آمده از دو مسیر مقایسه شد.

در مسیر اول روی داده‌های نگار چاه که نمونه مغزه برای آنها موجود است؛ روش کاهش بعد تحلیل سریع مؤلفه‌های مستقل اعمال و خروجی آن با استفاده از روش K نزدیک‌ترین همسایه دسته‌بندی می‌شود و میزان موفقیت به دست می‌آید.

در مسیر دوم برای همان داده‌های اولیه به کار رفته شده در مسیر اول، روش کاهش بعد تحلیل سریع مؤلفه‌های مستقل اعمال، سپس روی خروجی آن تبدیل کسینوسی گسسته به عنوان یک روش هموارساز (Simonoff, 1996) به اجرا گذاشته و خروجی پایانی با استفاده از روش K نزدیک‌ترین همسایه دسته‌بندی می‌شود و میزان نرخ موفقیت به دست می‌آید.

بهترین میزان موفقیت زمانی حاصل می‌شود که اطلاعات کمی از داده‌های اولیه به کار رفته شده باشد (در حدود ۱٪) که به تولید سیگنالی هموار خواهد انجامید که تنها رفتار سیگنال اولیه را در بر دارد. این نتیجه با فرض فاصله‌ای روش K نزدیک‌ترین همسایه سازگار است؛ زیرا نیمرخ‌های اولیه چاه‌نگاری دامنه‌های بسیار متفاوتی را برای ترکیب‌سنگی‌های یکسان (یکسان بودن ترکیب‌سنگی‌ها توسط مطالعات شرکت ملی نفت مناطق نفت خیز جنوب روی مقاطع نازک مغزه‌ها به اثبات رسیده است) ایجاد می‌کنند؛ این امر عملکرد دسته‌بندی خودکار را بر هم می‌زند.

$$F_k(\omega) = \sum_{n=0}^{N-1} x_n \left[\cos\left(-\frac{2\pi}{N}kn\right) + i \sin\left(-\frac{2\pi}{N}kn\right) \right] \quad (10)$$

که $k=1, 2, \dots, N-1$ است.

از سوی دیگر تنها می‌توان با یکی از بخش‌های حقیقی تبدیل گسسته فوریه (که این بخش با تابع کسینوس کنترل می‌شود) و یا بخش موهومی آن (که با تابع سینوسی کنترل می‌شود) کار کرد. این انتخاب سبب دو تبدیل کسینوسی گسسته و تبدیل سینوسی گسسته می‌شود که شبیه و وابسته به تبدیل گسسته فوریه هستند.

روش تبدیل کسینوسی گسسته برای تراکم‌سازی داده‌ها یا حذف و تضعیف نوفه است (Rao and Yip, 1990) در حالی که تبدیل سینوسی گسسته اساساً برای حل معادلات دیفرانسیلی پاره‌ای اعمال می‌شود (Martucci, 1994). در پردازش سیگنال و تصویر پرکاربردترین شیوه محاسبه تبدیل کسینوسی گسسته به دلیل قدرت تراکم ویژه انرژی و داده‌ها روش تبدیل کسینوسی گسسته-II هست (Rao and Yip, 1990) که به صورت رابطه زیر خلاصه می‌شود:

$$DCT - \prod_k(\omega) = \sum_{n=0}^{N-1} x_n \cos\left(\frac{\pi}{N}(n + \frac{1}{2})k\right) \quad (11)$$

کاربردهای تبدیل کسینوسی گسسته چندبعدی به مقدار زیادی از تعریف مشابه روش یک بعدی پیروی می‌کند و به سادگی با ضرب مجزای همه کسینوس‌ها در طول هر بعد به اجرا گذاشته می‌شود. برای نمونه اگر داده‌ها به صورت یک ماتریس باشند؛ اجرای روش تبدیل کسینوسی گسسته با اعمال رابطه زیر داده‌ها را در دو بعد

$$DCT_{k_1, k_2}(\omega) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x_{n_1, n_2} \cos\left(\frac{\pi}{N_1}(n_1 + \frac{1}{2})k_1\right) \cos\left(\frac{\pi}{N_2}(n_2 + \frac{1}{2})k_2\right) \quad (12)$$

که n_1, N_1, K_1 مربوط به اولین بعد و n_2, N_2, K_2 متناظر با بعد دوم است.

این ضرب‌ها می‌تواند برای محاسبات تبدیل کسینوسی گسسته‌های m بعدی به صورت زیر بر روی یابی شوند:

$$DCT_{k_1, k_2, \dots, k_m}(\omega) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \dots \sum_{n_m=0}^{N_m-1} x_{n_1, n_2, \dots, n_m} \prod_{j=1}^m \cos\left(\frac{\pi}{N_j}(n_j + \frac{1}{2})k_j\right) \quad (13)$$

اگر n تعداد‌های یک نگار چاه‌پیمایی و m تعداد نمونه‌های موجود در هر نیمرخ باشد؛ می‌توان داده‌های چنین موردی را به صورت یک ماتریس $A_{m \times n}$ نشان داد. با

اعمال تبدیل کسینوسی گسسته روی این ماتریس، ماتریس جدیدی ایجاد خواهد شد که همان‌طور که در شکل ۱ دیده می‌شود؛ بیشتر اطلاعات در پی این تبدیل جملات اول برای $n \geq 1$ قرار خواهند گرفت (یعنی جملات بعدی دارای دامنه ناچیز و متمایل به صفر هستند). دامنه اولین جملات ماتریس تبدیل کسینوسی گسسته بیشتر از جملات آخر است. جملات آخر دارای دامنه‌های مهمی نیستند و در پردازش مانند

یک آستانه‌گذاری، مقدار این دامنه‌ها به صفر تبدیل می‌شود (Battiatto et al., 2001).

توجه به این نکته اهمیت دارد که جملات آخر خیلی به صفر نزدیک هستند؛ اما این مقادیر کوچک هنوز می‌تواند روی ماشین‌های آموزش غیر نظارت یافته مانند تحلیل مؤلفه‌های اصلی (Abdi and William, 2010)، تحلیل مؤلفه‌های مستقل یا برخی از شبکه‌های عصبی مصنوعی مانند نقشه‌های خود سازمان یافته (Liu et al., 2006) تأثیر گذار باشد.

با صفر قرار دادن جملات آخر، تبدیل کسینوسی وارون (۱-تبدیل کسینوسی گسسته)، داده‌ها را به صورت زیر به حوزه اولیه خود بر می‌گرداند؛ یعنی:

$$TheIDCT_{n_1, n_2}(t) : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \text{is given by} \quad (14)$$

$$IDCT_{n_1, n_2}(t) = \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} a(k_1)a(k_2)x_{k_1, k_2} \cos\left(\frac{\pi}{N_1}(n_1 + \frac{1}{2})k_1\right) \cos\left(\frac{\pi}{N_2}(n_2 + \frac{1}{2})k_2\right)$$

که:

$$a(k_m) = \begin{cases} \frac{1}{\sqrt{N}}, & SEK = 0 \\ \frac{2}{\sqrt{N}}, & sek = 1, \dots, N-1 \end{cases} \quad (15)$$

۴- k نزدیک‌ترین همسایه

روش K نزدیک‌ترین همسایه یک قاعده تصمیم‌گیری برای دسته‌بندی نمونه‌هاست که نمونه‌ها را توسط نزدیک‌ترین آنها و نمونه‌های دسته‌بندی شده پیشین، دسته‌بندی می‌کند (Cover and Hart, 1967). این روش بسیار ساده و به میزان

دارد. این نتیجه با فرض فاصله‌های روش K نزدیک‌ترین همسایه سازگار است؛ زیرا نیمرخ‌های اولیه چانه‌نگاری دامنه‌های بسیار متفاوتی را برای ترکیب‌سنگی یکسان ایجاد می‌کنند و بنابراین عملکرد دسته‌بندی خودکار را بر هم می‌زند. به همین دلیل میزان موفقیت دسته‌بندی توسط روش K نزدیک‌ترین همسایه در مسیر دوم به شکل معناداری افزایش پیدا می‌کند و این امر به تعیین بهینه رخساره‌های پتروفیزیکی خواهد انجامید. شکل ۳ میزان موفقیت دسته‌بندی K نزدیک‌ترین همسایه را در دو حالت مختلف نشان می‌دهد.

با در نظر گرفتن شکل ۳ دیده می‌شود که اتحاد روش‌های تبدیل کسینوسی گسسته و تحلیل سریع مؤلفه‌های مستقل بهترین نرخ پیش‌گویی و سازگارترین نتایج را ایجاد می‌کند و تأثیر روش تبدیل کسینوسی گسسته روی دسته‌بندی خودکار از طریق پردازش سیگنال بسیار بیشتر از روش تحلیل سریع مؤلفه‌های مستقل است. این اظهار منطقی است؛ زیرا تحلیل چندمتغیره مانند تحلیل سریع مؤلفه‌های مستقل عموماً تفسیر مکانی داده‌ها را بهبود می‌بخشد. بنابراین کاربری منفرد روش‌های تحلیل چندمتغیره برای دسته‌بندی کننده‌های خودکار انتخاب ایده آلی نیست. برای درک بهتر تفاوت دو روش بالا و چگونگی تأثیر هر یک روی روش دسته‌بندی خودکار K نزدیک‌ترین همسایه، تأثیر هر یک از روش‌های تبدیل کسینوسی گسسته و تحلیل سریع مؤلفه‌های مستقل روی سیگنال به وجود آمده از مقادیر نگار گاما (SGR) که نمونه‌ی مغزه آن موجود است؛ در ادامه شرح داده می‌شود. سیگنال به وجود آمده از مقادیر لاگ SGR به عنوان نمونه انتخاب و نتایج پردازش‌ها روی آن به نمایندگی از همه لاگ‌ها به نمایش گذاشته شده است. در واقع نتایج نمایش داده شده روی سیگنال به وجود آمده از مقادیر لاگ SGR در سیگنال به وجود آمده از دیگر لاگ‌ها نیز دقیقاً صادق است و همه لاگ‌های نام برده شده همین روند پردازشی را طی می‌کنند.

شکل ۱۴ اثر کاربرد تبدیل کسینوسی گسسته را روی سیگنال پرتوی گاما (SGR) در چهار سطح از اطلاعات به کار رفته شده (۷۰٪)، (۵۰٪)، (۲۰٪) و (۱۰٪) را نشان می‌دهد. می‌توان کاهش نوفه و اطلاعات را در چهار سطح مختلف دید؛ اما هدف افزایش نرخ موفقیت دسته‌بندی خودکار است. برای این منظور بهترین انتخاب در ۱۰٪ اطلاعات برای تبدیل کسینوسی گسسته است. برای افزایش نرخ پیش‌گویی صحیح تنها نیاز است تا رفتار عمومی سیگنال حفظ شود. همچنین به این دلیل که سنگ‌هایی با ترکیب‌سنگی یکسان که یکسان بودن ترکیب‌سنگی آنها توسط مطالعات شرکت ملی نفت مناطق نفت خیز جنوب روی مقاطع نازک مغزه‌های موجود آنها به اثبات رسیده است؛ دارای مقادیر نمودار لاگ متفاوتی هستند که این امر به کاهش دقت و بازده در دسته‌بندی خودکار می‌انجامد؛ زیرا این روش‌ها اطلاعات مختص و خاص را درون فضایی بعدی بر پایه فاصله در نظر می‌گیرند و هر چه اختلاف مقادیر بیشتر باشد سبب ایجاد فاصله بیشتر برای مقادیر یک کلاس در فضای بعدی می‌شود و دسته‌بندی این مقادیر با فواصل بیشتر سبب کاهش دقت و بازده در دسته‌بندی خودکار خواهد شد. برای نمونه جدول ۲ برخی از مقادیر نگار SGR را برای نمونه مغزه‌هایی با ترکیب‌سنگی یکسان نشان می‌دهد.

این مقادیر بسیار متمایز برای ترکیب‌سنگی یکسان در جدول ۲ سبب کاهش دقت دسته‌بندی خودکار k نزدیک‌ترین همسایه و دیگر دسته‌بندی کننده‌های خودکار می‌شود. شکل ۵ سیگنال اولیه به وجود آمده از ۳۶ نمونه مغزه با ترکیب‌سنگی یکسان و مقادیر نگار SGR متفاوت و سیگنال به وجود آمده از اعمال تبدیل کسینوسی گسسته روی آن (سیگنال اولیه از مقادیر SGR) را نمایش می‌دهد.

روش تبدیل کسینوسی گسسته با هموارسازی و کاهش قابل قبول اختلاف مقادیر مختلف نگارها برای ترکیب‌سنگی‌های یکسان، روشی کارآمد برای افزایش میزان موفقیت دسته‌بندی خودکار K نزدیک‌ترین همسایه به شمار می‌رود. شکل ۶ سیگنال اولیه به وجود آمده از مقادیر نگار (SGR) برای مغزه‌های موجود و سیگنال به وجود آمده از اعمال تحلیل سریع مؤلفه‌های مستقل روی آن (سیگنال اولیه از مقادیر SGR) را نمایش می‌دهد.

همه آزمون‌ها روی نرم‌افزار متلب (Matlab R2013a) به اجرا گذاشته شده و جعبه‌ابزار تحلیل سریع مؤلفه‌های مستقل ۲،۵ به نشانی زیر مورد استفاده قرار گرفته است:

http://bsp.teitb.gr/members/downloads/fastica/FastICA_2.5.zip

برای اثبات عملکرد صحیح و نشان دادن میزان دقت افزایش یافته در این روش از مجموعه نمونه مغزه‌ها با ۱۹۷۴ عضو - که ترکیب‌سنگی همه نمونه مغزه‌ها توسط مطالعات شرکت ملی نفت مناطق نفت خیز جنوب روی مقاطع نازک مغزه‌ها مشخص شده است - در یک تقسیم تصادفی ۹۸۷ نمونه مغزه به عنوان یک مجموعه آزمون انتخاب می‌شود. پس از آموزش دیدن الگوریتم‌های بیان شده با ۱۹۷۴ نمونه مغزه که بر پایه ۱۲ نوع ترکیب‌سنگی مختلف حاصل از مطالعات مقاطع نازک برچسب گذاری شده‌اند؛ ترکیب‌سنگی‌های مجموعه آزمون با ۹۸۷ نمونه مغزه توسط الگوریتم‌ها پیش‌بینی و سپس ترکیب‌سنگی پیش‌بینی شده مجموعه آزمون با ترکیب‌سنگی‌های حقیقی به دست آمده از مطالعات مقاطع نازک روی مغزه‌ها قیاس می‌شوند؛ بر این اساس، درستی سنجی و میزان دقت افزایش یافته در هر یک از ۴ حالت مختلف در شکل ۳، مشخص و اثبات می‌شود. میزان موفقیت بر پایه نسبت تعداد ترکیب‌سنگی‌های پیش‌بینی شده درست به تعداد ترکیب‌سنگی‌های مجموعه آزمون بیان می‌شود. این روند بیان شده در هر گام به صورت تصادفی ۵۰ بار به اجرا گذاشته می‌شود و میانگین میزان پیش‌گویی، دامنه‌ها و مؤلفه‌های مستقل به دست می‌آید تا میزان موفقیت در پیش‌گویی نمونه مغزه‌ها و درستی سنجی الگوریتم‌ها به صورت هر چه دقیق‌تر به دست آید. توجه شود که همه مقادیر نشان داده شده در شکل‌ها، مقادیر میانگین است.

با اجرای رابطه ۹ برای همه مؤلفه‌های مستقل در داده‌ها، تحلیل سریع مؤلفه‌های مستقل به صورت ترکیب زیرخطی سازی می‌شود (Hyvärinen, 1999):

$$S = WX \quad (16)$$

که در آن، X داده‌های اولیه ثبت شده، W ماتریس ترکیب نیافته (un-mixture matrix) و S مؤلفه‌های مستقل هستند. وجود وارون ماتریس W توسط تقریب‌های مناسب در فرایند تحلیل سریع مؤلفه‌های مستقل تضمین می‌شود (Hyvärinen, 1999).

کاهش بعد در تحلیل سریع مؤلفه‌های مستقل با به کارگیری کمی از مؤلفه‌های مستقل برای بازسازی سیگنال حاصل می‌شود. برای نمونه برای بعد $d < 5$ می‌توان داده‌های اولیه را برای ۱۹۷۴ نمونه با استفاده از رابطه زیر بازسازی کرد:

$$S_{(d \times 1974)} = W_{(d \times 5)} \cdot X_{(5 \times 1974)} \quad (17)$$

خروجی در دسته‌بندی کننده K نزدیک‌ترین همسایه استفاده شد. در حقیقت این اولین مسیر این پژوهش است.

گام بعدی که هدف اصلی است؛ اجرای مشترک و واحد تحلیل سریع مؤلفه‌های مستقل و تبدیل کسینوسی گسسته برای آماده‌سازی داده‌ها برای دسته‌بندی توسط K نزدیک‌ترین همسایه است. در این حالت با اعمال تحلیل سریع مؤلفه‌های مستقل روی داده‌ها، خروجی آن به شکل ماتریسی $S'_{1974 \times 5}$ ایجاد شد؛ سپس با اعمال تبدیل کسینوسی گسسته روی این ماتریس، ماتریس $S'_{det1974 \times 5}$ ایجاد می‌شود که تعداد کمی از عضوهای این ماتریس ($S'_{det}(i,j)$) دامنه‌های قابل توجه‌ای دارند و در بلوک بالای سمت راست ماتریس قرار گرفته‌اند؛ دیگر اعضا دامنه‌هایی به نسبت کم دارند و در برخی مواقع متمایل به صفر هستند.

با استفاده از یک آستانه گذاری تجربی، عضوهای با مقادیر نزدیک به صفر برابر صفر قرار داده شد و در ادامه I- تبدیل کسینوسی گسسته برای بازگشت به حوزه تحلیل سریع مؤلفه‌های مستقل به اجرا گذاشته شد که بسته به انتخاب آستانه بازسازی داده‌ها این مرحله به عنوان یک مرحله هموارساز در نظر گرفته می‌شود و به ایجاد ماتریس هموار S_{sm} انجامید و در پایان روی این ماتریس دسته‌بندی کننده K نزدیک‌ترین همسایه اعمال شد.

بهترین میزان موفقیت دسته‌بندی توسط روش K نزدیک‌ترین همسایه زمانی حاصل می‌شود که اطلاعات کمی از داده‌های اولیه به کار رفته شده باشد (در حدود ۱٪) که به تولید سیگنالی هموار خواهد انجامید که تنها رفتار سیگنال اولیه را در بر

تعداد مؤلفه‌های مستقل ۱ تا ۵ به صورت مجزا انجام شد که نرخ موفقیت دسته‌بندی K نزدیک‌ترین همسایه در حالت‌های مختلف در جدول ۴ درج شده است. بر پایه مقایسه نتایج به دست آمده از جدول ۴، بالاترین نرخ موفقیت دسته‌بندی خودکار هنگامی رخ می‌دهد که با استفاده از ۳ مؤلفه مستقل صورت پذیرد. باید به این نکته توجه شود که نوع و ترتیب نمودارهای چاهی می‌تواند در محاسبات دو فرضیه مطرح شده تأثیرگذار باشد و عامل ایجاد تفاوت میان نتایج به دست آمده در این پژوهش با نتایج به دست آمده از پژوهش‌های دیگر شود.

۶- نتیجه‌گیری

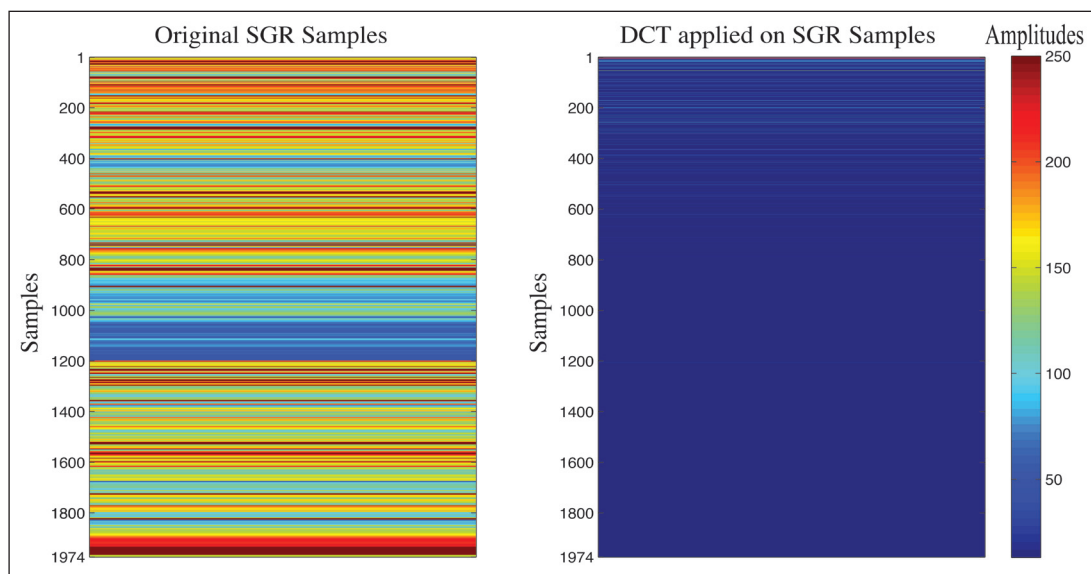
روش تبدیل کسینوسی گسسته هنگامی که به منظور کاهش نوفه در داده‌های نموداری چاه به کار برده می‌شود؛ انعطاف‌پذیری بیشتری نسبت به تحلیل سریع مؤلفه‌های مستقل دارد. این اختلاف در آستانه بهینه برای حذف نوفه قرار دارد. هنگامی که روش تحلیل سریع مؤلفه‌های مستقل برای کاهش بعد اعمال می‌شود؛ این عمل با کاهش بعد صورت می‌گیرد و به اجرا گذاشته می‌شود. به عبارت دیگر هنگامی که تبدیل کسینوسی گسسته برای کاهش نوفه اعمال می‌شود؛ انتخاب بهینه توسط یک آستانه تجربی تعیین می‌شود و بهترین نرخ موفقیت دسته‌بندی خودکار توسط K نزدیک‌ترین همسایه به دست می‌آید که عملکردش را کنترل می‌کند. برای K نزدیک‌ترین همسایه فرایند کاهش بعد نرخ موفقیت دسته‌بندی خودکار را تنها در حد ۲٪ بهبود می‌بخشد. نرخ درست دسته‌بندی خودکار K نزدیک‌ترین همسایه با استفاده از تحلیل سریع مؤلفه‌های مستقل و تبدیل کسینوسی گسسته بیشینه می‌شود که بهترین نتایج را در حدود ۱۲٪ و بدترین نتایج را در حدود ۲۰٪ بهبود می‌بخشد. بهترین نتایج در یک سیگنال بسیار هموار ایجاد می‌شود؛ احتمالاً به دلیل اینکه اختلاف دامنه هموار می‌شود و نقاطی را که متناظر با ترکیب سنگی یکسان است؛ گروه‌بندی می‌کند. مطالعات بعدی می‌تواند برای کنترل اثر واریوگرام‌های سیگنال در کاربرد این روش به اجرا گذاشته شود.

نمونه B در شکل ۶ برخی از اثرات تصنعی را نشان می‌دهد که در اعمال روش تحلیل سریع مؤلفه‌های مستقل ایجاد شده است. نمونه P در شکل ۶ اختلاف دامنه‌ای کم میان سیگنال اولیه و سیگنال به وجود آمده از اعمال تحلیل سریع مؤلفه‌های مستقل را نشان می‌دهد. نمونه T در شکل ۶ نشان می‌دهد که مقدار زیادی از دامنه‌ها در سیگنال به وجود آمده از اعمال تحلیل سریع مؤلفه‌های مستقل حذف شده‌اند که این امر سبب ایجاد یک اختلاف دامنه‌ای کامل با سیگنال اولیه شده است. این جدایش میان سیگنال‌ها نشان می‌دهد که اعمال روش تحلیل سریع مؤلفه‌های مستقل با حذف بسامدهای بالا و نیز تغییر سطوح بسامدی مختلف همراه است. همان‌گونه که دیده می‌شود؛ کاهش نوفه با استفاده از روش تحلیل سریع مؤلفه‌های مستقل ساده و کارآمد است؛ اما در افزایش نرخ پیشگویی دسته‌بندی خودکار مؤثر نیست.

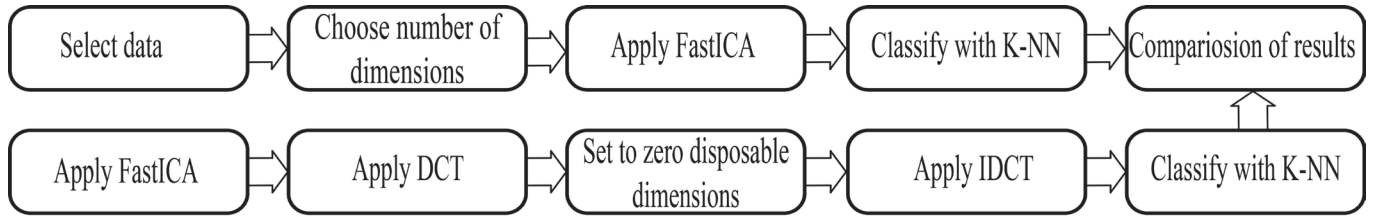
در این پژوهش می‌توان فرضیه‌ای را در نظر گرفت که تقدم اعمال روش‌های تحلیل سریع مؤلفه‌های مستقل و تبدیل کسینوسی گسسته چه مقدار می‌تواند روی میزان نرخ موفقیت دسته‌بندی با استفاده از روش K نزدیک‌ترین همسایه تأثیرگذار باشد. برای تحلیل اثر چنین فرضیه‌ای ۲ عملکرد دسته‌بندی خودکاری که در شکل ۳ نشان داده شده است؛ برای ۳ حالت مورد آزمون قرار گرفت: I) هنگامی که تحلیل سریع مؤلفه‌های مستقل پیش از تبدیل کسینوسی گسسته اعمال شد؛ II) هنگامی که تحلیل سریع مؤلفه‌های مستقل هنگام تبدیل کسینوسی گسسته اعمال شد؛ III) هنگامی که تحلیل سریع مؤلفه‌های مستقل پس از تبدیل کسینوسی گسسته اعمال شد. جدول ۳ نرخ موفقیت را برای هر ۲ عملکرد در ۳ حالت I، II و III نشان می‌دهد.

جدول ۳ نشان می‌دهد که تقدم عملکرد روش‌های تبدیل کسینوسی گسسته و تحلیل سریع مؤلفه‌های مستقل روی نرخ موفقیت دسته‌بندی خودکار K نزدیک‌ترین همسایه تأثیرگذار است و در حالت II بهتر عمل می‌کند.

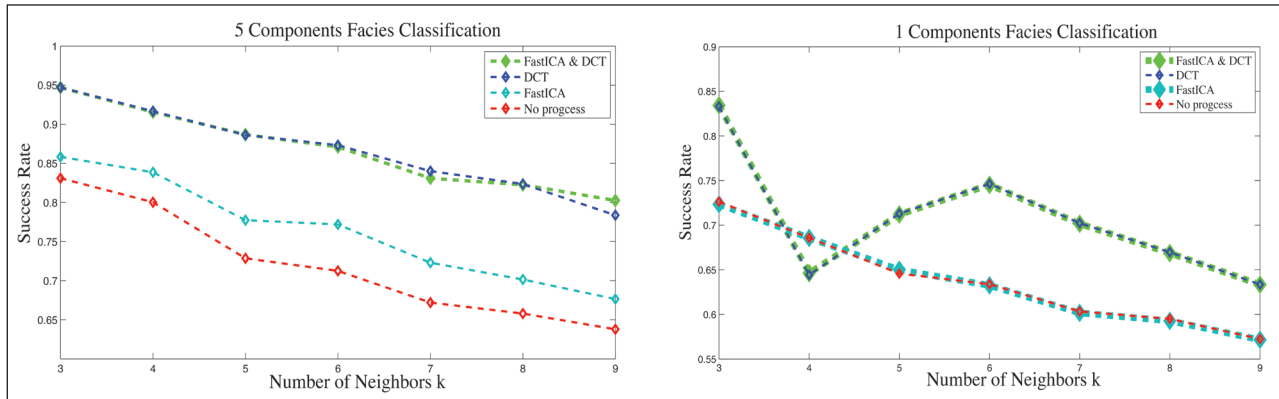
همچنین می‌توان فرضیه دیگری در این پروژه در نظر گرفت که تعداد مؤلفه‌های مستقل به چه میزان می‌تواند روی دسته‌بندی خودکار K نزدیک‌ترین همسایه تأثیر بگذارد. برای تحلیل اثر این فرضیه دسته‌بندی خودکار K نزدیک‌ترین همسایه با



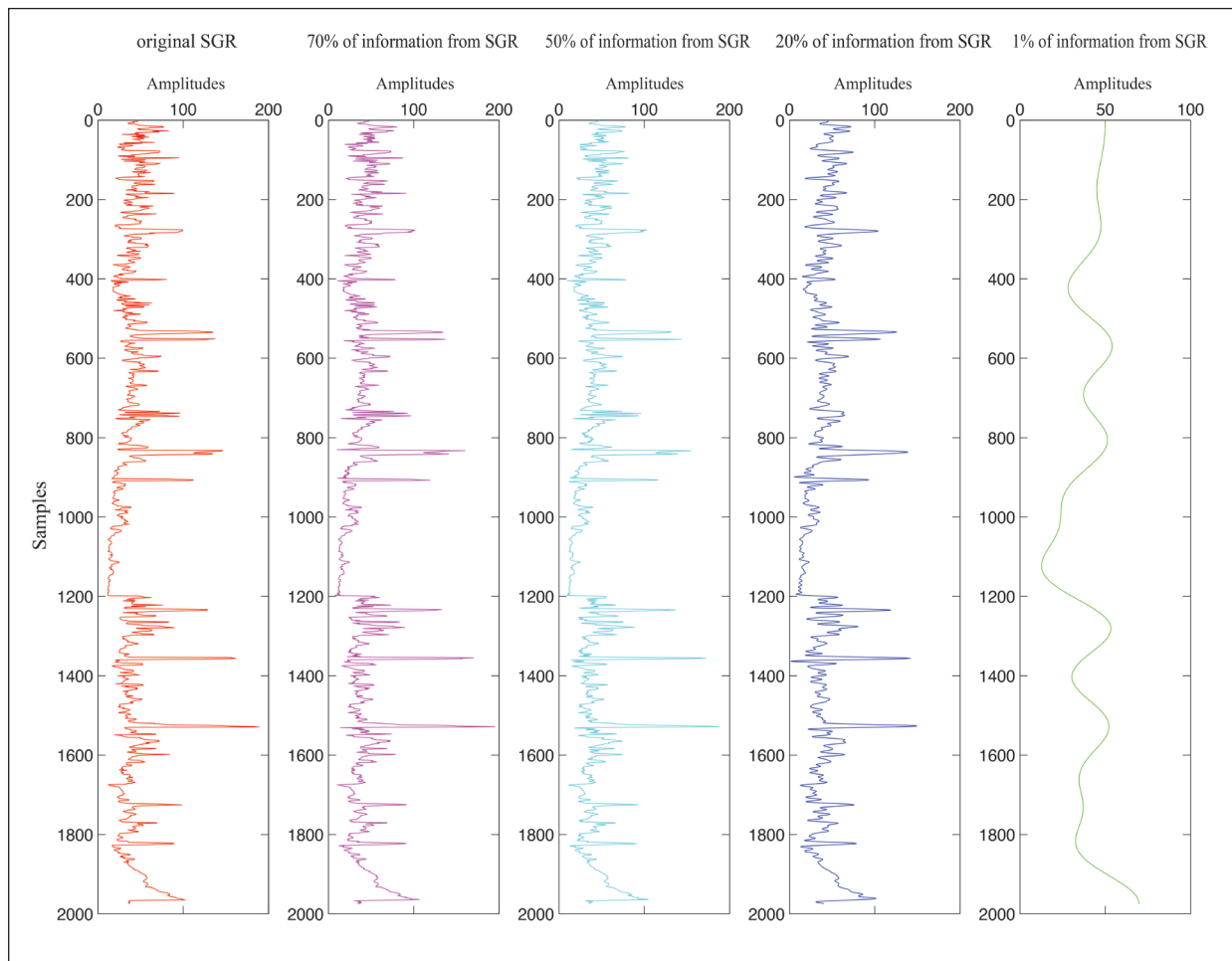
شکل ۱- مقایسه قدرت تبدیل کسینوسی گسسته در مقایسه با نمونه‌های داده اولیه.



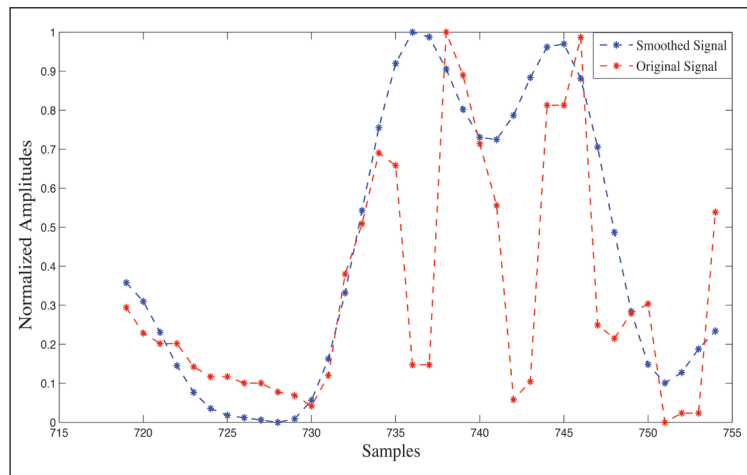
شکل ۲- روش‌های به کار رفته در روند پژوهش.



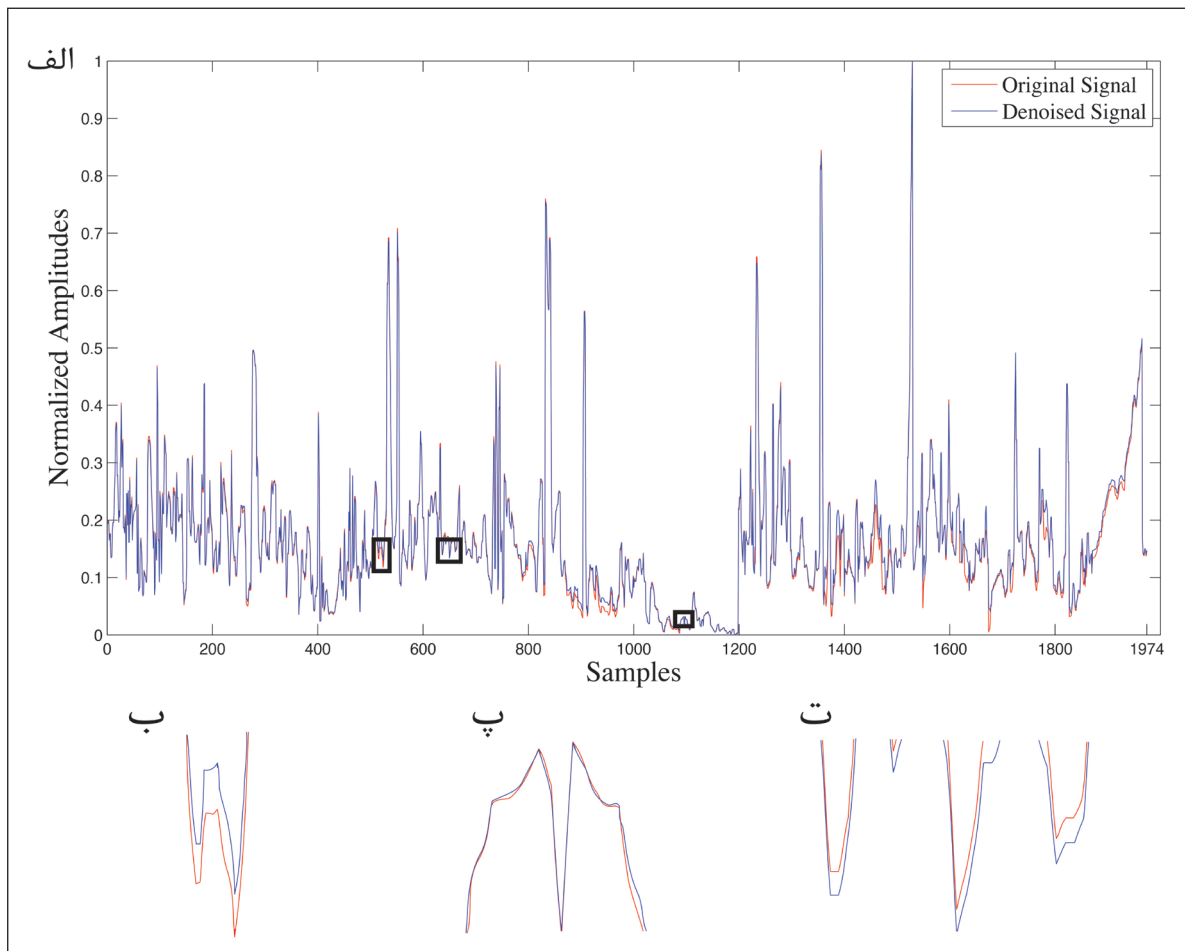
شکل ۳- میزان موفقیت دسته‌بندی خودکار در دو حالت مختلف.



شکل ۴- بازسازی سیگنال SGR با استفاده از سطوح مختلف اطلاعات.



شکل ۵- هموارسازی سیگنال SGR دارای ترکیب سنگی یکسان.



شکل ۶- سیگنال اولیه و حذف نوفه شده از راه تحلیل سریع مؤلفه‌های مستقل؛ الف) سیگنال کامل؛ ب، پ و ت) پنجره‌های مستطیلی که تفاوت دو سیگنال را آشکارتر نشان می‌دهد.

جدول ۱- مشخصه نمونه مغزه‌های انتخابی.

نام چاه	بازه ژرفایی	متراژ مغزه	تعداد کل نمونه مغزه‌ها	تعداد نمونه مغزه‌های انتخابی
MN-317	۳۴۸۳/۶۵۰۰ - ۳۶۶۴/۶۳	۱۸۰/۹۸	۱۲۷	۱۰۲
MN-330	۳۲۶۸/۳۷۳۸ - ۳۷۳۸/۸۵	۴۷۰/۰۸	۱۲۰۸	۱۰۹۶
MN-331	۳۲۲۶/۳۵۸۸ - ۳۵۸۸/۲۹	۳۶۲/۲۱	۱۰۳۲	۹
MN-338	۳۴۳۰/۳۸۹۲ - ۳۸۹۲/۸۴	۴۶۲/۰۷	۸۳۷	۷۶۷
مجموع	-	-	۳۲۰۴	۱۹۷۴

جدول ۲- نمونه‌هایی از مقادیر SGR برای یکی از ترکیب‌سنگی‌ها.

ماسه سنگ	
ژرفا (m)	نگار گاما (API)
۳۲۶۲/۳۰۰۴	۸۴/۴۷
۳۳۳۸/۰۴۳۱	۱۵/۶۱
۳۳۶۲/۸۸۴۲	۱۰۵/۲۰
۳۳۸۶/۶۵۸۶	۵۷/۶۲
۳۴۶۴/۲۳۰۱	۸۰/۰۳

جدول ۳- درصد پیشگویی صحیح بیشینه نسبت به تقدم اعمال تبدیل کسینوسی گسسته و تحلیل سریع مؤلفه‌های مستقل.

	۱ مؤلفه (%)	۵ مؤلفه (%)
پیش از تبدیل کسینوسی گسسته	۷۷/۰۸۱	۸۹/۶۹۶
هنگام تبدیل کسینوسی گسسته	۷۷/۲۴	۹۰/۲۶۸
پس از تبدیل کسینوسی گسسته	۶۶/۹۳۳	۸۱/۳۴۶

جدول ۴- نرخ موفقیت دسته‌بندی خودکار طبق تعداد مؤلفه‌های مستقل.

۱ مؤلفه (%)	۲ مؤلفه (%)	۳ مؤلفه (%)	۴ مؤلفه (%)	۵ مؤلفه (%)
۷۵/۷۲۸	۷۶/۱۵۳	۷۷/۹۰۴	۷۷/۸۰۴	۷۷/۸۸۴

References

- Abdi, H. and Williams, L. J., 2010- Principal Component Analysis. *WIREs Comp Stat* 2, 433–459.
- Amaziane, B., Bourgeat, A., Jurak, M., 2006- Effective macrodiffusion in solute transport through heterogeneous porous media. *Multiscale Modeling and Simulation* 5, 184–204.
- Avseth, P., Mukerji, T. and Mavko, G., 2005- Quantitative Seismic Interpretation. Applying Rock Physics to Reduce Interpretation Risk. Cambridge University Press, Cambridge, New York, Melbourne.
- Battiato, S., Mancuso, M., Bosco, A. and Guarnera, M., 2001- Psychovisual and statistical optimization of quantization tables for DCT compression engines. In: *Proceedings of the 11th International Conference on Image Analysis and Processing, ICIAP'01, Palermo, Italy*, p. 602.
- Bhatia, N. and Vandana, A., 2010- Survey of nearest neighbor techniques. *International Journal of Computer Science and Information Security* 2 (8), 302–305.
- Blinn, J. F., 1993- What's the deal with the DCT. *IEEE Computer Graphics and Applications* 13 (4), 78–83.
- Burden, R. L., Faires, J. D., 1985- Numerical Analysis, Third Edition Prindle, Weber and Schmidt, Boston.
- Carrasquilla, A. and Leite, M. V., 2009- Fuzzy logic in the simulation of sonic log using as input combinations of gamma ray, resistivity, porosity and density well logs from Namorado Oilfield. In: *Proceedings of the 11th International Congress of the Brazilian Geophysical Society, Salvador, Brazil*.
- Coconi-Morales, E., Ronquillo-Jarillo, G. and Campos-Enríquez, J. O., 2010- Multi-scale analysis of well-logging data in petrophysical and stratigraphic correlation. *Geofísica Internacional* 49 (2), 55–67.
- Comon, P., 1994- Independent component analysis: a new concept?. *Signal Processing*, 36; 287–314.
- Cover, T. M. and Hart, P. E., 1967- Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13 (1), 21–27.
- Doyen, P. M., 2007- Seismic reservoir characterization: an earth modelling perspective. EAGE Publications, Houten, The Netherlands.
- Dubrule, O., 1994- Estimating or choosing a geostatistical model. In: Dimitrakopoulos, R. (Ed.), *Geostatistics for the Next Century*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 3–14.
- Duda, R. and Hart, P., 1973- Pattern Classification and Scene Analysis. Wiley, New-York.
- Farina, A. and Studer, F. A., 1984- Application of Gram-Schmidt algorithm optimum radar signal processing. *IEEE Proceedings Part F* 131, 139–145.
- Franklin, J. N., 1968- Matrix Theory. Englewood Cliffs: Prentice-Hall. 292 pp.
- Grana, D., Pirrone, M. and Mukerji, T., 2012- Quantitative log interpretation and uncertainty propagation of petrophysical properties and facies classification from rock-physics modeling and formation evaluation analysis. *Geophysics* 77, WA45–WA63.
- Hyvärinen, A., 1999- Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks* 10 (3), 626–634.
- Hyvärinen, A., Karhunen, J. and Oja, E., 2001- Independent Component Analysis. John Wiley and Sons, Toronto 481 pp.
- Liu, Y., Weisberg, R. H., Mooers, C. N. K., 2006- Performance evaluation of the self-organizing map for feature extraction. *Journal of Geophysical Research* 111, C05018, <http://dx.doi.org/10.1029/2005JC003117>.
- MacQueen, J. B., 1967- Some methods for classification and analysis of multivariate observations. In: Le Cam, L.M., Neyman, J. (Eds.), *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, pp. 281–297.
- Martucci, S. A., 1994- Symmetric convolution and the discrete sine and cosine transforms. *IEEE Transactions on Signal Processing* SP-42, 1038–1051.
- Messina, A. and Langer, H., 2011- Pattern recognition of volcanic tremor data on Mt. Etna (Italy) with KAnalysis—a software program for unsupervised classification. *Computers and Geosciences* 37, 953–961.
- Mitchell, T., 1997- Machine Learning. McGraw-Hill Higher Education, New York 432 pp.
- Oppenheim, A. V., Schafer, R. W. and Buck, J. R., 2009- Discrete-Time Signal Processing, 3th ed. Prentice Hall, NJ 1120 pp.
- Rao, K. R. and Yip, P., 1990- Discrete Cosine Transform: Algorithms, Advantages, Applications. Academic Press, Boston 512 pp.
- Rosati, I. and Cardarelli, E., 1997- Statistical pattern recognition technique to enhance anomalies in magnetic surveys. *Journal of Applied Geophysics* 37 (2), 55–66.
- Russell, S. and Norvig, P., 2002- Artificial Intelligence: A Modern Approach. Prentice Hall, Essex, England 478 pp.
- Rutherford, S. R. and Williams, R. H., 1989- Amplitude versus offset variations in gas sands. *Geophysics* 54 (06), 680–688.
- Sanchetta, A. C., Leite, E. P. and Honório, B. C. Z., 2013- Facies recognition using a smoothing process through Fast Independent Component Analysis and Discrete Cosine Transform. *Computers and Geosciences* 57, 175–182.

- Schuerman, J., 1996- Pattern Classification: A Unified View of Statistical and Neural Approaches. Wiley and Sons, New York 392 pp.
- Simonoff, J. S., 1996- Smoothing Methods in Statistics. Springer, New York 368 pp.
- Toussaint, G. T., 2005- Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining. International Journal of Computational Geometry and Applications 15 (2), 101–150.
- Turlapaty, A. C., Anantharaj, V. G. and Younan, N. H., 2010- A pattern recognition based approach to consistency analysis of Geophysical datasets. Computers and Geosciences 36, 464–476.

Optimization determine of petrophysical facies class using Fast independent component analysis and discrete cosine transform based k-nearest neighbors in the Marun oil field, Asmari reservoir

S. A. Hasheminejad¹ and K. Ahmadi^{2*}

¹M.Sc., Department of Petroleum Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran

²Ph.D., Department of Petroleum Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran

Received: 2016 November 29

Accepted: 2017 May 23

Abstract

This paper aims to optimally determine petrophysical facies according to well log data. Using the automatic classification method of K-NN (K-Nearest Neighbours), petrophysical facies can be determined even though not optimally. For optimal determination of facies, the K-NN method is combined with FastICA (Fast Independent Component Analysis) and DCT (Discrete Cosine Transform) methods. This increases the success rate of the K-NN method. It also brings about optimal determination of petrophysical facies after which modelling and description of hydrocarbon reservoirs can be done. The research is performed in two different ways: In the first approach, the Fast ICA method is applied to data and then classified by the K-NN method. In the second approach, FastICA and DCT methods are applied to data and then classified by the K-NN method. Finally, the success rate of classification by the K-NN method is evaluated in both approaches to optimally determine petrophysical facies. Such evaluations indicate that application of the second method to data significantly enhances the success rate of the classification by the K-NN method, thereby leading to optimal determination of petrophysical facies, which is the very aim of this study. The utilized data including sonic log (DT), gamma rays (SGR), density (FDC or RHOB), neutron porosity (CNL or NPHI), and deep induction logs (ILD), belongs to the Marun oil field in southern Iran.

Keywords: Discrete Cosine Transform, Fast Independent Component Analysis, Determination of petrophysical facies, K-Nearest Neighbour, Smoothing.

For Persian Version see pages 25 to 34

*Corresponding author: K. Ahmadi; E-mail: petroyaran@gmail.com