مدل سازی پراکنده سلسله مراتب اشتراکی

نویسنده: مهران کریمیان ریزی –mehrankarimian97@ms.tabrizu.ac.ir
مریم فریور – farivar.maryam@yahoo.com

چکیده: مدل سازی انعطاف پذیر یک چارچوب قدرتمند برای تجزیه و تحلیل داده ها و پردازش داده ها است. به طور سنتی، کدگذاری در این چارچوب با حل یک مشکل رگرسیون خطی حلقوی ***L-1*** معمولا ***Lasso*** نامیده می شود.
در این کار ما ابتدا خواص القا کننده مدل ***Lasso*** را در سطح ویژگی فردی با خصوصیات انعکاسی بلوک مدل گروه ***Lasso*** ترکیب می کنیم، جایی که گروه های پراکنده از ویژگی ها به صورت مشترک کدگذاری می شوند و یک الگوریتم اسپارسیتی را که سلسله مراتبی ساختار یافته را به دست می اورد.
این نتیجه در سلسله مراتبی ***Lasso*** است که نشان می دهد مزایای مدل سازی عملی مهم است. سپس این رویکرد را به مورد مشارکتی گسترش می دهیم، جایی که مجموعه ای از سیگنال های به طور همزمان کدگذاری یک الگوی اسپارسیتی(ریزش) مشابه در سطح بالاتر (گروهی) دارند، اما نه لزوما در پایین تر. سپس سیگنال ها گروه های فعال یا گروه ها را به اشتراک می گذارند اما لزوما همان مجموعه فعال نیستند. این برای برنامه های کاربردی مانند جداسازی منبع بسیار مناسب است. یک روش بهینه سازی کارآمد، که همگرایی را به بهینه ی جهانی تضمین می کند، برای این مدل های جدید طراحی شده است. ارائه پایه چارچوب جدید و رویکرد بهینه سازی با نمونه های تجربی و نتایج نظری اولیه تکمیل می شود.

کلیدواژه: *lasso*، سلسه مراتب اشتراکی، اسپارسیتی

۱ – مقدمه

به‌علاوه بر جذاب بودن در سطح نظری، نشان داده‌شده است که مدل‌سازی سیگنال پراکنده مجر به نتایج متعددی در پردازش سیگنال شده است. مدل استاندارد فرض می‌کند که یک سیگنال می‌تواند به‌صورت مؤثر توسط ترکیب خطی چندگانه اتم از یک فرهنگ لغت داده‌شده یا یادگیرنده ارائه شود. اتم‌های انتخاب‌شده ازآنچه معمولاً به‌عنوان مجموعه فعال شناخته می‌شود تشکیل می‌شود که توانایی آن به‌طور قابل‌توجه ای کوچک‌تر از اندازه فرهنگ لغت و ابعاد سیگنال است. در سال‌های اخیر نشان داده‌شده است که افزودن محدودیت‌های ساختاری به این مجموعه فعال در سطح مقاومت نمایش و در سطح
تفسیر سیگنال ارزش دارد. (به‌ویژه هنگامی‌که مجموعه فعال برخی از خواص فیزیکی سیگنال نشان می‌دهد) [۱] و منابع موجود در آن را ببینید. این امر منجر به کدگذاری پراکنده گروهی یا ساختاری می‌شود، بجای این‌که اتم‌ها به‌عنوان تک‌سلول‌ها بررسی کنیم ، اتم‌ها به‌صورت گروهی دسته‌بندی می‌شوند و چند گروه در یک‌زمان فعال هستند. یک روش جایگزین برای اضافه کردن ساختار (و استحکام) به مسئله این است که کدگذاری هم‌زمان سیگنال‌های چندگانه را در نظر بگیریم و درخواست کنیم که همه آن‌ها یک مجموعه فعال را به اشتراک بگذارند. این‌یک رویکرد فیلتر همگانی طبیعی برای کد نویسی پراکنده است، [۲] و منابع موجود در آن را ببینید. در این کار ما این مدل ها را در چند جهت گسترش می‌دهیم. اولاً ما یک مدل پراکنده سلسله‌مراتبی را ارائه می‌دهیم که در آن تنها تعداد کمی از گروه‌های پراکنده‌ی اتم‌ها در یک‌زمان فعال هستند، بلکه هر گروه از اسپاسیتی داخلی نیز برخوردار است. در سطح مفهومی، این بدان معنی است که سیگنال توسط چند گروه (دسته‌ها) نشان داده‌شده است، و در داخل هر گروه فقط چند عضو در یک‌زمان فعال هستند.
یک نمونه ساده از این‌یک قطعه موسیقی است (برنامه‌های کاربردی متعدد در ژنومیک که در آن فقط چند ابزار فعال هستند (هر یک از وسیله‌ها یک گروه است) و موسیقی واقعی توسط دستگاه به‌طور مؤثر توسط چند اتم از زیر

فرهنگ لغت(گروه)مربوط به آن نشان داده‌شده است. به‌این‌ترتیب، این روش پیشنهادی چارچوب کدگذاری پراکنده را برای اجرای مؤثر جداسازی منابع ارائه می‌دهد. درحالی‌که منابع منفرد (دسته‌ها / گروه‌ها) که سیگنال را تولید می‌کنند، جایی که منابع فردی سیگنال تولیدشده در همان زمان را شناسایی می‌کنند که کارایی آن را بازسازی می‌کنند. (کد پراکنده درون گروه). یک روش بهینه‌سازی کارآمد برای حل این چارچوب کدگذاری پراکنده سلسله‌مراتبی پیشنهادشده است. سپس، ما یک‌قدم فراتر از این می‌رویم. اکنون تصور کنید که چندین رکورد از دو ابزار مشابه داشته باشیم (یا پنجره‌های زمانی مختلف با رکورد مشابه) که هر بار آهنگ‌های مختلفی پخش می‌کنیم. سپس، اگر ما این روش جدید کدگذاری پراکنده را باهم به کار ببریم، انتظار داریم که رکوردهای مختلف گروه‌های مشابه ای را به اشتراک بگذارند (ازآنجایی‌که از ابزارهای مشابه استفاده می‌کنند)، اما هر یک از این‌ها الگوی اسپارسیتی منحصربه‌فرد در داخل گروه دارند.(ازآنجاکه هر رکورد ملودی متفاوتی است) ما یک چارچوب کد نویسی پراکنده پیشنهاد می‌دهیم که دقیقاً به این موضوع پاسخ می‌دهد. یک روش بهینه‌سازی کارآمد برای این حالت نیز به دست می‌آید. در بقیه این مقاله، ما این مدل‌های جدید و بهینه‌سازی مربوطه خود را معرفی می‌کنیم، مثال‌های ارائه‌شده از آن‌ها را نشان می‌دهد و مسیرهای احتمالی تحقیقاتی که توسط این چارچوب‌های جدید، ازجمله برخی از نظریه‌های نظری ارائه‌شده، ارائه می‌دهیم.

۲- کد نویسی سلسله مراتبی مشترک

۲-۱ سابقه و هدف

فرض کنید ما مجموعه‌ای از نمونه‌های داده $x_i \in \mathbb{R}^m. j = 1. \dots . n$ و یک فرهنگ لغت از اتم‌های p، مونتاژ شده به‌عنوان ماتریس $D \in \mathbb{R}^{m\times p}$ و $D = [d_1 d_2 \dots d_p]$ را داریم. هر x_j نمونه می‌تواند به‌عنوان $X_j = Da_j + \varepsilon. a_j \epsilon R^p. \varepsilon \in R^m$ نوشته شود. به‌عنوان یک ترکیب خطی از اتم در فرهنگ لغت D به‌علاوه برخی از اختلال εاست. فرض اولیه اساسی در کدگذاری پراکنده این است که برای همه یا بیشتر j، بازسازی بهینه a_j تنها چند عنصر غیر صفر دارد.به‌طور رسمی اگر هزینه l_0 را به‌عنوان شمردن عدد عناصر غیر صفر a_j تعریف کنیم، $\|a_j\|_0 := |\{k : a_{kj} \neq 0\}|$، ما انتظار داریم که برای همه یا بیشتر $\|a_j\|_0 \ll p$باشد. بهینه‌سازی l_0 غیر محدب و معروف به NP-hard است، بنابراین تقریباً محدب بجای آن در نظر گرفته می‌شود که از هزینه معمول l_1 استفاده می‌کند

$$\min_a \|a\|_1 \quad s.t. \quad \|X_j - Da\|_2^2 \leq \varepsilon \qquad (۱)$$

تقریب بالا به‌عنوان Lasso شناخته‌شده است [۳]. یک نوع رایج این است که از نسخه بدون محدودیت استفاده کنید :

$$\min_a \frac{1}{2}\|X_j - Da\|_2^2 + \lambda\|a\|_1 \qquad (۲)$$

جایی که λ یک پارامتر است که معمولاً توسط اعتبار سنجی متقابل پیدا می‌شود. تنظیم‌کننده $\|\cdot\|_1$ باعث ایجاد محدودیت در راه‌حل a_j می‌شود.

این نه تنها ازنقطه‌نظر قانونی، بلکه همچنین از یک نقطه انتخاب مدل مطلوب است، جایی که فرد می‌خواهد ویژگی‌ها یا عوامل مرتبط (اتم‌ها) را شناسایی کند که هر نمونه x_j را منطبق می‌کنند. بااین‌حال، در بسیاری از موارد یکی از آن‌ها تمایل به نشان دادن عوامل مربوط به آن نه به‌عنوان یک اتم، بلکه به‌عنوان گروهی از اتم‌ها دارد. با توجه به یک فرهنگ لغت از اتم‌های p، گروه‌ها را از طریق شاخص‌های خود تعریف می‌کنیم، $g \subseteq \{1. \dots . p\}$با توجه به گروه g، زیرمجموعه اتم‌های D متعلق به آن به‌صورت D_g و مجموعه مربوطه از ضریب بازسازی خطی به‌عنوان ag مشخص می‌شود. تعریف $\mathcal{G} = \{g_1. \dots . g_{|\mathcal{G}|}\}$یک افراز از $\{1. \dots . p\}$ است. مسئله گروه Lasso در [۴] به‌صورت زیر معرفی شد:

$$\min_a \frac{1}{2}\|X_j - Da\|_2^2 + \lambda\psi_{\mathcal{G}}(a). \qquad (۳)$$

جایی که $\psi_{\mathcal{G}}$ گروهی از تنظیم‌کننده Lasso است که ازلحاظ $\mathcal{G}$ به‌عنوان $\psi_{\mathcal{G}(a)} = \sum_{g\epsilon\mathcal{G}}\|a_g\|_2$ تعریف‌شده است. توجه داشته باشید که $\psi_{\mathcal{G}}$ را می‌تواند به‌عنوان l_1 در هنجارهای اقلیدسی بردارها که توسط ضرایب متعلق به یک گروه a_g تشکیل‌شده است، دیده شود. این‌یک تعمیم از تنظیم‌کننده l_1است. همان‌طور که از ماده خاص G = {1,2,…,P} به وجود می‌آید، و به همین ترتیب اثر آن بر گروه‌های a نیز یک تعمیم طبیعی ازآنچه با Lasso به دست می‌آید،این است که اتم‌ها را در گروه‌ها روشن یا خاموش می‌کند.

۲-۲- سلسله‌مراتب Lasso

گروه Lasso در سطح یک تک ضرایب با محدودیت (اسپارسیتی) در یک سطح گروه‌بندی می‌کند، درحالی‌که در داخل هر گروه، راه‌حل مساوی است. (درواقع آن را به حداقل مربعات درون گروه کاهش می‌دهد) [۵].

همان‌طور که ما علاقه‌مند به حفظ محدودیت (اسپارسیتی) در سطح ضریب می‌باشیم، ما به‌سادگی تنظیم‌کننده L1 را با گروه کنترل تنظیم مجدد می‌کنیم[۶]، منجر به مدل پیشنهادی سلسله‌مراتبی Lasso (HiLasso) می‌شود.

ما به این تنظیم‌کننده به‌عنوان L1 + L2 اشاره می‌کنیم. در بخش ۳ ما بهینه‌سازی کارآمد برای (۲,۴) پیشنهاد می‌کنیم، درحالی‌که در بخش ۴ ما تجربی فضایل این مدل را نشان می‌دهیم.

۲-۳ سلسله‌مراتب همکاری Lasso

نمونه‌های x_jیکسان اجزای فعال از فرهنگ لغت را داشته باشند، یعنی شاخص‌های ضرایب غیر صفر در a_jبرای همه نمونه‌های مجموعه‌ای مشابه هستند. چنین وابستگی را در مسئله رگرسیون منظم l_1 به‌اصطلاح مشارکت (همچنین به نام "چندوظیفه" و یا "هم‌زمان") مشکلی کدگذاری جزئی به وجود می‌آورد[۲-۷].

به‌طور خاص، اگر ما ماتریس ضرایب $A = [a_1 \dots . a_n]$ مرتبط با بازسازی نمونه $X = [x_1 \dots . x_n]$، مدل کدگذاری مختلط مشارکتی توسط:

$$\min_a \frac{1}{2}\|X - DA\|_F^2 + \lambda \sum_{k=1}^{p} \|a^k\|_2 \qquad (۴)$$

که در آن a^k ردیف k از A است، یعنی بردار n مقادیر مختلفی است که ضریب مربوط به اتم k ام برای هر نمونه j است. اگر ما اکنون این ایده را به گروه Lasso توسعه دهیم، ما یک فرمول مشترک برای Lasso را به دست می‌آوریم،

$$\min_a \frac{1}{2}\|X - DA\|_F^2 + \lambda \psi_{\mathcal{G}}(A) \qquad (۵)$$

جایی که تنظیم‌کننده $\psi_{\mathcal{G}}$ برای یک ماتریس به‌عنوان $\psi_{\mathcal{G}(a)}(A) = \sum_{g \in \mathcal{G}} \|A_g\|_F$ تعریف‌شده است، که a_g تشکیل‌شده توسط تمام ردیف متعلق به گروه g است. ما این نکته‌ها را انتخاب کردیم زیرا این تنظیم‌کننده طبیعی گسترش قانون تنظیم‌کننده در (۲,۳) برای مورد مشترک است. به‌طور کامل از این نظر، این ترکیب هنوز در نوشته‌ها موردبررسی قرار نگرفته است. در این مقاله ما یک گام روبه‌جلو حرکت می‌کنیم و همراه با فرمت سلسله‌مراتبی رفتار می‌کنیم. مدل ترکیبی که برای این مشکل پیشنهاد می‌شود (C-HiLasso) می‌تواند به‌صورت زیر نوشته شود:

$$\min_a \frac{1}{2}\|X - DA\|_F^2 + \lambda_2 \psi_{\mathcal{G}}(A) + \sum_{j=1}^{n} \lambda_1 \|a\|_1 \qquad (۶)$$

گروه مشارکتی Lasso یک مورد خاص از مدل ما است، زمانی که λ_1 صفر است. از سوی دیگر، می‌توان برای هر x_i مستقل Lasso را با تنظیم λ_2به صفر به دست آورد. این فرمول جدید به‌ویژه هنگامی مناسب است که بردارها اجزاء را ازدست‌داده باشند. در این صورت ترکیب اطلاعات از همه نمونه‌ها بسیار مهم است تا منجر به نمایندگی صحیح و انتخاب مدل (گروه) شود. این را می‌توان با تغییر اندکی در اصطلاح داده‌ها در (۲,۶) انجام داد. برای هر بردار داده x_iیک خطای بازسازی با استفاده از تنها عناصر مشاهده‌شده محاسبه می‌شود. توجه داشته باشید که اجزاء گمشده بر سایر شرایط معادله تأثیر نمی‌گذارند.

۳- بهینه‌سازی

۳-۱ مشکل سیگنال-سیگنال HiLasso

در دهه گذشته، بهینه‌سازی مشکلات شکل (۲,۲) و (۲,۳) به‌شدت موردبررسی قرارگرفته است و الگوریتم‌های بسیار کارآمد برای حل آن‌ها وجود دارد. اخیراً، Wrightو همکارانش ارائه یک چارچوب، SpaRSA، برای حل مسئله کلی را در زیر مفروضات معقول پیشنهاد داد. [۸]

$$\min_a f(a) + \lambda \psi(a) \qquad (۷)$$

برای تضمین همگرایی F باید یک عملکرد صاف و محدب باشد، درحالی‌که ψ فقط باید در $\mathbb{R}^n$ محدود باشد. وقتی تنظیم‌کننده، ψ، جداشدنی است، بهینه‌سازی می‌تواند در هر گروه به موضوعات کوچک‌تر تقسیم شود. این چارچوب زمانی قوی می‌شود که این زیرمسیله ها بتوانند به‌طور مؤثر حل شوند. این مورد از تنظیمات Lasso و گروه Lasso است، اما تنظیم‌کننده L1 + L2 پیشنهادی فوری نیست. در این کار ما SpaRSA را با روش متناوب جهت تعدیل (ADMOM) [۹] ترکیب می‌کنیم تا مشکل HiLasso را حل کند. الگوریتم SpaRSA یک ترتیب از تکرار

$\{X^t\}_{t\in\mathbb{N}}$را تولید می‌کند که در شرایط خاصی به راه‌حل (۳,۸) تبدیل می‌شود. در هر تکرار، X^{t+1}به دست می‌آید:

$$\min_z (z-x^t)^T\nabla f(x^t)+\frac{\alpha^t}{2}\|z-x^t\|_2^2+\lambda\psi(z) \quad (۸)$$

برای برخی از توالی پارامترهای $\{\alpha^t\}_{t\in\mathbb{N}}$ با $\alpha^t\epsilon\mathbb{R}$ است. شرایطی که الگوریتم همگرایی آن‌ها را به انتخاب t بستگی دارد، برای جزئیات بیشتر به [۸] مراجعه کنید. آسان است نشان دهد که (۳,۹) برابر است با :

$$\min_z \frac{1}{2}\|z-u^t\|_2^2+\frac{\lambda}{\alpha^t}\psi(z) \quad (۹)$$

که در آن $u^t=X^t-\frac{1}{\alpha^t}f(X^t)$ است : در این فرمول جدید، روشن است که اولین اصطلاح در تابع هزینه می‌تواند به‌صورت المان‌های جداشده باشد. بنابراین هنگامی‌که تابع تنظیم (z) جداشدنی است، بنابراین بهینه‌سازی کلی است و می‌توان (۳,۱۰) را به‌صورت مستقل برای هر گروه حل کرد:

$$\min_{z_g}\frac{1}{2}\|z_g-u^t\|_2^2+\frac{\lambda}{\alpha^t}\psi_g(z_g) \quad (۱۰)$$

که در مورد HiLasso، این می‌شود،

$$\min_{b\epsilon\mathbb{R}^{|9|}}\frac{1}{2}\|b-w\|_2^2+\frac{\lambda_2}{\alpha^t}\|b\|_2+\frac{\lambda_1}{\alpha^t}\|b\|_1 \quad (۱۱)$$

جایی که

$$w=u_g^t \text{ and } u^t=a^t-\frac{1}{a^t}D^T(Da^T-x)$$

این‌یک SOCP است که می‌توان از حل‌کننده‌های عمومی استفاده کرد. بااین‌حال، این مشکل باید چندین مرتبه در تکرار SpaRSA حل شود، بنابراین برای حل آن بسیار مؤثر است.

برای این منظور از روش ADMOM استفاده می‌کنیم [۹]. ایده این است که مسئله مساوی مصنوعی محدودشده حل شود: جایی که $\widetilde{\lambda_\iota}=\lambda_i/a^t$ الگوریتم مجموعه‌ای از تکرارهای $\{b^t.\beta^t.p^t\}_{t\epsilon\mathbb{N}^+}$ را تولید می‌کند که به حداقل روش لاگرانژین تکمیلی همگرا می‌شود:

$$L_C(b.\beta.p)=\frac{1}{2}\|b-w\|_2^2+\widetilde{\lambda_2}\|b\|_2+\widetilde{\lambda_1}\|b\|_1+p^T(b-\beta)+\frac{c}{p}\|b-\beta\|_2^2 \quad (۱۲)$$

که در آن عناصر p به‌اصطلاح چند ضربدرهای لاگرانز و C ثابت هستند. در هر تکرار، متغیرها b و β به‌روزرسانی می‌شوند، دریک‌زمان، با حداقل رساندن لاگرانژین تکمیلی و اجازه دادن به باقی‌مانده:

$$b^{t+1}=\operatorname*{argmin}_b\frac{1}{2}\|b-w\|_2^2+\widetilde{\lambda_1}\|b\|_1+b^Tp+\frac{c}{2}\|b-\beta\|_2^2$$

$$\beta^{t+1}=\operatorname*{argmin}_\beta\widetilde{\lambda_2}\|\beta\|_2-\beta^Tp+\frac{c}{2}\|b^{t+1}-\beta\|_2^2.$$

$$p^{t+1}=p+c(b^{t+1}-\beta^{t+1}) \quad (۱۳)$$

برای راحتی در علامت‌گذاری، عناصر فوق را برای تکرارها در مرحله t صرف نظرکردیم، فقط به صراحتاً آن‌ها را در مرحله t + 1 نشان دادیم. بروز رسانی برای b، به مختصات b به مقادیر اسکالر جداشدنی است.

شرایط بهینگی برای هرکدام از این مشکلات اسکالر منجر به یک متغیر ساده از اپراتور شناخته‌شده نرم‌افزاری$S(\omega_I.\lambda)=\operatorname{sgn}\omega_i\,max\{0.|w_i|-\lambda\}$ می‌شود.

برای راحتی، ما از علامت $S(\omega_I.\lambda)$ برای نشان دادن بردار به‌دست‌آمده در هنگام استفاده از عملگر نزدیک به آستانه (با پارامترλ) به هر عنصر w استفاده می‌کنیم. از سوی دیگر، به‌روزرسانی β به زیر مقیاس عددی تفکیک شدنی نیست. بااین‌حال، شرایط بهینه آن توسط $\acute{\beta}+\widetilde{\lambda_2}\partial\|\acute{\beta}\|_2-\acute{b}\ni 0$ داده می‌شود، که دقیقاً همان چیزی است که منجر به اپراتور انقباض بردار، s_v، در [۴] برای گروه Lasso شرح داده‌شده است.(درواقع خیلی ساده تراست، هیچ ضرب ماتریسی وجود ندارد):

$$S_v(b.\widetilde{\lambda_2})=\left[1-\frac{\widetilde{\lambda_2}}{\|b\|_2}\right]_+b \quad (۱۴)$$

سپس هر دو به‌روزرسانی را می‌توان در فرم بسته و بسیار مؤثر محاسبه کرد:

$$b = \frac{1}{c+1} S\left(w + c\beta - p.\widetilde{\lambda_1}\right).\beta = \frac{1}{c} S_v\left(p + cb.\widetilde{\lambda_2}\right). \quad (۱۵)$$

الگوریتم بسیار قوی است و در چندین تکرار برای بهینه آن همگرا می‌شود، درنتیجه یک رویکرد بسیار کارآمد برای حل مشکل (۳,۱۱) به دست می‌آید. سپس چارچوب SpaRSA یک رویکرد بسیار جالب برای پیشنهاد Hi-Lasso می‌شود. الگوریتم کامل در الگوریتم ۱ خلاصه‌شده است. هنگامی‌که یک گروه کل فعال نیست، یک سرعت اضافی افزایش می‌یابد، با دور زدن ADMOM به دست می‌آید.

از شرایط مطلوب (۳,۱۱) به این معنی است که اگر ۰ یک راه‌حل است وقتی‌که $\lambda_1 = 0$(گروه استاندارد Lasso)، آن نیز یک راه‌حل در مورد کلی است.

این را می‌توان با ارزیابی $S_v(w.\lambda_1) > 0$ بررسی کرد.

۳-۲ همکاری بهینه‌سازی HiLasso

اکنون ما یک الگوریتم بهینه‌سازی را برای حل کارآمد همکاری *HiLasso* پیشنهاد می‌کنیم. ایده اصلی این است که از *ADMOM* برای تقسیم مسئله کلی به دو مسئله فرعی استفاده کنید: یکی که مسئله چند سیگنال را به رگرسیون *L1* تک سیگنال رها می‌کند، و دیگری که پرونده چند سیگنال را به‌عنوان یک گروه مشابه با *Lasso* مطرح می‌کند. به‌این‌ترتیب ما از جدایی هر اصطلاح به‌عنوان نشان داده‌شده در شکل ۱ استفاده می‌کنیم.

ما یک مشکل بهینه‌سازی محدود را تعریف می‌کنیم:

$$min \frac{1}{2}\|X - DA\|_F^2 + \lambda_1 \sum_j \|a_j\|_1 + \lambda_2 \psi_g(B) \;\; s.t\; A = B \quad (۱۶)$$

تکرارهای *ADMOM* توسط (مافوق دینگز برای متغیرهای تکرار t را برای راحتی نشانه‌گذاری حذف کردیم) داده می‌شود.

$$A^{t+1} = \underset{A}{argmin} \frac{1}{2}\|X - DA\|_F^2 + \lambda_1 \sum_j \|a_j\|_1 + Tr(B^T P^{t+1}) + \frac{C}{2}\|B - A\|_F^2$$

$$B^{t+1} = \underset{B}{argmin} \frac{C}{2}\|B - A^{t+1}\|_F^2 + Tr(B^T P^{t+1}) + \lambda_2 \psi_g(B)$$

$$P^{t+1} = P + c(A - B) \quad (۱۷)$$

حل برای A^{t+1}: مسئله (۳,۱۳) می‌تواند با به‌روزرسانی یک ستون از ماتریس A در زمان از هم جدا کرد.

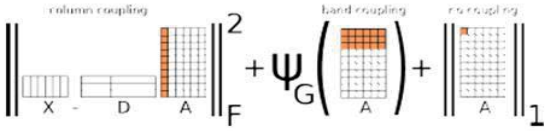
$$\min_{aj} \|x - Da_j\|_2^2 + p_j^T a_j + \frac{c}{2}\|a_j - b\|_2^2 + \lambda_1 \|a_j\|_1 \quad (۱۸)$$

این مشکل را می‌توان با استفاده از چارچوب *SpaRSA* حل کرد.

ایده این است که سه معیار اول هزینه را به‌صورت *f(0)* در معادله (۳,۸) در نظر بگیریم. هزینه محاسبات مربوطه برابر با یکی از *Lasso* است، ازآنجاکه معیار استاندارد *L1* است.

حل برای B^{t+1}: مشکل ارائه‌شده توسط (۳,۱۴) گروهی قابل‌تفکیک، به‌عنوان نتیجه مستقیم تفکیک‌پذیری ψ_g است. بنابراین، ما نیاز داریم که مشکلات بهینه‌سازی *lGl* را حل کنیم.

$$\min_{B_g} \frac{C}{2}\|B_g - A_g^{t+1}\|_F^2 + Tr\left(P^{t+1} B_g^T\right) + \lambda_2 \|B_g\|_F$$



جایی که A_g، B_g و P_g زیر ماتریس $n \times |g|$ از A، دسته p مربوط به گروه g است. ما آن‌ها را به‌عنوان بردارهای ستون (هرکدام با اجزای $n \times |g|$) بیان می‌کنیم، با همپوشانی ستون‌هایشان، به دست آوردن b_g. β_g و p_g به ترتیب، و مسئله بهینه‌سازی را در فرم برداری به‌صورت زیر بازنویسی کنید :

$$\min_b \lambda_2 \|b\|_2 - p_g^T b + \frac{c}{2}\|a_g^{t+1} - b\|_2^2 \quad (۱۹)$$

این مشکل همانند (۳,۱۲) است و می‌تواند به‌سادگی تغییر متغیرها به یک مشکل گروهی *Lasso* کاهش یابد و به‌این‌ترتیب با استفاده از آستانه گذاری حل می‌شود.

۴- نتایج تجربی

ما با مقایسه مدل ما با استاندارد *Lasso* و گروه *Lasso* با استفاده از داده‌های مصنوعی شروع می‌کنیم. ما دیکشنری‌های $|\mathcal{G}|$، D_i با ۶۴ اتم بعد ۶۴، با *i.i.d* نوشته‌های گاوسی درست می‌کنیم. ستون‌های نرمال شده دارای یک واحد $L2$ بودند. سپس ما به‌طور تصادفی دو گروه را انتخاب کردیم تا در هر زمان (بر روی تمام سیگنال‌ها) فعال شوند.

مجموعه‌ای از $N = 200$ تست سیگنال تولید شد، یک در هر گروه فعال، به‌عنوان ترکیب خطی از $k << 64$ عناصر واژه‌نامه، $x_j^i = D_i a_j^i$ است.

این سیگنال‌ها نیز نرمال شده‌اند. مخلوط‌ها به‌وسیله جمع‌کردن این سیگنال‌ها و (درنهایت) اضافه کردن نویز گازی از انحراف استاندارد σ ایجاد شد.تست سیگنال‌های تولیدشده دارای ساختار سلسله‌مراتبی پراکنده است و درحالی‌که گروه‌ها را به اشتراک می‌گذارند، لزوماً محدودیت درون گروه‌ها به اشتراک نمی‌گذارند. ما یک فرهنگ لغت را با پیوند دادن زیرمجموعه‌ها ساختیم، $D = [D_1 \dots D_{|\mathcal{G}|}]$و از آن برای حل مشکلات *Lasso*، گروه *Lasso*، *HiLasso* و *C-HiLasso* استفاده کنید. جدول ۱ خلاصه خطای میانگین مربع (*MSE*) و فاصله همدمینگ ضرایب بازدهی را نشان می‌دهد.

جدول ۱ : خطای میانگین مربع و فاصله همدمینگ ضرایب بازدهی

σ	Lasso	Glasso	HiLasso	C-HiLasso
۰,۱	۴۱,۷/۲۲,۰	۱۱۷,۳/۳۶۱,۶	۳۳,۰/۱۹,۸	۱۶,۳/۱۳,۳
۰,۲	۵۶,۴/۲۱,۶	۱۱۸,۲/۳۷۸,۳	۳۹,۹/۲۲,۷	۲۴,۹/۱۷,۱
۰,۴	۹۶,۵/۲۲,۷	۱۳۷,۸/۳۴۰,۳	۶۵,۶/۱۹,۵	۵۹,۵/۲۷,۴
K	Lasso	Glasso	HiLasso	C-HiLasso
۸	۳۸,۸/۲۲,۰	۱۱۸,۴/۳۱۸,۲	۲۷,۲/۱۹,۵	۹,۶/۱۶,۲
۱۲	۱۲۰,۰/۳۶,۲	۱۱۶,۶/۳۵۰,۴	۷۰,۴/۲۶,۵	۴۱,۳/۲۹,۱
۱۶	۱۶۴,۱/۴۳,۹	۱۰۹,۳/۳۳۸,۶	۱۱۰,۰/۳۲,۲	۵۵,۱/۳۵,۰
|g|	Lasso	Glasso	HiLasso	C-HiLasso
۴	۱۰۸,۰/۲۷,۸	۱۹۱,۶/۲۲۱,۷	۱۰۰,۹/۲۹,۸	۷۴,۲/۳۰,۲
۸	۱۲۰,۰/۳۶,۲	۱۱۶,۶/۳۵۰,۴	۷۰,۴/۲۶,۵	۴۱,۳/۲۹,۱

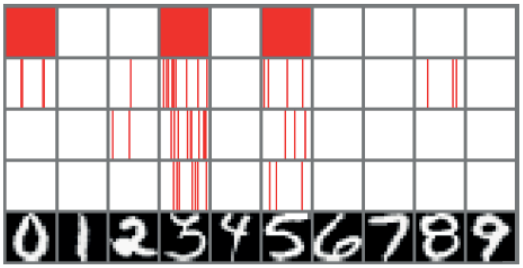
ما مشاهده می‌کنیم که مدل ما می‌تواند از ساختار سلسله‌مراتبی داده‌ها و همچنین ساختار همکاری استفاده کند. از دیدگاه مدل‌سازی، ما متوجه می شویم که گروه *Lasso* به‌طور کلی بلوک های صحیح را انتخاب می‌کند اما درون آن‌ها یک راه‌حل نادرست به دست نمی‌دهد.از سوی دیگر، *Lasso* راه‌حلی ارائه می‌دهد که دارای عناصر غیر صفر متعلق به گروه‌هایی است که در سیگنال اصلی فعال نبودند، منجر به انتخاب مدل اشتباه می‌شود.

HiLasso یک راه‌حل نادر است که اتم‌ها را از گروه‌های صحیح تشکیل می‌دهد اما هنوز برخی از اشتباهات جزئی را ارائه می‌دهد. در مورد همکاری، در تمام موارد مورد آزمایش، هیچ ضریبی در خارج از گروه‌های فعال صحیح انتخاب‌نشده و ضرایب بهبودیافته به‌طور مداوم بهترین هستند.

این استحکامات از این واقعیت است که گروه‌های فعال به‌طور مشترک با استفاده از اطلاعات موجود در تمام سیگنال‌ها پیدا می‌شوند. ما داده‌های عددی *USPS* را در نظر می‌گیریم که در چارچوب مدل‌سازی نزولی نشان داده‌شده است. [۱۰] در اینجا سیگنال‌ها بردارهایی هستند که دارای تراکم خاکستری پیچیده ۱۶ × ۱۶ تصویر است. ما دو رقم را انتخاب کردیم و آن‌ها را برای ایجاد یک تصویر مخلوط جمع کردیم. ما ۲۰۰ نمونه مخلوط تصادفی ایجاد کردیم و سپس آن‌ها را با روش‌های مختلف تجزیه‌وتحلیل کردیم. در این مورد هیچ حقیقی فعال زمین وجود ندارد و ما به‌عنوان یک ضریب عملکرد، خطای جدایی تعریف‌شده به‌عنوان $\frac{1}{NR}\sum_{i=1}^{R}\sum_{j=1}^{N}\left\|X_j^i - \hat{X}_j^i\right\|_2^2$ که جزء X_j^iمربوط به منبع i در سیگنال j است، استفاده می‌شود و $\hat{X}_j^i$ بهبودیافته است. با استفاده از تست تقسیم آموزش تمرین برای *USPS* ما برای اولین بار یک فرهنگ لغت را برای هر رقم آموختیم. سپس یک فرهنگ لغت را باهم پیوند دادن آن‌ها ایجاد کردیم. در شکل ۲ ما خطای جداسازی را در موقعیت‌های مختلف نشان می‌دهیم. همان‌طور که در حالت ترکیبی، تنها روش مشارکتی توانست با موفقیت منابع فعال واقعی را شناسایی کند. در شکل ۲ برخی از نمونه‌های بازتولیدی مجموعه‌های فعال برای هر روش نشان داده‌شده است.

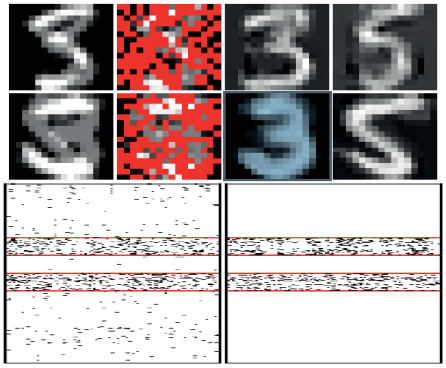
شکل ۲ : نمونه بازتولیدی مجموعه‌های فعال

Digits	Lasso	Glasso	HiLasso	C-HiLasso
۳+۵	۷۴,۱	۸۰,۱	۶۸,۶	۶۳,۴
۳+۵+ n	۸۷,۹	۹۵,۴	۹۲,۹	۷۷,۳
۲+۷	۶۱,۱	۶۰,۸	۵۸,۷	۴۲,۶
۲+۷+n	۷۵,۴	۶۵,۲	۶۴,۷	۵۳,۷



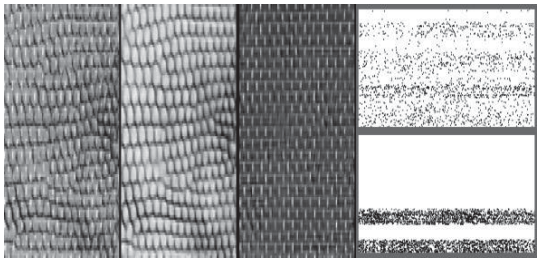
ما همچنین از مجموعه داده‌های ارقام برای آزمایش داده‌های گم‌شده استفاده می‌کنیم. به‌طور تصادفی، به‌طور متوسط ۶۰٪ از پیکسل‌ها در هر تصویر مخلوط را حذف کردیم و سپس *C-Hilasso* را اعمال کردیم. الگوریتم قادر است به‌درستی تشخیص دهد کدام رقم در تصاویر وجود دارد. در شکل ۳ ما چند مثال را نشان می‌دهیم. توجه داشته باشید که این‌یک مشکل کاملاً متفاوت ازآنچه معمولاً در نوشته‌های تکمیل ماتریس مطرح‌شده است. در اینجا ما قصد نداریم سیگنال‌هایی را که همه متعلق به یک زیرمجموعه ناشناخته منحصربه‌فرد هستند، نادیده بگیریم، اما سیگنال‌هایی که ترکیبی از دو فضای غیر منحصربه‌فرد هستند که به‌طور خودکار از فرهنگ لغت موجود انتخاب می‌شوند.

شکل ۳: نمونه



چنین فضاهای نامعلومی برای همه سیگنال‌های مورد سؤال هستند(سطح درشت سلسله‌مراتب)، اما لزوماً دقیق همان اتم‌ها نیستند و بنابراین لزوماً متعلق به همان زیر فضاهای نیستند. هر دو سطح سلسله‌مراتب به‌طور خودکار شناسایی می‌شوند، به‌عنوان‌مثال، گروه‌ها مربوط به "۳" و "۵" و اتم‌های دقیق (زیر فضاهای) در هر گروه هستند، این آخرین موارد برای هر سیگنال در مجموعه ممکن است. درحالی‌که ما در نظر داریم که زیربستی های احتمالی را می‌توان از فرهنگ لغت ارائه‌شده انتخاب کرد، در بخش ۵ ما در مورد یادگیری چنین لغت نامه ها نیز بحث خواهیم کرد. در چنین مواردی مشکل تکمیل ماتریس استاندارد یک مورد خاص از چارچوب *C-HiLasso* (با یک گروه واحد و تمام سیگنال‌های دارای یک مجموعه فعال، زیر فضا در گروه) می‌شود، به‌طور طبیعی، پرسش‌های نظری زیادی را برای این مدل عمومی جدیدتر ارائه می‌دهد . درنهایت *C-HiLasso* را برای جدا کردن بافته‌ای همپوشانی در یک تصویر استفاده کردیم. ما ۸ نمونه از مجموعه داده *Brodatz* را انتخاب کردیم و یک فرهنگ لغت را برای هر یک از آن‌ها آموزش دادیم (این ۸ گروه از فرهنگ لغت را تشکیل می‌دهند). سپس یک تصویر را به‌عنوان مجموع دو نمونه(تصویر تست در مرحله آموزش استفاده نشد) ایجاد کردیم. در شکل ۴ ما نتایج را نشان می‌دهیم. فاصله همینگ برای *C-HiLasso* = ۰,۰۰۳ است که نشان می‌دهد که گروه‌های درست و تنها آن‌ها در تمام زمان‌ها انتخاب می‌شوند.

شکل ۴: نتایج شبیه‌سازی



۵- بحث

در این مقاله چارچوب جدیدی از کد نویسی پراکنده سلسله‌مراتبی مشترک ایجادشده است، که در آن سیگنال‌های متعدد در رمزگذاری آن‌ها، تقسیم‌بندی گروه‌های کد (مدل) و داشتن (احتمال عدم پیوند)

نمایندگی‌های نادر در داخل گروه‌های مربوطه، همکاری می‌کنند. یک رویکرد بهینه‌سازی کارآمد ایجادشده است که همگرایی را به حداقل جهانی تضمین می‌کند و نمونه‌هایی از قدرت این چارچوب نشان داده‌شده است. در سطح عملی، ما در حال حاضر در زمینه برنامه‌های کاربردی این چارچوب پیشنهادشده در تعدادی از جهات، ازجمله جداسازی ابزارهای مشترک در موسیقی، و همچنین طبقه‌بندی سیگنال‌ها، به دنبال توانایی نشان داده‌شده به‌طور جمعی گروه‌ها / مدل‌های صحیح را انتخاب می‌کنیم.

در سطح نظری، یک کل خانواده از مشکلات جدید با این چارچوب پیشنهادی باز می‌شود. یکی از مهم‌ترین ویژگی‌های کلی انتخاب گروه‌های صحیح و درنتیجه آن، انتخاب مدل مناسب و جداسازی منبع است.

به‌عنوان‌مثال، در مورد دو گروه (به‌این‌ترتیب هیچ محدودیتی در سطح گروهی) و یک سیگنال واحد که توسط ترکیب خطی‌اتم از هر گروه تشکیل‌شده است، در نظر بگیریم.

سپس، آسان است نشان دهد که یکپارچگی متقابل بین گروه‌ها نقش مهمی ایفا می‌کنند. اجازه دهید $\mu_i\ i = 1.2$ ، انسجام داخلی اتم‌های گروه i و $\mu_i\ i$یکی بین گروه‌ها (همبستگی نرمال شده حداکثر بین اتم گروه ۱ با اتم گروه ۲). سپس آسان است که نشان دهیم که منحصربه‌فرد بودن جدایی می‌تواند تضمین شود اگر $(2k_1 - 1)\mu_1 + 2k\mu_{1.2} < 1\ and\ (2k_2 - 1)\mu_2 + 2k\mu_{1.2} < 1$،

با k_i سطوح محدودیت مربوطه در هر گروه (این ضعیف‌تر است که دقیق‌تر از آن توسط [۱۱] توسعه‌یافته است).این نیاز به محدودیت واقعی در سطح گروه و سطح مشترک است. توجه داشته باشید که البته با توجه به یک گروه فعال تنها یک نمونه خاص از مدل ما است (به‌این‌ترتیب برای کارهایی که در این مورد وجود دارد) به [۱۰]، بنابراین چارچوب نظری کلی برای چارچوب سلسله‌مراتبی پیشنهادشده ما به‌طور خودکار شامل چندین نتیجه موجود در برنامه‌نویسی جزئی است.

درنهایت، ما همچنین یک چارچوب برای یادگیری فرهنگ لغت برای برنامه‌نویسی چندگانه سلسله‌مراتبی مشترک ایجاد کرده‌ایم، به این معنی که بهینه‌سازی به‌طور هم‌زمان در فرهنگ لغت و کد است. انتظار می‌رود که این امر به معنای یادگیری فرهنگ لغت استاندارد باشد که موجب بهبود عملکرد قابل‌ملاحظه‌ای می‌شود (دوباره برای یک مورد خاص با یک گروه واحد در یک‌زمان) [۱۰].

مراجع

[1] R. Jenatton, J. Audibert, and F. Bach, “Structured variable selection with sparsity-inducing norms,” Tech. Rep.arXiv:0904.3523v1, INRIA, 2009.

[2] J. Tropp, “Algorithms for simultaneous sparse approximation. part ii:convex relaxation,” *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[3] R. Tibshirani, “Regression shrinkage and selection via the LASSO,” *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[4] M. Yuan and Y. Lin, “Model selection and estimationinregression with grouped variables,” *Journal of the Royal Statistical Society, Series B*, vol. 68, pp. 49–67, 2006.

J. Peng, J. Zhu, A. Bergamaschi, W. Han, D. Noh, J. Pollack, and P. Wang, “Regularized multivariate regression for identifying master predictors with application to integrative genomics study of breast cancer,” Annals of Applied Statistics.To appear.

[7] B. Turlach, W. Venables, and S. Wright, “Simultaneous variable selection,” Technometrics, vol. 27, pp. 349–363, 2004.

[8] J. Wright, R. Nowak, and M. Figueiredo, “Sparse reconstruction by separable approximation,” Trans. Sig. Proc., vol. 57, no. 7, pp. 2479–2493, 2009.

[9] D. Bertsekas and J. Tsitsiklis, Parallel and Distributed
Comptutation: Numerical Methods, Prentice Hall, 1989.

[10] I. Ramirez, P. Sprechmann, and G. Sapiro, “Classification and clustering via dictionary learning with structured incoherence and shared features,” in CVPR, 2010.

[11] J. Starck, M. Elad, and D. Donoho, “Image decomposition via the combination of sparse representations and a variational approach,” IEEE Transactions on Image Processing, vol. 14, pp. 1570–1582, 2004.