دوفصلنامه محاسبات و سامانه‌های توزیع شده
سال هشتم، شماره ۲، شماره پیاپی ۱۶، صفحه ۱ تا ۱۱، سال ۱۴۰۴

یک سیستم توصیه گر شغل هیبریدی مبتنی بر رزومه و سوابق افراد متقاضی با استفاده از الگوریتم‌های K-EM و SVM

رضا مولایی فرد[۱] و جواد محمدزاده[۲]
[۱] گروه مهندسی ایران کامپیوتر، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
[۲] گروه مهندسی کامپیوتر، دانشکده هوش مصنوعی، واحد کرج، دانشگاه آزاد اسلامی، کرج، ایران

چکیده

با توجه به رشد روزافزون مشاغل و افراد جویای کار، وجود سیستمی که بتواند مشاغل مناسب را به افراد جویای کار در بستر وب فراهم آورد، لازم و ضروری به نظر می‌رسد. بهترین راه برای انجام این کار استفاده از سیستم‌های توصیه گر می‌باشد. سیستم‌های توصیه گر می‌تواند پیشنهاداتی به کاربران ارائه دهند که موردعلاقه آن‌ها باشد. در این تحقیق به ارائه روش جدیدی به‌منظور بهبود سیستم‌های توصیه گر درزمینه پیشنهاد شغل به کاربران می‌پردازیم. روش کار بدین‌صورت می‌باشد که ما ابتدا سوابق مربوط به مشاغل افراد را جمع‌آوری می‌کنیم، سپس این داده‌ها را با استفاده از الگوریتم‌های داده‌کاوی و برحسب علاقه هر فرد طبقه‌بندی می کنیم. سپس این داده ها را با استفاده از سیستم توصیه گر مبتنی بر فیلترینگ مشارکتی در اختیار کاربر قرار خواهیم داد. نتایج حاصل از ارزیابی روش پیشنهادی حاکی از عملکرد بالای این سیستم پیشنهادی می باشد. به‌نحوی‌که روش پیشنهادی توانست میزان دقت ۹۲٪ و میزان فراخوانی ۹۶٪ را به دست آورد و به‌طورکلی این سیستم تا حدود ۹۰٪ می‌تواند توصیه‌هایی به کاربران پیشنهاد دهند که تا درصد بالایی می‌تواند موردعلاقه کاربر هدف باشد.

واژه‌های کلیدی: سیستم توصیه گر، شغل و سوابق، ماشین بردار پشتیبان، فیلترینگ مشارکتی، طبقه‌بندی، داده‌کاوی

تاریخچه مقاله:
تاریخ ارسال: ۱۴۰۴/۰۵/۲۵
تاریخ اصلاحات: ۱۴۰۴/۰۸/۲۷
تاریخ پذیرش: ۱۴۰۴/۱۰/۱۰
تاریخ انتشار: ۱۴۰۴/۱۲/۲۵
ایمیل نویسنده مسئول: molaefard@gmail.com

۱- مقدمه

با توجه به رشد روزافزون افراد جویای کار، وجود سیستمی که بتواند شغل مناسب افراد را بر اساس رزومه و سوابق این افراد، ارائه دهد لازم و ضروری به نظر می‌رسد. با توجه به بازدهی کاری افراد در محیط‌هایی که موردعلاقه کاربر باشد، وجود سیستمی که بتواند شرایطی را فراهم بیاورد که افراد بتوانند شغل موردعلاقه خود را در محیطی متناسب با روحیات و توانایی خود پیدا کنند می‌تواند موجب رشد و توسعه اقتصادی کشور، همچنین رضایت شغلی افراد شود. یکی از مهم‌ترین راه‌ها برای پیشنهاد شغل به افراد، استفاده از سیستم‌های توصیه گر می‌باشند. سیستم‌های توصیه گر یا پیشنهاددهنده، سیستم‌هایی هستند که با گرفتن اطلاعات محدودی از کاربر و خصوصیاتی مانند سوابق افراد، می‌توانند پیشنهادهایی به افراد دهند که موردعلاقه کاربر هدف باشد. سیستم‌های توصیه گر شخصی برای حل مشکل اضافه بار و استخراج اطلاعات موردنیاز کاربران و به‌طور گسترده در بسیاری از حوزه‌ها کاربرد دارند. سیستم‌های توصیه گر در حوزه استخدام در دهه‌های اخیر از رشد چشمگیری برخوردار بوده‌اند. مشخصات کاربر و فن‌آوری توصیه در سیستم‌های توصیه گر اخیراً به‌شدت موردتوجه قرارگرفته و بررسی‌های زیادی بر روی این سیستم قرار گرفته است و برای برخی موارد در صنایع نیز اجرایی شده است. درحالی‌که بازار بسترهای شغلی آنلاین در حال رشد است، کاربران در انتخاب شغل خود مشکل دارند. هنگامی‌که کاربران برای کار در وب‌سایت‌های مشاغل آنلاین درخواست می‌دهند، اطلاعات مختلف مانند محتویات شغل و شرایط استخدام را بررسی می‌کنند تا از جزئیات کار مطلع شوند. هنگامی‌که کاربران شغلی را انتخاب می‌کنند، به‌جای مشاهده و پشتیبانی عنوان شغل، بر جزئیات مختلف مربوط به شغل تمرکز می‌کنند. بااین‌حال، وب‌سایت‌های شغلی آنلاین موجود معمولاً مشاغل را با استفاده از اطلاعات ترجیحی کمی مانند رتبه‌بندی توصیه می‌کنند. بااین‌حال، اگر خدمات توصیه فقط با استفاده از اطلاعات کمی ارائه شود، عملکرد توصیه به‌طور مداوم روبه‌زوال است؛ بنابراین، خدمات توصیه شغلی باید با استفاده از اطلاعات

مختلف در مورد شغل، خدمات شخصی ارائه دهند. در سیستم توصیه گر شغل، افراد می‌توانند به جستجوی شغل موردعلاقه خود بپردازند. بین درآمد و اشتغال رابطه مستقیم و معناداری وجود دارد. از دو منظر می‌توان به این موضوع نگاه کرد؛ کارفرمایان و متقاضیان کارفرمایان اغلب اطلاعات موردنیاز متقاضیان، مانند توانایی‌ها و مهارت‌های آن‌ها را جمع‌آوری می‌کنند که از آن‌ها برای تصمیم‌گیری در مورد قابلیت اشتغال داوطلبان استفاده می‌شود. این فرایند زمان‌بر و نسبتاً پیچیده است. بر اساس دیدگاه متقاضیان، هر کس در مورد عواملی مانند حقوق، موقعیت، پاداش و انواع شغل اولویت‌ها و اولویت‌هایی دارد. متقاضیان می‌توانند از ابزارها و روش‌های متفاوتی برای یافتن شغل مناسب استفاده کنند، مانند روش سنتی جستجوی خانه به خانه که دشوار و زمان بر است. از طرف دیگر، از سیستم‌های توصیه کار می‌توان برای کمک به متقاضیان برای جستجوی سریع‌تر و راحت‌تر شغل استفاده کرد. علاوه بر این، هنگام جستجوی مشاغل موجود، ترجیحات متقاضی توسط سیستم‌های توصیه گر شغل مورد توجه قرار می‌گیرد. تاکنون تحقیقاتی بر روی سیستم‌های توصیه گر شغل صورت گرفته که در اکثر این تحقیقات مشکلاتی مانند عدم پیش‌بینی دقیق برای افراد بدون رزومه همچنین مشکل شروع سرد که مشکل اساسی سیستم‌های توصیه گر می‌باشد، وجود داشته‌اند که در این پژوهش سعی شد با استفاده از سیستم توصیه گر مبتنی بر فیلترینگ مشارکتی و سیستم توصیه گر مبتنی بر دانش همچنین طبقه‌بندی اطلاعات افراد با استفاده از ماشین بردار پشتیبان، این مشکلات حل شوند. در این پژوهش به ارائه روش جدیدی به‌منظور توصیه شغل به افراد بر اساس توانایی، رزومه و سوابق افراد با استفاده از ابزارهای طبقه‌بندی پرداخته می‌شود که می‌تواند شغل موردعلاقه کاربر را به وی پیشنهاد دهد.

۲- کارهای مرتبط

ژائو و همکاران خود در سال ۲۰۲۱ به ارائه روشی به‌منظور حل مشکل عدم تطابق بین مشاغل و افراد و سپس توصیه شغل مناسب به این افراد با استفاده از سیستم‌های آنلاین توصیه گر پرداختند. این محققان عقده داشتند که یکی از مشکلات اصلی در سناریوی استخدام آنلاین، عدم مطابقت مناسب بین افراد جویای کار و شغل‌های موجود، می‌باشد. این محققان در مقاله خود از تکنیک‌های مورداستفاده در سیستم های توصیه گر مبتنی بر جاسازی برای مقیاس وسیع برای انتخاب شغل برای افراد استفاده کردند. روش کار بدین‌صورت بود که برای یادگیری جاسازی جامع برای پست‌های شغلی و داوطلبان از طریق سطوح نمایندگی که از متن خام و اطلاعات مکانی و معنایی افراد استفاده کردند. سپس مجموعه‌ای از ترکیب ادغام شغل و افراد برای ایجاد و آموزش شاخص استفاده می‌شود که از جستجوی تقریبی نزدیک‌ترین همسایه برای بازیابی افراد جویای کار پشتیبانی می‌کند. پس از اولین مرحله بازیابی افراد، یک مدل مرتب‌سازی مرحله دوم که از سایر اطلاعات زمینه‌ای استفاده می‌کند برای ایجاد نتیجه تطبیق نهایی استفاده شد. نتایج ارزیابی آفلاین و آنلاین نشان می‌دهد که سیستم پیشنهادی مبتنی بر جاسازی دو مرحله‌ای از نظر نرخ کلیک و کیفیت در مقایسه با موارد به‌دست‌آمده نسبت به سایر سیستم‌های موجود بهبود چشمگیری دارد [۱].

جیابلی و همکاران خود در سال ۲۰۲۱ به ارائه روش جدیدی به‌منظور بهبود سیستم‌های توصیه گر در زمینه پیشنهاد شغل آنلاین پرداختند. سیستم این محققان که Skills2Job نام دارد، می‌تواند پیشنهادات شغلی مناسبی را به افراد ارائه دهد. در این سیستم توصیه گر، پیشنهادهای شغل بر اساس مجموعه ایی از مهارت‌های کاربران، مناسب‌ترین مشاغل را از مجموعه ایی بزرگ از مشاغل آنلاین مشخص و سپس به افراد جویای کار توصیه می‌کند. این محققان داده‌های ارسال‌شده در سه کشور انگلستان، فرانسه و آلمان را پردازش کردند، سپس سیستم را برای چند جداسازی آموزش دادند، سپس به ارزیابی مهارت افراد برای توصیه این شغل‌ها به افراد پرداختند. نتایج حاصل از ارزیابی Skills2Job نشان می‌دهد که این سیستم پیشنهادی می‌تواند نتایج را با دقت بالایی به افراد توصیه کند [۲].

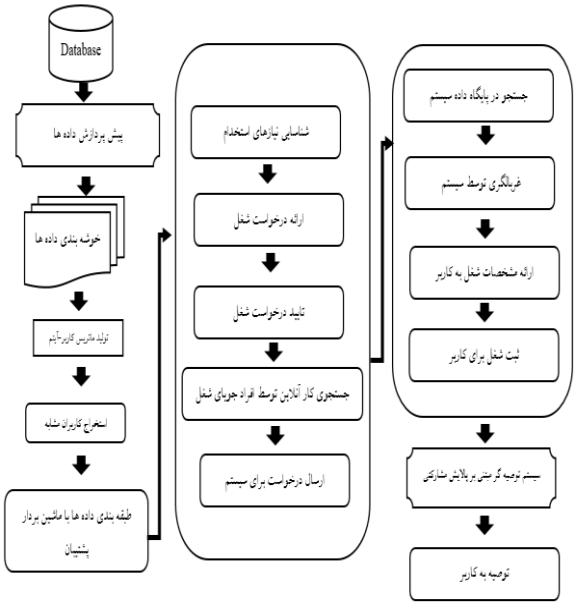
جها و همکاران خود در سال ۲۰۲۱ به ارائه روشی به‌منظور تولید یک سیستم توصیه گر شغل با استفاده از وزن دهی کاربران پرداخت. این محققان معتقد بودند که در جستجوی حجم عظیمی از اطلاعات موردنیاز برای استخدام در اینترنت،

یک کارجو همیشه ساعت‌ها را صرف یافتن اطلاعات جدید می‌کند. برای کاهش این کار پر زحمت، آن‌ها یک سیستم توصیه برای جستجوی شغل آنلاین طراحی و پیاده‌سازی کردند. روش کار بدین‌صورت بود که الگوریتم فیلترینگ مشارکتی مبتنی بر کاربر و را برای مقایسه یک الگوریتم بهتر انجام دادند. همچنین اطلاعات پیشین شامل رزومه و جزئیات استخدام اطلاعات را در نظر گرفتند. وزن کاربران متقاضی درخواست (کاربرانی که مشاغل نامزد را درخواست کرده بودند) و وزن کاربران مورداستفاده در مشاغل دلخواه را در الگوریتم استفاده کردند. درنهایت، مدل پیشنهادی از طریق مطالعه آزمایشی که از داده‌های واقعی استفاده می‌کرد تأیید شد. نتایج توصیه‌شده می‌تواند امتیاز بالاتری از دقت و فراخوانی را به دست آورد و با ترجیحات کاربران قبلی ارتباط بیشتری دارد. [۳].

لی و همکاران خود در سال ۲۰۲۱ به بررسی سیستم‌های توصیه گر شغل پرداختند. این محققان معتقد بودند درحالی‌که بازار بسترهای شغلی آنلاین در حال رشد است، کاربران در انتخاب شغل خود مشکل‌دارند. هنگامی‌که کاربران برای کار در وب‌سایت‌های مشاغل آنلاین درخواست می‌کنند، اطلاعات مختلف مانند محتویات شغل و شرایط استخدام را بررسی می‌کنند تا از جزئیات کار مطلع شوند. هنگامی‌که کاربران شغلی را انتخاب می‌کنند، به‌جای مشاهده و پشتیبانی عنوان شغل، بر جزئیات مختلف مربوط به شغل تمرکز می‌کنند. بااین‌حال، وب‌سایت‌های شغلی آنلاین موجود معمولاً مشاغل را با استفاده از اطلاعات ترجیحی کمی مانند رتبه‌بندی توصیه می‌کنند. بااین‌حال، اگر خدمات توصیه فقط با استفاده از اطلاعات کمی ارائه شود، عملکرد توصیه به‌طور مداوم روبه‌زوال است؛ بنابراین، خدمات توصیه شغلی باید با استفاده از اطلاعات مختلف در مورد شغل، خدمات شخصی ارائه دهند. روش پیشنهادی روشی را پیشنهاد می‌کند که با ارائه اطلاعات ترجیحی کیفی، مانند جزئیات مربوط به کار، عملکرد توصیه را بهبود می‌بخشد. برای این منظور، این مطالعه یک مدل‌سازی موضوعی را بر روی محتوای شغل مشخصات کاربر انجام می‌دهد. همچنین، با استفاده از تکنیک‌های LDA برای بررسی موضوعات از محتوای شغل و استخراج ترجیحات کیفی استفاده می‌کند. آزمایش‌ها نشان می‌دهد که روش پیشنهادی عملکرد توصیه بهتری در مقایسه با روش توصیه سنتی دارد [۴].

۳- تحلیل روش پیشنهادی

در روش پیشنهادی به ارائه روش جدیدی به‌منظور بهبود سیستم‌های توصیه گر در زمینه پیشنهاد شغل به کاربران می‌پردازیم. روش کار بدین‌صورت می‌باشد که ما ابتدا یک پایگاه داده از سوابق افراد جویای کار را جمع‌آوری می‌کنیم. سپس این اطلاعات را مورد پیش‌پردازش قرار می‌دهیم. هدف از پیش‌پردازش کردن داده‌ها، خارج کردن داده‌ها از حالت خام و تبدیل این داده‌ها به داده‌های مفید و قابل‌استفاده برای ابزارهای داده‌کاوی می‌باشد. سپس برای به دست آوردن اطلاعات مشابه داده‌ها را خوشه‌بندی می‌کنیم. سپس میزان علاقه هر کاربر را به شغل‌ها موردنظر را بر روی ماتریس کاربر-آیتم ترسیم می‌کنیم. مشخصات یک کاربر شامل دو نوع داده است: داده‌های تعاملات (داده‌های خود کاربر) و داده‌های ارتباطات اجتماعی (داده‌های دوستان کاربر). علاوه بر این، مشخصات کاربران و شرح مشاغل به چندین قسمت به نام زمینه شغل موردنظر تقسیم می‌شود. بخش دوم این مطالعه روشی را برای برآورد اهمیت هر زمینه از کاربران و مشاغل در انجام توصیه‌های شغلی ارائه می‌دهیم. در نهایت، بخش سوم به استفاده از الگوریتم یادگیری ماشین به‌منظور بهبود نتایج به‌دست‌آمده با اقدامات تشابه اختصاص داده‌شده است، ما SVM خطی را آموزش دادیم. نتایج ما نشان می‌دهد که استفاده از این روش یادگیری تحت نظارت، عملکرد سیستم توصیه گر مبتنی بر محتوا را افزایش می‌دهد. نمایی از روش پیشنهادی را در شکل (۱) مشاهده می‌کنید.



شکل ۱. چهارچوب روش پیشنهادی

۳-۱-داده‌ها

مجموع داده‌های مورداستفاده در این پژوهش شامل یک پایگاه داده از سوابق و رزومه مربوط به افراد جویای کار شامل ۳۰۷۰۰ رکورد می‌باشد که از این مجموع داده فایلی حاوی ۲۰۰۰ رکورد را به‌صورت رندوم استخراج کردیم که از این فایل برای روش پیشنهادی استفاده کردیم. همچنین از پایگاه داده ایی شامل ۵۰۰ شغل برای توصیه و شبیه‌سازی روش پیشنهادی استفاده گردیده است.

۳-۲-پیش‌پردازش داده‌ها

در اولین مرحله از روش پیشنهادی باید پایگاه داده خود را مورد پیش‌پردازش قرار دهیم. علت استفاده از عمل پیش‌پردازش داده‌ها خارج کردن داده‌ها از حالت خام و تبدیل داده‌ها به داده‌های مورداستفاده برای الگوریتم‌های طبقه‌بندی، خوشه‌بندی و داده‌کاوی می‌باشد. پیش‌پردازش داده‌ها در طی ۴ مرحله صورت می‌گیرد. این مراحل شامل پاکسازی جهت پاک کردن داده‌های ناقص و بدون استفاده، یکپارچه‌سازی داده‌ها برای تجمیع داده‌ها، کاهش ابعاد برای حذف مقادیر اضافی و تکراری و تبدیل داده‌ها برای تبدیل داده ا به داده‌های مفید و قابل‌فهم برای ابزارهای داده‌کاوی می باشد [۵،۶،۷،۸].

۳-۳- خوشه‌بندی داده‌ها

در این قسمت باید داده‌ها را خوشه‌بندی کنیم. روش خوشه‌بندی مورداستفاده الگوریتم K-means می‌باشد که یکی از مهم‌ترین الگوریتم‌های خوشه‌بندی می‌باشد. ابتدا داده‌ها را با الگوریتم K-means خوشه‌بندی می‌کنیم، سپس الگوریتم K-means را برای بهبود نتایج بهتر با الگوریتم الکترومغناطیس ترکیب می‌کنیم که ترکیب این دو الگوریتم، الگوریتم K-EM را تشکیل می‌دهند که می‌تواند نتایج بسیار دقیقی را به همراه داشته باشد. این الگوریتم پارامتر K را به‌عنوان ورودی گرفته و مجموعه n شیء را به k خوشه افراز می‌کند. به‌طوری‌که سطح شباهت داخلی خوشه‌ها زیاد است و سطح شباهت اشیاء بیرون خوشه کم باشد. شباهت هر خوشه نسبت به متوسط اشیاء آن خوشه سنجیده شده که این متوسط مرکز خوشه نامیده می‌شود [۹،۱۰،۱۱].

این الگوریتم بدین‌صورت کار می‌کند:

ورودی:k، تعداد خوشه است. پایگاه داده x شامل n شیء،

$$\{x_1,x_2,x_3,\ldots,x_n\}, x_i \in R^m X = \quad (۱)$$

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

مراحل اجرای الگوریتم خوشه‌بندی k-means به‌صورت زیر می‌باشد:

۱. به‌صورت تصادفی، kداده دلخواه را به‌عنوان 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 اقلیدسی در نظر می‌گیریم:

$$d(x_i, m_i) = \|x_i - m_i\| = \sqrt{\sum_{i=1}^{m}(x_{ik} - m_{jk})^2} \quad (۲)$$

که در آن X_i، i امین داده ورودی و m_j مرکز (میانگین) خوشه j ام است. داده مذکور در خوشه ایی قرار می‌گیرد که کمترین

فاصله را با آن خوشه داده باشد. مرکز خوشه نیز میانگین حسابی داده‌های آن خوشه است.

۱. مراکز خوشه‌ها را به‌روز می‌کنیم یعنی برای هر خوشه میانگین اعضای خوشه را به دست می‌آوریم.

۲. با توجه به مراکز جدید خوشه‌ها به مرحله ۲ برمی‌گردیم فرایند فوق را تا جایی ادامه می‌دهیم که هیچ تغییری در خوشه‌ها رخ ندهد (در این حالت الگوریتم پایان‌یافته است).

برای پیدا کردن K خوشه، مسئله بهینه سای تابع عملکرد، روی‌داده‌ها و مکان خوشه‌ها تعریف می‌شود. تابع عملکرد مورداستفاده برای این هدف، به‌صورت رابطه زیر انجام می‌شود [۱۷،۱۸،۱۹،۲۰].

$$f(O,c)=\sum_{i=1}^{n} min\left\{\|O_i - C_j\|^2 \mid j=1,2,..,K\right\} \quad (۳)$$

سپس الگوریتم K-means را با الگوریتم الکترومغناطیس ترکیب می‌کنیم. الگوریتم الکترومغناطیس از سازوکار جاذبه – دافعه نظریه الکترومغناطیس برای تعیین جواب بهینه استفاده می‌کند. این فرایند شامل ۴ مرحله اصلی می‌باشد: مقداردهی اولیه الگوریتم، انجام جستجوی محلی برای یافتن بهینه‌ی محلی، محاسبه نیروی کل وارده بر هر ذره و حرکت طول برای نیرو.

در مرحله اول فرایند، اندازه ذرات به‌طور کاملاً تصادفی از ناحیه موجه انتخاب می‌شوند که این حدود بین بازه [۱و۰] تنظیم می‌گردد. سپس مقدار تابع مقدار تابع هدف نقاط محاسبه‌شده و ذره‌ی دارای بهترین مقدار ذخیره می‌شود.

در مرحله دوم، جستجوی محلی برای یافتن بهینه‌ی محلی برای هر ذره، اجرا می‌شود. در این مرحله می‌توان از هر روش جستجوی محلی برای افزایش کارایی الگوریتم استفاده کرد.

مرحله سوم فرایند، به محاسبه نیروی کل وارده بر ذره اختصاص دارد. نیروی الکترواستاتیک بین دو نقطه‌ی بارها به‌طور مستقیم متناسب با مربع فاصله‌ی بین بارهاست. بار هر ذرهi ، قابلیت تعیین قدرت جذب یا دفع ذره‌ی i ام را دارد. این بار به‌صورت رابطه زیر محاسبه می‌شود.

$$q^i = exp\left(-n\ \frac{f(x^i)-f(x^{best})}{\sum_{j=1}^{popsize}\left(f(x^j)-f(x^{best})\right)}\right) \quad i=1,2,\dots,popsize \quad (۴)$$

بعد از محاسبه کل نیروی F یک‌ذره، در مرحله چهارم و نهایی به حرکت در طول جهت نیرو اختصاص دارد. ذره i در جهت نیرو با یک طول گام تصادفی حرکت می‌کند. طول گام تصادفی ϑ عرض می‌شود که از توزیع یکنواخت بین بازه‌ی صفر و ۱ تولیدشده است. به‌منظور حفظ موجه بود، نیروی اعمال‌شده روی هر ذره، نرمالایز می‌شود [21,22,23,24,25].

$$x^i = x^i + \vartheta\frac{f^i}{\|f^i\|}(RNG) \quad i=1,2,\dots,popsize \quad (۵)$$

مرحله دوم تا چهارم آن‌قدر تکرار می‌شود تا الگوریتم به یک توقف برسد.

گام‌های اساسی الگوریتم پیشنهادی عبارت است از:

فاز ۱. اجرای الگوریتم و تولید جمعیت اولیه

۱. انتخاب تصادفی k نقطه به‌عنوان مراکز ثقل اولیه.

۲. تخصیص هر نقطه به مرکز ثقل.

۳. به‌روزرسانی مکان هر مرکز ثقل به‌وسیله محاسبه میانگین مقدار نقاط تخصیص داده‌شده به مرکز.

۴. تکرار مراحل ۲ و ۳ تا زمانی که شرایط توقف ارضا شود، یعنی بیشترین تعداد تکرارها حاصل شود یا تغییری در مراحل مرکز ثقل حاصل نشود.

۵. استفاده از خروجی فاز اول به‌عنوان جواب کاندید برای فاز دوم.

فاز ۲. اجرای الگوریتم الکترومغناطیس

۱. محاسبه مقدار برازندگی برای همه جواب‌های کاندید.

۲. محاسبه نیروی کل وارد بر هر ذره.

۳. حرکت ذرات با توجه به نیروی برآیند وارد بر آن‌ها.

۴. تکرار مراحل ۲ و ۳ تا زمانی که شرایط توقف ارضا شود.

۴-۳- ماتریس کاربر-آیتم

در این مرحله باید داده‌های خوشه‌بندی‌شده از مرحله قبل را بر روی ماتریس کاربر_آیتم پیاده‌سازی کنیم تا علاقه هر کاربر به هر شغل را موردبررسی قرار دهیم. برای انجام این کار C را به‌عنوان همه کاربران و S را به‌عنوان تمام آیتم‌ها که می‌تواند

به کاربر پیشنهاد شود در نظر می‌گیریم. تابع سودمندی u، سودمندی آیتم s را برای کاربران c بیان می‌کند. مجموعه کل سفارشات را با R نشان داده می‌شود که به‌صورت $C \times R \rightarrow$ ت یف می‌نماییم. سپس برای هر کاربر $c \in C$ آیتم مثل $s \in S$ را که سودمندی کاربر را ماکزیمم نماید به‌صورت قراردادی طبق رابطه زیر تعریف می‌نماییم.

$$\forall_c \in \boldsymbol{C}, \quad \boldsymbol{S_c} = \arg\max \boldsymbol{u}\,(\boldsymbol{C},\boldsymbol{S}) \qquad (۶)$$

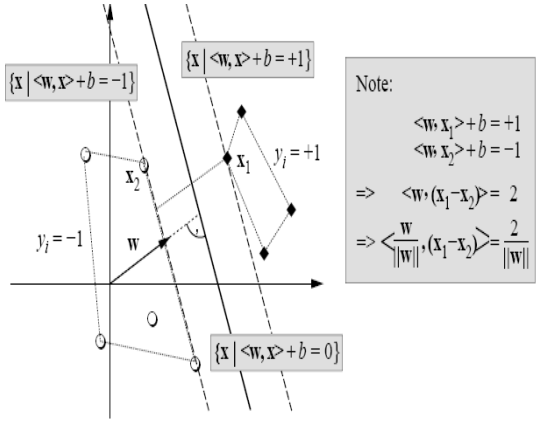
۳-۵- استخراج کاربران مشابه

پس‌ازاینکه کاربر جدید خوشه یا کلاس تعیین شد، همسایگان آن‌که شامل کاربران موجود در آن خوشه می‌باشد، استخراج می‌شوند. نظرات این همسایگان در پیشنهاد نهایی فیلم به کاربر جدید مؤثر هستند؛ اما همه همسایگان به یک اندازه به کاربر جدید شباهت ندارند و باید از یک معیار شباهت برای همسایگان نزدیک‌تر استفاده خواهد شد. فرض می‌کنیم کاربران سیستم به‌صورت مجموعه $U=\{u_1, u_2, u_3, \dots, u_m\}$ با خصوصیات $D=\{d_1, d_2, d_3, \dots, d_n\}$ و مجموعه فیلم‌ها به‌صورت $I=\{i_1, i_2, i_3, \dots, i_k\}$ تعریف شوند. سپس شباهت یک کاربر جدید و هر یک از همسایگان بر اساس رابطه زیر محاسبه می گردد.

$$\boldsymbol{d(x,y)} = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \qquad (۷)$$

۳-۶- طبقه‌بندی داده‌ها با ماشین بردار پشتیبان

در این مرحله باید داده‌های به‌دست‌آمده از مراحل قبل را جهت دسته‌بندی شغل‌ها و افراد، طبقه‌بندی کنیم تا بتوانیم شغل‌ها را به افراد توصیه کنیم. روش پیشنهادی موردنظر استفاده از الگوریتم SVM می‌باشد که می‌تواند دسته‌بندی دقیقی را بر روی داده‌ها انجام دهد. در روش ماشین بردار پشتیبان، بردارهای ورودی به یک فضای چند بعدی نگاشت می‌شوند. پس از آن، یک ابرسطح ساخته خواهد شد که با حداکثر فاصله ممکن، بردارهای ورودی را از هم جدا خواهد کرد [۱۲]. به این ابرسطح، ابرسطح با حداکثر مرز جداکننده گفته می‌شود. همان‌گونه که در شکل ۲ نشان داده‌شده است، دو ابرسطحِ موازی در دو سمتِ ابرسطح با حداکثر مرز جداکننده ساخته خواهد شد که داده‌های مربوط به دو طبقه را به‌گونه‌ای از هم مجزا می‌کنند که هیچ داده‌ای در مرز بین این دو ابرسطح قرار نمی‌گیرد. ابرسطح با حداکثر مرز جداکننده، ابرسطحی است که فاصله بین دو ابرسطحِ موازی را به حداکثر می‌رساند [۱۳،۱۴]. فرض بر این است که هرچقدر مرز جداکننده یا در واقع، فاصله بین دو ابرسطح موازی بیشتر باشد، خطای طبقه‌بندی هم کمتر خواهد بود. در پایان این مرحله داده ها دسته‌بندی و به دو طبقه مجزا تقسیم می‌شوند که از طریق این تقسیم‌بندی می‌توان داده‌ها را جهت داده‌های صحیح و غیر صحیح تقسیم نمود.



شکل ۲. نحوه ساخت ابرسطح جداکننده بین دوطبقه داده در فضای دوبعدی

۳-۷- توصیه داده‌ها به کاربران با سیستم توصیه گر فیلترینگ مشارکتی

درنهایت و در مرحله آخر داده‌های به‌دست‌آمده در مراحل قبل که شغل‌های می‌باشند را باید به افراد توصیه کنیم. سیستم مورداستفاده برای توصیه‌ها استفاده از سیستم توصیه گر فیلترینگ مشارکتی می‌باشد. این سیستم توصیه گر کار خود را با یافتن کاربرانی که در اولویت‌ها و سوابق خرید اشتراکاتی با کاربر جاری دارند شروع می‌کنند، در ادامه با جمع‌آوری اطلاعات مربوط به آیتم‌های موردعلاقه این گروه و حذف آن دسته آیتم‌ها

که قبلاً توسط کاربر خریداری شدند، لیستی از توصیه‌ها را ارائه می‌کند. به این گروه از سیستم‌ها، سیستم‌های مبتنی بر پالایش مشارکتی گفته می‌شود که ازجمله پرکاربردترین سیستم‌ها در راستای تولید توصیه به کاربران هستند [۱۵،۱۶]. مکانیزم اصلی الگوریتم پالایش مشارکتی به این شرح است که با استفاده از معیار تشابه افراد، اولویت‌های گروه‌های زیادی از کاربران ثبت می‌شود. در ادامه کاربرانی که با کاربر جاری اولویت‌های یکسانی دارند به‌عنوان همسایگان او انتخاب می‌شوند سپس میانگین اولویت‌ها محاسبه‌شده و تابع اولویت نهایی سعی در توصیه آیتمی می‌کند که کاربر امتیازدهی نکرده است، درنهایت تعدادی داده باقی خواهد ماند که این داده‌ها می‌توانند موردعلاقه کاربر باشند [۲۶،۲۷،۲۸،۲۹،۳۰].

۴- نتایج و شبیه‌سازی

برای ارزیابی سیستم‌های توصیه گر معمولاً از دو معیار دقت و میزان فراخوانی استفاده می‌گردد. در این تحقیق نیز از این دو معیار برای ارزیابی سیستم پیشنهادی استفاده گردید. میزان دقت و فراخوانی را از دو رابطه زیر به دست می‌آوردیم.

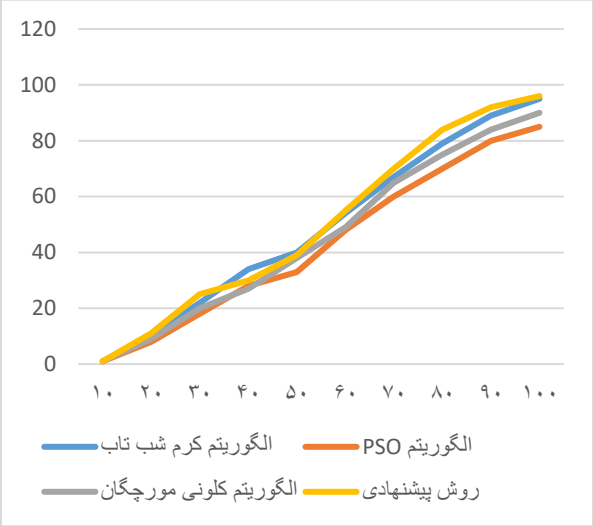
دقت با استفاده از رابطه زیر محاسبه می‌گردد.

$$\boldsymbol{Precision} = \frac{|\{\boldsymbol{relevant\ item} \cap \{\boldsymbol{retrved\ item}\}|}{|\{\boldsymbol{retrived\ item}\}|} \quad (۸)$$

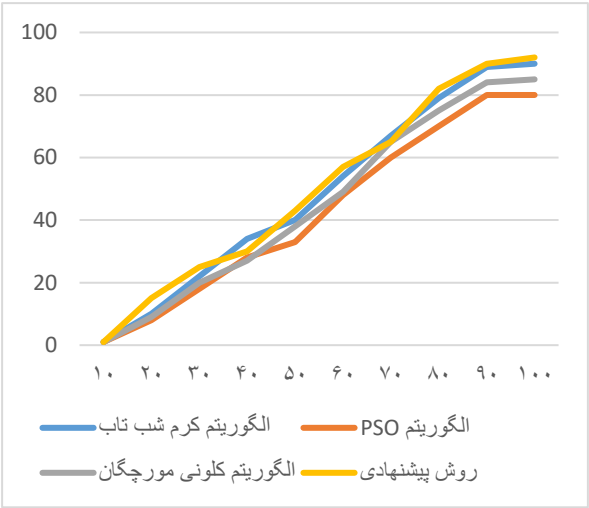
فراخوانی با استفاده از رابطه زیر محاسبه می‌گردد.

$$\boldsymbol{Recall} = \frac{|\{\boldsymbol{relevant\ item} \cap \{\boldsymbol{retrved\ item}\}|}{|\{\boldsymbol{relevant\ item}\}|} \quad (۹)$$

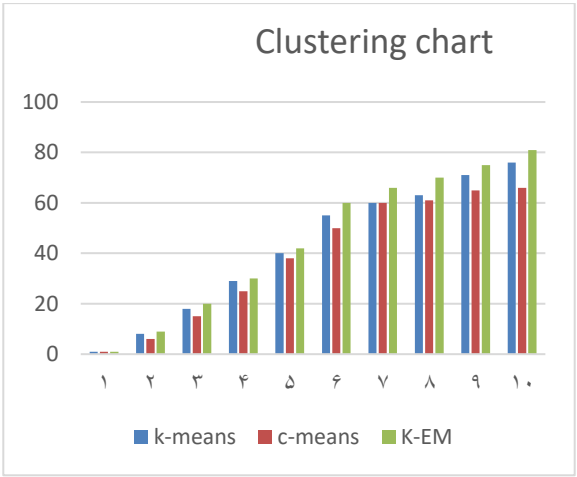
به‌منظور ارزیابی روش پیشنهادی الگوریتم‌های موجود درروش پیشنهادی با سایر الگوریتم‌های موجود مقایسه گردیدند که نتایج به‌دست‌آمده را در شکل‌های زیر مشاهده می‌کنید.



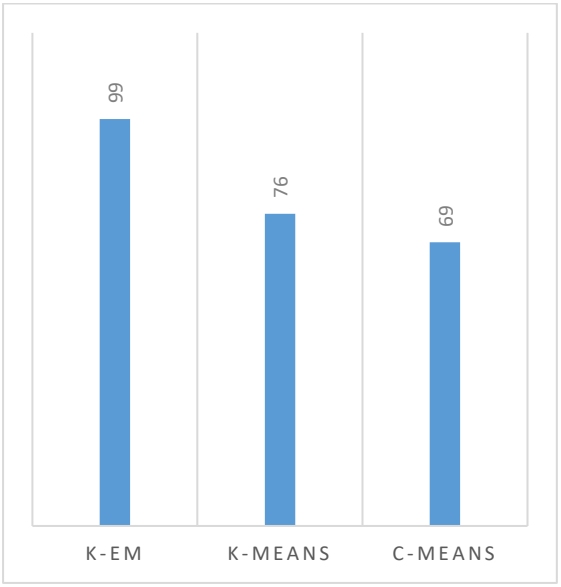
شکل ۳. مقایسه میزان دقت روش پیشنهادی و سایر روش ها



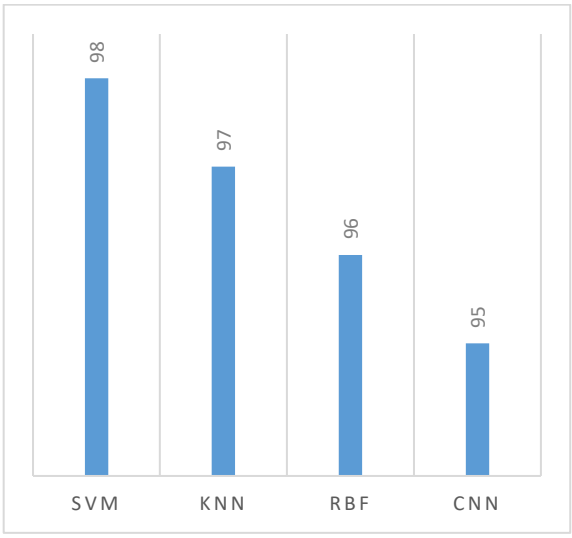
شکل ۴. مقایسه میزان فراخوانی روش پیشنهادی و سایر روش ها



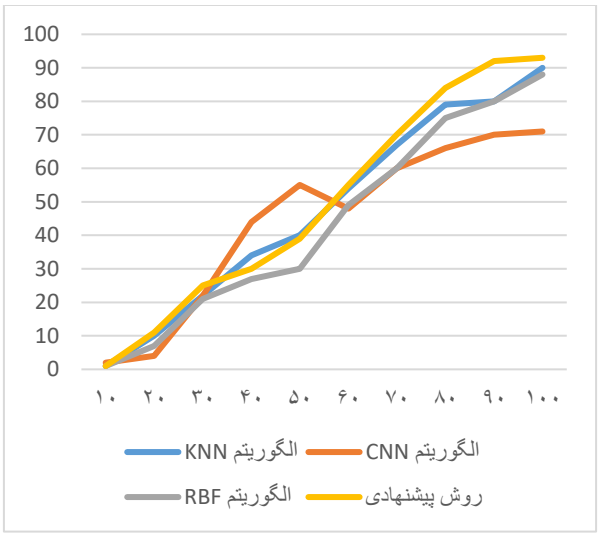
شکل ۵. مقایسه خوشه‌بندی روش پیشنهادی و سایر روش‌ها



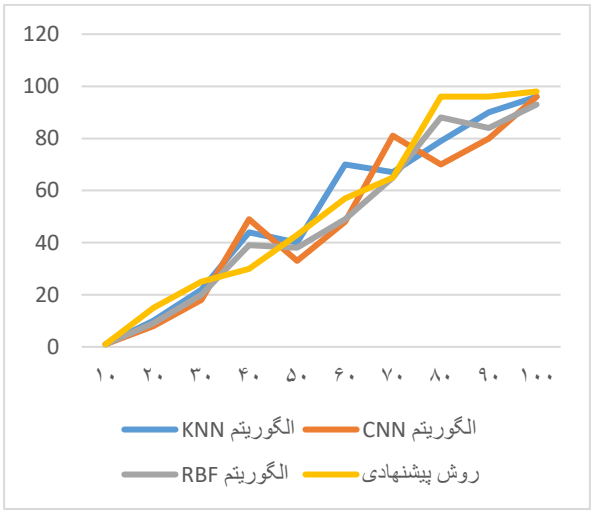
شکل ۶. مقایسه میزان کارایی خوشه بندی مورد استفاده در روش پیشنهادی و سایر الگوریتم های خوشه بندی



شکل ۷. مقایسه الگوریتم SVM و سایر روش های مشابه



شکل ۸. مقایسه میزان F1-Score روش پیشنهادی و سایر روش ها



شکل ۹. مقایسه ROC Curve روش پیشنهادی و سایر روش ها

۶- نتیجه‌گیری

با توجه به رشد روزافزون افراد جویای کار، وجود سیستمی که بتواند در دست‌یابی افراد به شغل موردنظر خود کمک و راهنمایی کند، لازم و ضروری به نظر می‌رسد. برای انجام این کار می‌توان با شخصی‌سازی سیستم ها، از اطلاعات گذشته افراد جویای کار برای پیشنهاد شغل موردعلاقه به این افراد استفاده نمود. یکی از بهترین و کاربردی‌ترین راه‌ها برای شخصی‌سازی سیستم ها استفاده از سیستم های توصیه گر می‌باشد. با توجه به کارایی سیستم‌های توصیه گر و قدرت روزافزون این سیستم‌ها در پیش‌بینی علایق کاربران در این تحقیق از یک سیستم توصیه گر شغل برای پیش‌بینی شغل موردعلاقه کاربران استفاده گردید. در این تحقیق با استفاده از الگوریتم‌های داده‌کاوی، اطلاعات ارزشمند افراد از میان حجم داده‌های بسیار بالا استخراج گردید و در قالب یک سیستم توصیه گر به افراد معرفی شدند. نتایج حاصل از تحقیق حاکی از درصد دقت بالای این سیستم در پیش‌بینی و پیشنهاد شغل‌های موردعلاقه هر فرد بود، به‌نحوی‌که در سنجش دو معیار دقت و میزان فراخوانی برای پیشنهاد شغل توانست به ترتیب درصدهای ۹۲٪ و ۹۶٪ را به دست آورد. با توجه به نتایج به‌دست‌آمده این سیستم تا ۹۰٪ می‌تواند شغل‌های موردعلاقه کاربران را به آن‌ها پیشنهاد دهد.

۷- منابع

1. *Zhao, J., Wang, J., Sigdel, M., Zhang, B., Hoang, P., Liu, M., & Korayem, M. (2021). Embedding-based Recommender System for Job to Candidate Matching on Scale. arXiv preprint arXiv:2107.00221.*
2. *Giabelli, A., Malandri, L., Mercorio, F., Mezzanzanica, M., & Seveso, A. (2021). Skills2Job: A recommender system that encodes job offer embeddings on graph databases. Applied Soft Computing, 101, 107049.*
3. *Jha, A., Purbia, L., & Ranawat, R. S. (2021). ONLINE JOB RECOMMENDER SYSTEM. International Journal of Advanced Research in Computer Science, 12.*
4. *Li, Q., Jeon, S., Lee, C., & Kim, J. K. (2021). A Study on the Job Recommender System Using User Preference Information. Journal of Information Technology Services, 20(3), 57-73.*
5. *Ranjan, K. G., Prusty, B. R., & Jena, D. (2021). Review of preprocessing methods for univariate volatile time-series in power system applications. Electric power systems research, 191, 106885.*
6. *Liberda, D., Pięta, E., Pogoda, K., Piergies, N., Roman, M., Koziol, P., ... & Kwiatek, W. M. (2021). The Impact of Preprocessing Methods for a Successful Prostate Cell Lines Discrimination Using Partial Least Squares Regression and Discriminant Analysis Based on Fourier Transform Infrared Imaging. Cells, 10(4), 953.*
7. *Frye, M., Mohren, J., & Schmitt, R. H. (2021). Benchmarking of Data Preprocessing Methods for Machine Learning-Applications in Production. Procedia CIRP, 104, 50-55.*
8. *Kakkar, M., Jain, S., Bansal, A., & Grover, P. S. (2021). Combining data preprocessing methods with imputation techniques for software defect prediction. In Research Anthology on Recent Trends, Tools, and Implications of Computer Programming (pp. 1792-1811). IGI Global.*
9. *Harsh, A. (2016). Automatic K-Expectation-Maximization (K-EM) Clustering Algorithm for Data Mining Applications.*
10. *Mehdizadeh, E., Teimouri, M., Zaretalab, A., & Niaki, S. T. A. (2017). A combined approach based on k-means and modified electromagnetism-like mechanism for data clustering. International Journal of Information Technology & Decision Making, 16(05), 1279-1307.*
11. *.Hung, W. L., & Chang-Chien, S. J. (2017). Learning-based EM algorithm for normal-inverse Gaussian mixture model with application to extrasolar planets. Journal of Applied Statistics, 44(6), 978-999.*
12. *Shankar, K., Lakshmanaprabu, S. K., Gupta, D., Maseleno, A., & De Albuquerque, V. H. C. (2020). Optimal feature-based multi-kernel SVM approach for thyroid disease classification. The journal of supercomputing, 76(2), 1128-1143.*
13. *Gopi, A. P., Jyothi, R. N. S., Narayana, V. L., & Sandeep, K. S. (2020). Classification of tweets data based on polarity using improved RBF kernel of SVM. International Journal of Information Technology, 1-16.*
14. *Kouziokas, G. N. (2020). SVM kernel based on particle swarm optimized vector and Bayesian optimized SVM in atmospheric particulate matter forecasting. Applied Soft Computing, 93, 106410..*
15. *Molaee Fard, R. (2021). Provide a video recommendation system using collaborative filtering and data mining methods. Journal of Advances in Computer Research, 12(1), 13-25..*
16. *Kashani, S. M. Z., & Hamidzadeh, J. (2020). Feature selection by using privacy-preserving of recommendation systems based on collaborative filtering and mutual trust in social networks. Soft Computing, 24(15), 11425-11440.*
17. *He, Z., Shen, X., Zhou, Y., & Wang, Y. (2024, January). Application of K-means clustering based*

on artificial intelligence in gene statistics of biological information engineering. In Proceedings of the 2024 4th International Conference on Bioinformatics and Intelligent Computing (pp. 468-473).

18. *Abdullah, A. A., Ahmed, A. M., Rashid, T., Veisi, H., Rassul, Y. H., Hassan, B., ... & Shamsaldin, A. S. (2024). Advanced clustering techniques for speech signal enhancement: A review and metanalysis of fuzzy c-means, k-means, and kernel fuzzy c-means methods. arXiv preprint arXiv:2409.19448.*
19. *. Khan, F., Khan, O., Parvez, M., Ahmad, S., Yahya, Z., Alhodaib, A., ... & Ağbulut, Ü. (2024). K-means clustering optimization of various quantum dots and nanoparticles-added biofuels for engine performance, emission, vibration, and noise characteristics. Thermal Science and Engineering Progress, 54, 102815.*
20. *Emambocus, B. A. S., Jasser, M. B., Zet, L. C., Ajibade, S. S. M., Wong, R. T., Chua, H. N., & Rafsanjani, A. S. (2024, March). A clustering algorithm employing salp swarm algorithm and K-means. In 2024 20th IEEE International Colloquium on Signal Processing & Its Applications (CSPA) (pp. 108-113). IEEE.*
21. *Zhu, C., Zhang, Y., Wang, M., Deng, J., Cai, Y., Wei, W., & Guo, M. (2024). Optimization, validation and analyses of a hybrid PV-battery-diesel power system using enhanced electromagnetic field optimization algorithm and ε-constraint. Energy Reports, 11, 5335-5349.*
22. *Formisano, A., & Tucci, M. (2024). Machine Learning Approaches for Inverse Problems and Optimal Design in Electromagnetism. Electronics, 13(7), 1167.*
23. *Fang, Y., Liu, K., Zhang, X., Lin, G., Wang, H., Dong, L., & Li, S. (2025). Application of Improved Electromagnetism-like Mechanism Algorithm on Massive Remote Sensing Image Screening. Applied Sciences (2076-3417), 15(7).*
24. *Ibrahim, A. M., & Tawhid, M. A. (2023). Chaotic electromagnetic field optimization. Artificial Intelligence Review, 56(9), 9989-10030.*
25. *Nouainia, A., Mejri, F., & Aguili, T. (2022). Analysis and optimization of electromagnetic leaks using genetic algorithms. Journal of Electromagnetic Waves and Applications, 36(17), 2519-2535.*
26. *molaee Fard, R. M., & Yarahmadi, P. (2022). Providing a Recommendation System for Recommending Articles to users using Data Mining Methods. Journal of Computer & Robotics, 15(2), 49-58.*
27. *molaee Fard, R. M., & Mosleh, M. A web recommender system for user nose pages using DBSCAN clustering algorithm and machine learning SVM method.*
28. *Kalideen, M. R., & YAĞLI, C. (2025). Machine Learning-based Recommendation Systems: Issues, Challenges, and Solutions. Journal of Information and Communication Technology, 2, 06-12.*
29. *Ionescu, B., Müller, H., Stanciu, D. C., Andrei, A. G., Radzhabov, A., Prokopchuk, Y., ... & Stein, B. (2025, September). Overview of imageclef 2025: Multimedia retrieval in medical, social media and content recommendation applications. In International Conference of the Cross-Language Evaluation Forum for European Languages (pp. 290-314). Cham: Springer Nature Switzerland.*
30. *Shih, K., Han, Y., & Tan, L. (2025). Recommendation system in advertising and streaming media: Unsupervised data enhancement sequence suggestions. arXiv preprint arXiv:2504.08740.*

رضا مولایی فرد دکترای رشته کامپیوتر در گرایش نرم افزار از دانشگاه علوم و تحقیقات تهران و مربی دانشگاه آزاد اسلامی واحد دزفول فعال در زمینه تحیل داده ها و زمینه پژوهشی یادگیری ماشین، داده کاوی و بهینه سازی. ایمیل :

molaefard@gmail.com

جواد محمد زاده دکترای رشته کامپیوتر از دانشگاه تهران و عضو هیئت علمی دانشگاه آزاد اسلامی واحد کرج و دارای زمینه پژوهشی شبکه های پیچیده پویا، سیستم های توزیع شده و الگوریتم های فراابتکاری. ایمیل:

mahammadzadeh@kiau.ac.ir

روش ارجاع: ر.مولایی فرد، ج. محمدزاده، یک سیستم توصیه گر شغل هیبریدی مبتنی بر رزومه و سوابق افراد متقاضی با استفاده از الگوریتم‌های K-EM وSVM . دوفصلنامه محاسبات و سامانه های توزیع شده، سال هشتم، شماره ۲، شماره پیاپی ۱۶، صفحه ۱ تا ۱۱، سال ۱۴۰۴.

How to cite: R.molaee fard, J. mohammadzadeh, A Hybrid Recommendation System Based on Resume and Record of the Applicant Using SVM and k-EM Algorithm. Journal of Distributed Computing and Sustem (JDCS), vo1 8, Issue 2, page 1-11, 2026.

A Hybrid Recommendation System Based on Resume and Record of the Applicant Using SVM and k-EM Algorithm

Reza molaee fard[1] and Javad mohammadzadeh[2]

[1] Department of Computer Engineering, SR.C., Islamic Azad University, Tehran, Iran
[2] Department of Computer Engineering, Ka.C., Islamic Azad University, Karaj, Iran

Abstract

Given the increasing number of jobs and job seekers, it seems necessary to have a system that can provide suitable jobs to job seekers on the web. The best way to do this is to use recommender systems. Recommender systems can provide users with suggestions that are of interest to them. In this research, we present a new method to improve recommender systems in the field of job suggestions to users. The working method is that we first collect the records related to the jobs of individuals, then we classify this data using data mining algorithms and according to the interest of each individual. Then we will provide this data to the user using a recommender system based on collaborative filtering. The results of the evaluation of the proposed method indicate the high performance of this proposed system. The proposed method was able to achieve an accuracy rate of 92% and a recall rate of 96%, and overall, this system can offer recommendations to users up to 90% of the time that can be of interest to the target user to a high degree.