EasyChair Preprint

№ 1668

Use of combination of hierarchical algorithms and evolutionary learning algorithms for Improved the Accuracy of Clustering of Complex Dynamical Networks Nodes

Alireza Farhang

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

October 16, 2019

پنجمین کنفرانس ملّی
محاسبات توزیعی و پردازش داده های بزرگ
۲۲ آبان ۱۳۹۸
5th National Conference on Distributed Computing and Big Data Proccessing
ISC
دانشگاه شهید مدنی آذربایجان

استفاده از ترکیب الگوریتم سلسله مراتبی و الگوریتم یادگیری تکاملی جهت بهبود دقت در خوشه بندی گره ها در شبکه‌های پیچیده پویا

علیرضا فرهنگ[۱]
[۱] دانشجوی دکتری مهندسی کامپیوتر، گروه مهندسی کامپیوتر دانشگاه آزاد میانه، میانه، ایران

چکیده

خوشه بندی در حوزه‌های مختلفی از جمله شبکه های پیچیده پویا کاربرد دارد. گره های حاضر در انجمن‌های یک شبکه به احتمال زیادی علایق مشترکی دارند. خوشه بندی در طراحی سیستم‌های توصیه کاربرد زیادی دارد. در این روش ابتدا گره هایی که در گذشته فعالیت مشابهی با گره فعال داشته اند شناسایی می‌شود سپس بر اساس امتیازاتی گره همتا، پیش بینی می‌کند گره فعال چه گره هایی را می‌پسندد و آنها را به خوشه توصیه می‌نماید. بر همین اساس ما در این پژوهش سعی داریم از سیستم های خوشه بندی برای بهبود خوشه بندی گره های شبکه های پیچیده پویا استفاده نماییم. هدف اصلی ما در این پژوهش بهبود خوشه بندی گره های شبکه های پیچیده پویا با استفاده از سیستم های خوشه بندی می باشد. از جمله فرضیات این پژوهش این است که سیستم خوشه بندی می تواند یک روش کارا برای خوشه بندی گره های شبکه های پیچیده پویا ارائه می دهد و همچنین بر اسا سوابق مکانی و زمانی گره می تواند به عنوان پارامتری جهت خوشه بندی آن ها مورد استفاده قرار بگیرد. در نهایت در این مقاله سعی داریم استفاده از ترکیب الگوریتم سلسله مراتبی و الگوریتم یادگیری تکاملی دقت در خوشه بندی گره ها در شبکه‌های پیچیده پویا را بهبود بخشیم.

کلمات کلیدی: خوشه بندی، شبکه های پیچیده پویا، الگوریتم سلسه مراتبی، الگوریتم یادگیری تکاملی.

۱- مقدمه

سیستم های خوشه بندی رویکردی برای مواجهه با مشکلات ناشی از حجم فراوان و روبه رشد اطلاعات سرریز اطلاعات می باشد و هدفش کمک به کاربر خود برای تصمیم گیری در میان این حجم از اطلاعات است. در طول دهه ی گذشته حجم بسیار زیادی از اطلاعات بر روی پایگاه داده ها و سرورهای اطلاعاتی قرار گرفته اند. ساخت سیستمی هوشمند برای اینکه بتواند در جستجو، مرتب سازی، دسته بندی و فیلترینگ اطلاعات به گره خود کمک کند امری ضروری و اجتناب ناپذیر بود، بنابراین این سیستم ها به وجود آمدند و اخیراً پیاده سازی آنها در اینترنت رو به افزایش است. در این فصل از پژوهش ما ابتدا مفاهیم پایه ای سیستم های خوشه بندی را شرح داده و سپس به مروری بر ادبیات سیستم های خوشه بندی و شبکه های پیچیده پویا می پردازیم.

سیستم های خوشه بندی از تکنیک های مختلفی برای تحلیل بهره می برند که می توان به روش های احتمالی، شبکه های بیزی، الگوریتم نزدیکترین همسایه ها، الگوریتم های الهام گرفته از زیست مثل شبکه های عصبی ، الگوریتم های ژنتیک و مدل های فازی اشاره کرد[۶]. مشکل شروع سرد هنگامی رخ می دهد که امکان ایجاد توصیه های قابل اعتماد به دلیل عدم وجود توصیه های اولیه برای گره مشابه وجود ندارد. ما می توانیم سه نوع مشکل شروع سرد را از هم متمایز کینم: جامعه ی جدید، آیتم جدید و کاربر جدید. نوع آخر در سیستم های خوشه بندی بسیار مهم هستند. مشکل جامعه ی جدید اشاره به سختی به دست آوردن مقدار کافی داده برای ساخت توصیه های قابل اعتماد در زمان شروع یک سیستم خوشه بندی میکند. دو راه معقول برای مقابله با این مشکل وجود دارد. تشویق گره برای رتبه بندی از طریق ابزارهای مختلف، انتخاب سیستم خوشه بندی بر پایه ی فیلترینگ مشترک وقتی گره و داده های زیادی وجود دارند. مشکل آیتم جدید به خاطر این

مطرح می شود که آیتم های جدیدی که در سیستم خوشه بندی وارد می شوند معمولاً رتبه بندی اولیه ای ندارند، و بنابراین آنها به احتمال زیاد به گره توصیه نمی شوند. به نوبه ی خود، یک آیتم که توصیه نمی شود، توسط یک بخش بزرگی از جامعه نادیده گرفته می شود. پس بخاطر اینکه آنها از این مورد غافل هستند آن را ارزیابی نمیکنند. یک راه حل عمومی برای این مشکل داشتن مجوعه ای از گره با انگیزه که مسئولیت رتبه بندی هر آیتم جدید در سیستم را به عهده می گیرند می باشد. مشکل کاربر جدید، نشان دهنده ی یکی از مشکلات بزرگ در مواجهه با سیستم های خوشه بندی در عمل است. زمانی که گره جدید در سیستم های خوشه بندی هنوز هیچ رتبه بندی انجام نداده اند، نمی توانند توصیه های اختصاصی برپایه ی فیلترینگ مشترک برپایه ی حافظه دریافت کنند. وقتی گره اولین رتبه بندی خود را انجام می دهند آنها از سیستم خوشه بندی انتظار دارند که به آنها توصیه های شخصی سازی شده پیشنهاد کند، اما تعداد رتبه بندی ها معرفی شده در سیستم خوشه بندی معمولاً برای قادر ساختن سیستم برای ایجاد توصیه های قابل اعتماد کافی نیستند. بنابراین گره جدید احساس میکنند که سیستم خوشه بندی سرویسی که آنها انتظار دارند را ارائه نمی کنند و به همین دلیل آنها ممکن است استفاده از آنرا متوقف کنند. استراتژی عمومی برای مقابله با مشکل کاربر جدید شامل استفاده از اطلاعات اضافی برای مجموعه ای از رتبه بندی ها به منظور قادر ساختن به ایجاد توصیه هایی بر پایه ی داده های در دسترس برای هر کاربر است. با مشکل شروع سرد معمولاً هنگام استفاده از رویکردهای ترکیبی مواجه می شویم.

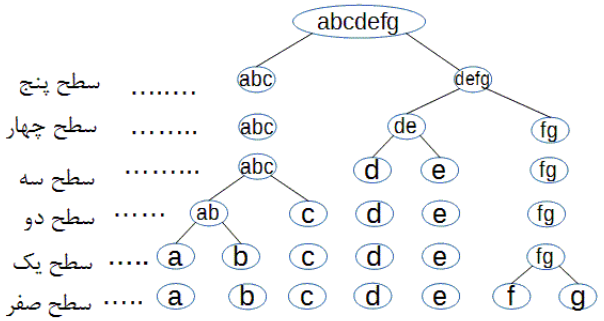
اخیرا رشد قابل توجهی در سایت های شبکه های اجتماعی رخ داده که تاثیر قابل توجهی بر جوانب مختلف زندگی داشته است. سیستم های خوشه بندی ها یا RS ها هم به شدت در حال رشد هستند که تاثیر خود را در فعالیت های مالی فوری با توجه به اینکه یک توصیه خوب اغلب تبدیل به یک خرید واقعی می شود، نشان داده اند. اگر چه هم شبکه های اجتماعی و هم RS ها در ابتدای کار مسیر کاملا جداگانه ای داشتند و محققان هم با آنها جداگانه برخورد می کردند اخیرا این دو بسیار به هم نزدیک شده اند. یکی از مهم ترین جوانب به کار برده شده در تجارت الکترونیک موفق، توانایی تشخیص محصولات و خدماتی است که افراد یا گره ممکن است به آنها علاقه مند باشند و سپس پیشنهاد این آیتم ها به گره ، به منظور اینکه ترغیب به خرید شوند، می باشد[۷]. این همان مسئله اصلی است که RS آن را هدف گرفته و مسئله مهمی نیز هست. هنگامی که بحث بازرگانی مطرح است مزایای اقتصادی مهم و فوری در یک پیشنهاد خوب در زمینه افزایش فروش و ایجاد بازار مشهود است. زمانی که بحث کاربر یا خریداران بالقوه مطرح است این روزها خریداران اغلب با تنوع انتخاب ها و گزینه ها در دنیای کسب و کار آنلاین روبرو هستند در حالی که زمان و منابع کمی برای سرمایه گذاری در فرایند انتخابی دارند. بنابراین نیازی روز افزون به پشتیبانی از پیشنهاد به کاربر، برای فائق آمدن بر این مشکلات و پیشنهاداتی به گره خصوصی در زمینه آیتم های گوناگون مانند کتاب، فیلم، موسیقی و اخبار احساس می شود. به علاوه ایده اصلی که در پس Recommender System ها وجود دارد را می توان نه تنها در تماس مستقیم با خرید و کسب و کار به کار برد بلکه می توان آنها را به مفاهیم گسترده تری نیز بسط داد، برای مثال به دنیای شبکه های اجتماعی که در آن این سیستم ها می توانند بسیار مفید باشند. به عبارتی می توان گفت که سیستم های پیشنهاد دهنده سیستم هایی هستند که در پیدا کردن و انتخاب نمودن آیتم های مورد نظر گره به آنها کمک می کنند. طبیعی است که این سیستم ها بدون در اختیار داشتن اطلاعات کافی و صحیح در مورد گره و آیتم های مورد نظر آنها (بعنوان مثال فیلم، موزیک، کتاب و ...) قادر به پیشنهاد دادن نمی باشند. بنابراین یکی از اساسی ترین اهداف آنها جمع آوری اطلاعات گوناگون در رابطه با سلایق گره و آیتم های موجود در سیستم است. منابع و روش های گوناگونی برای جمع آوری چنین اطلاعاتی وجود دارد. یک روش جمع آوری اطلاعات بصورت صریح که در آن کاربر صراحتا اعلام می کند که به چه چیز هایی علاقه دارد (بعنوان نمونه با امتیاز دادن (rate) به یک موزیک). روش دیگر روش ضمنی است که کمی دشوار تر است و در آن سیستم باید سلایق کاربر را با کنترل و دنبال کردن رفتار ها و فعالیت های او بیابد (بعنوان مثال باید ببینید کاربر به چه موزیک هایی بیشتر گوش می دهد، چه صفحاتی را بازدید می کند، با چه کسانی در ارتباط است و ...). علاوه بر اطلاعات ضمنی و صریح برخی از سیستم ها نیز هستند که از اطلاعات شخصی گره استفاده می کنند. بعنوان مثال سن، جنسیت و ملیت گره می تواند منبع خوبی برای شناخت کاربر و ارائه پیشنهاد به وی باشد. به این گونه از اطلاعات Demographic Information گفته می شود که گروهی از سیستم های پیشنهاد دهنده مبتنی بر همین اطلاعات بنا شده اند. با ظهور web و گسترش شبکه های اجتماعی در سال های اخیر محققین به منبع اطلاعاتی دیگری برای بهبود کیفیت پیشنهاد ها پی بردند که همان اطلاعات موجود در شبکه های اجتماعی بود و بر همین اساس کار های تحقیقاتی زیادی در این حوزه پایه ریزی گردید. اهمیت سیستم های پیشنهاد دهنده وقتی بیش از پیش مشخص می شود که گره با انبوهی از اطلاعات و آیتم ها مواجه باشند.

۲- تقسیمات مقاله

مراحل روش پیشنهادی سیستم خوشه بندی را به ترتیب در اینجا آورده‌ایم. از بین آن‌ها، بخش‌هایی که نیاز به توضیح و بسط بیشتری دارند، به طور جداگانه آورده خواهد شد.

تبدیل نقاط خام اولیه به نقاط توقف: نقاط خام اولیه شامل تعداد زیادی داده لحظه‌ای است که به طور متناوب ثبت شده است. منظور از نقاط توقف، محل‌هایی است که کاربر طی زمان زیادی (مثلا ۳۰ دقیقه) در شعاع محدودی (مثلا ۲۰۰ متر) سپری کرده است. با این تبدیل، حجم داده‌ها به طور چشمگیری کاهش می‌یابد و به حالت دنباله‌ای از نقاط در می‌آید.

- خوشه‌بندی چندلایه نقاط توقف برای یافتن مکان‌های مهم: برای یافتن مکان‌های مهم گره ، نقاط توقف آن‌ها را خوشه‌بندی کرده‌ایم. ما در این مرحله، برای اولین بار در سیستم‌های پیش‌بینی، از الگوریتم خوشه‌بندی سلسله مراتبی و از نوع پایین به بالا استفاده کردیم. این الگوریتم در پایین‌ترین سطح، هر نقطه را یک خوشه در نظر می‌گیرد. سپس در هر مرحله حرکت به سمت بالا، نزدیک‌ترین خوشه‌ها را با هم ترکیب می‌کند. سپس با در نظر گرفتن برش‌هایی از سطوح ساخته شده، مکان‌های کاربر تعیین می‌شود. در ادامه توضیحات بیشتری درباره این الگوریتم خواهیم داد.
- زدن برچسب مکانی به نقاط توقف و تبدیل دنباله اصلی نقاط توقف به دنباله‌های روزانه: نقاط توقف اولیه در این مرحله با استفاده از برچسب‌هایی که در خوشه‌بندی به‌دست آوردند، برچسب زده می‌شوند. سپس دنباله‌ی اصلی تبدیل به دنباله‌های روزانه می‌شوند تا برای ساخت الگوهای متداول و پیش‌بینی آماده شوند. نحوه‌ی تبدیل دنباله‌ها، برچسب زدن مکانی و فیلتر کردن دنباله‌ها، در ادامه شرح داده خواهد شد.
- ساخت دنباله‌های پرتکرار از دنباله‌های روزانه: به منظور استخراج الگوی حرکتی افراد، لازم است تا دنباله‌های پرتکرار را از روی دنباله‌های گذشته کاربر استخراج نماییم. این کار با الگوریتم مبتنی بر اپریوری -یعنی ساخت کاندیدا و شمارش آن‌ها- انجام می‌شود. با این تفاوت که به خاطر وجود سلسله‌مراتب روی مکان‌های موجود، هر مکان دارای روابطی با مکان‌های کوچکتر و بزرگتر از خود است. این موضوع تغییراتی در در الگوریتم به وجود می‌آورد که در ادامه به آن‌ها خواهیم پرداخت.
- ارایه معیار شباهت برای تطابق و انتخاب در پیش‌بینی: سیستم پیش‌بینی مکان آتی، برای پیش‌بینی موقعیت کاربر در دنباله جدید و فعلی، شباهت آن را با دنباله‌ها پرتکرار موجود در سیستم محاسبه می‌کند. سپس، از میان دنباله‌های متداول، آن‌هایی که شباهت بیشتری به دنباله فعلی دارند، انتخاب شده و از روی آن‌ها رای گیری برای آیتم بعدی انجام می‌شود. در این بین، انتخاب دنباله‌های متداول از سطوح مختلف هم دارای اولویت‌بندی خواهد بود.



شکل ۱: خوشه بندی سلسله مراتبی

در ابتدای الگوریتم به تعداد نقاط توقف خوشه داریم که هرکدام یک عضو دارند. فاصله دو خوشه را از فرمول زیر که میانگین فاصله‌های اعضای دو خوشه است محاسبه می‌کنیم که در آن، $|p - p'|$ به معنی فاصله اقلیدسی دو نقطه p' و p است.

$$Average_Dist(C1, C2) = \frac{1}{n1.n2} \sum_{p \in C1} \sum_{p' \in C2} |p - p'|$$

همان‌طور که گفته شد، در هر مرحله از خوشه‌بندی فاصله همه خوشه‌ها با یکدیگر محاسبه شده و از بین آن‌ها، آن زوج خوشه‌هایی که کمترین فاصله را دارند با هم ترکیب می‌شوند. این فاصله کمینه در هر سطح i به نام $(minz_i)$ در مکان‌یابی استفاده می‌شود. فرض کنید تعداد سطح‌های در نظر گرفته شده برای برش، k باشد. قبل از خوشه‌بندی یک لیست از مقادیر آستانه به طول k در نظر می‌گیریم. در طی خوشه‌بندی هرگاه فاصله $minz_i$ بیش از آستانه کنونی شود، خوشه‌های سطح i ام به عنوان یکی از برش‌های دلخواه انتخاب می‌شوند.

در انتهای این الگوریتم، خوشه‌هایی که هر نقطه توقف در سطوح مختلف متعلق به آن‌ها است، مشخص می‌شوند. به این ترتیب برچسب‌های مکانی نقاط توقف، که لیستی مرتب هستند، زده می‌شوند. همان‌طور که در شکل ۱ نشان داده‌شده است، خوشه‌های به‌دست آمده به خاطر روش ساخت خود از یک ساختار درختی پیروی می‌کنند. ما در این مرحله، فقط با استفاده از سطوح منتخب خوشه‌ها درختی می‌سازیم. این درخت

بعدا در استخراج توالی‌های پرتکرار استفاده می‌شود. در درخت ساخته شده، هریک از راس‌ها (خوشه‌ها) دارای یک پدر بوده و یک راس مشترک به نام ریشه پدر راس‌های بالاترین سطح خواهد بود. زدن برچسب مکانی به نقاط توقف و تبدیل دنباله اصلی نقاط توقف به دنباله‌های روزانه، در این مرحله، داده‌های موجود که دنباله‌ی نقاط توقف هستند، برچسب مکانی زده می‌شوند. همان‌طور که در مرحله خوشه‌بندی دیدیم، هر نقطه توقف می‌تواند در سطح‌های مختلف و حداکثر به تعداد کل سطح‌ها، عضو خوشه‌های مختلف باشد. بنابراین هر نقطه توقف به تعداد سطح‌های خود برچسب خواهد داشت و از آنجا که ترتیب این شناسه‌های مکانی هم اهمیت دارد، این شناسه‌ها در یک لیست مرتب به نام شناسه سطوح ذخیره می‌شود. سپس، لیست اصلی نقاط توقف را با استفاده از اطلاعات زمانی نقاط توقف به دنباله‌های روزانه تبدیل می‌کنیم. در این مرحله، لازم است تا دنباله‌های روزانه فیلتر شود زیرا در مواردی نقاط توقف متوالی دارای برچسب یکسان هستند و نیاز است در این موارد یکی از دو نقطه متوالی حذف شود. هم‌چنین دنباله‌های روزانه‌ای که در انتهای فیلترینگ، طولشان کمتر از ۲ شود، از میان دنباله‌های روزانه حذف می‌شوند.

دنباله‌های پرتکرار طبق الگوریتم کاوش دنباله های پرتکرار مبتنی بر اپریوری که قبلا اشاره شد ساخته می‌شوند. پارامتر «حداقل مقدار پشتیبان» به عنوان ورودی این الگوریتم داده می‌شود و ساخت کاندیدا و شمارش آنها برای طول های مختلف دنباله انجام می‌شود.

این الگوریتم ابتدا عناصر یکه را شمارش می‌کند و در واقع دنباله‌های کاندید به طول یک را می‌سازد؛ سپس کاندیدا را شمارش می‌کند و آنهایی که تکرارشان در روزهای مختلف از عدد حداقل مقدار پشتیبان بیشتر باشند، به عنوان دنباله‌های متداول به طول یک انتخاب می‌شوند. سپس از ترکیب دنباله‌های متداول به طول یک، کاندیدای به طول دو را می‌سازد و دفعات وجود آن‌ها را در دنباله‌های روزانه کاربر –با توجه به تعریف ج- شمارش می‌کند. این‌بار دنباله‌های متداول به طول دو ساخته می‌شود. این کار تا زمانی ادامه می‌یابد که در یک مرحله نتوان هیچ کاندیدی ایجاد کرد. ساختار کلی برنامه ی اجرایی باید فعالیت ویروسی را پیش برده و موضوعات اخلاقی و قانونی را مورد توجه قرار دهد. در مواردی که برای مثال موضوعات اخلاقی مورد نظر قرار نگیرند، ممکن است برنامه اجرایی ویروسی به یک پیامد منفی برای سازمان اقدام کننده منجر شود. عامل مهم دوم به خصوصیات خدمات یا محصولی که باید بازاریابی شود مرتبط است. برخی از محصولات و خدمات برای برنامه های اجرایی بازاریابی ویروسی نسبت به برخی دیگر از خدمات مناسبتر هستند. عنصر مهم دیگر پیامی است که منتقل میشود. پیامهایی که باعث پرورش تخیل شده و برای دریافت کننده ایجاد سرگرمی کرده و یا حتی باعث فریب دریافت کننده میشوند، برای ارسال به طور داوطلبانه محتملتر هستند. به طور کلی یک پیام باید به طور مؤثری دریافت کننده را مجذوب کند تا بتواند وی را به یک انتقال دهنده تبدیل کند.

۳- قواعد نوشتاری

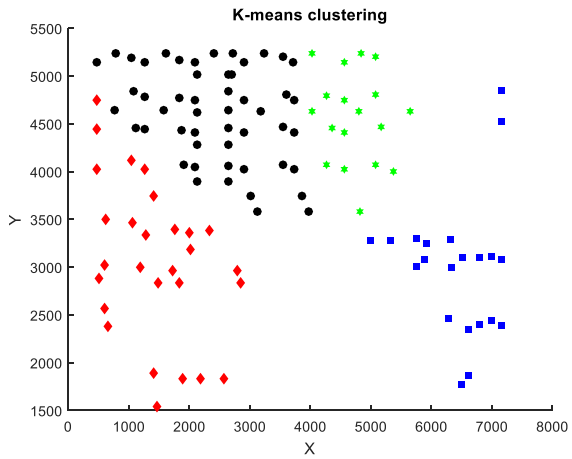
روش خوشه‌بندی بر مبنای نمونه اولیه، نوعی از خوشه‌بندی است که در آن خوشه مجموعه‌ای از اشیاء است که هر شیء به نمونه اولیه نزدیک است (بیشتر شبیه است) که این نمونه اولیه خوشه را در میان خوشه‌های دیگر تعریف می‌کند و درواقع هر خوشه توسط نمونه اولیه اش شناخته می‌شود. برای داده‌های پیوسته نمونه اولیه اغلب نقطه مرکزی (میانه) تمام داده‌های موجود در خوشه می‌باشد. در مورد داده‌های دیگر که در آنها نقطه مرکزی بی معنی به نظر می‌رسد، نمونه اولیه حد متوسطی از داده‌هاست که نماینده داده‌های آن خوشه می‌باشد.

در این مقاله از روش سلسه مراتبی و الگوریتم تکاملی برای بهبود دقت خوشه بندی در شبکه های پیچیده پویا استفاده می کنیم. و K تعداد خوشه های ما می باشد که از دو شروع می کنیم و به تدریج بیشتر می کنیم تا حدی که به یک میزان قابل قبول در خوشه بندی و شناسایی نویزهای موجود برسیم.

دوماً تکنیک K-means یک روش خوشه‌بندی بر مبنای نمونه اولیه است. روش خوشه‌بندی بر مبنای نمونه اولیه نوعی از خوشه‌بندی است که در آن خوشه مجموعه‌ای از اشیاء است که هر شیء به نمونه اولیه نزدیک است (بیشتر شبیه است) که این نمونه اولیه، نماینده هر خوشه را در میان خوشه‌های دیگر تعریف می‌کند و درواقع هر خوشه توسط نمونه اولیه اش شناخته می‌شود. نمونه اولیه خوشه ها در این تحقیق مختصات یک مکان عمومی فرضی است که در آن چگالی داده ها بیشتر است. برای داده‌های پیوسته نمونه اولیه اغلب نقطه مرکزی (میانه) تمام داده‌های موجود در خوشه می‌باشد. درمورد داده‌های غیرعددی که در آنها نقطه مرکزی بی معنی به نظر می‌رسد، نمونه اولیه حد متوسطی از داده‌هاست که نماینده داده‌های آن خوشه می‌باشد.

با توجه به ماهیت داده‌های استفاده شده در این تحقیق و همچنین نحوه استفاده از خوشه بندی در روش پیشنهادی، روش خوشه بندی K-means در مقایسه با سایر روش‌های خوشه بندی مانند خوشه بندی

تجمیعی* یا خوشه بندی تقسیمی†، برای کاربرد در این تحقیق مناسب تر باشد. روش خوشه بندی تجمیعی با n خوشه یک عضوی برابر با تعداد داده ها شروع کرده و در ادامه محصولات مشابه را در K خوشه با تعداد متغیر از داده ها ادغام می‌کند.هر داده یک خوشه است و چگالی داده ها در نظر گرته نمی شود. روش خوشه بندی تقسیمی نیز با یک خوشه کلی n عضوی شامل تمام داده ها شروع کرده و به K خوشه شامل نقاط مشابه می‌رسد. همچنین زمان انجام محاسبات برای خوشه بندی در روش K-means در مقایسه با روش‌های خوشه بندی خوشه بندی تجمیعی یا خوشه بندی تقسیمی کمتر بوده و به تعداد نقاطی که بالاترین چگالی را در بین نقاط دارند (m<n) بستگی دارد. بنابراین می‌توان گفت روش خوشه بندی K-means درمیان سایر روش‌های خوشه بندی برای کاربرد در این تحقیق مناسب تر است. روش خوشه‌بندی K-means خود دارای چندین تکنیک است که دو نمونه از آن‌ها که بیشتر برجسته است K-means و K-medoid می‌باشد. K-means معرف یک نمونه اولیه به عنوان نقطه مرکزی است و عموماً روی داده‌های پیوسته با n بعد اعمال می‌شود. K-medoid معرف نمونه اولیه به عنوان حد متوسط از داده‌هاست که نماینده بیشتر نقاط موجود در گروه می‌باشد و نظر به اینکه فقط نیازمند مقدار نزدیکی به نقاط است، می‌تواند در حیطه وسیع مورد استفاده قرار گیرد. با این وصف، یک نقطه مرکزی اکثراً با نقاط داده‌ای واقعی ارتباط زیادی ندارد، ولی یک حد متوسط با این تعریف، باید نقاط داده‌ای واقعی باشد. روش خوشه‌بندی K-means ساده است. در ابتدا K نقطه مرکزی اولیه انتخاب می‌شود، که K یک پارامتر تعیین شده توسط کاربر است، به این معنی که، K برابر با تعداد خوشه‌های مطلوب است. سپس هر نقطه به نزدیک ترین نقطه مرکزی اختصاص می‌یابد وهر مجموعه از نقاط اختصاص یافته به یک خوشه اختصاص می‌یابد. سپس نقطه مرکزی هر خوشه با اختصاص هر نقطه به آن خوشه به روزرسانی می‌شود. به روزرسانی مراکز خوشه به صورت یافتن نقطه مرکزی جدید با استفاده از میانه نقاط جدید اختصاص یافته به خوشه انجام می‌شود. این قدم‌های تخصیص و به روز رسانی تا زمانی که هیچ یک از نقاط مرکزی درخوشه‌ها تغییر نکند و با تخصیص هر نقطه، نقاط مرکزی یکسانی برگردادنده شود و یا تمامی‌نقاط به خوشه‌ها تعلق یابند، ادامه می‌یابد. درشکل ۴-۲ خوشه بندی نقاط موجود در ناحیه مشخص شده با استفاده از الگوریتم خوشه بندی K-means نشان داد شده است.



شکل ۲:خوشه بندی نقاط موجود در ناحیه مشخص شده

همانطور که در شکل ۲ نشان داده شده است، نقاط موجود در ناحیه بر اساس چگالی نقاط به چهار خوشه تقسیم شده است که هر یک از خوشه ها نشان دهنده یک مکان عمومی می باشد. نقاط تخصیص یافته به هر یک از مکان های عمومی با رنگ متمایزی مشخص شده اند. پس از خوشه بندی مطابق روش پیشنهادی به ایجاد توصیه هایی برای گره شبکه های اجتماعی به منظور بازدید از مکان های عمومی محبوب، با توجه به بازدید گره گذشته از این مکان ها خواهیم پرداخت. بدین منظور از الگوریتم های Apriori و FP-Growth که از روش های تکنیک کاوش قوانین انجمنی می باشد، استفاده شده است. کاوش قوانین انجمنی با توجه به حضور گره در مکان های عمومی به تصمیم‌گیری در مورد ارتباط بین مکان های عمومی و ایجاد توصیه برای گره شبکه اجتماعی خواهد پرداخت. به همین منظور ابتدا باید گراف حرکت مربوط به گره شبکه اجتماعی که داده های مربوط به آن ها در دوره زمانی گذشته جمع آوری شده است رسم شود.

۴- نتیجه

با استفاده از داده های مجموعه تست، به ارزیابی میزان دقت الگوهای ایجاد شده برای هر یک از کارمندان می پردازیم. نتیجه ارزیابی در جدول ۵-۱ نشان داده شده است.

* Agglomerative Clustering

† Divisive Clustering

جدول ۱-ارزیابی دقت الگوریتمهای رده بندی اعمال شده

الگوریتم اعمال شده	Accuracy	Precision	Recall
روش مقاله	۹۲,۳٪	۹۲,۸٪	۹۳,۱٪
جنگل تصادفی	۹۱,۸٪	۹۱,۹٪	۹۲٪
شبکه های عصبی	۸۹,۱٪	۸۹,۵٪	۸۹,۹٪
Naïve Bays	۹۰,۱٪	۹۰,۵٪	۹۰,۸٪
درخت تصمیم	۸۵,۵٪	۸۶,۱٪	۸۶,۹٪
Rough Set	۸۶,۳٪	۸۶,۶٪	۸۷,۱٪

همانطور که در جدول ۱ مشاهده می شود روش ارائه شده در این مقاله بالاترین میزان دقت را در تشخیص الگوی استفاده از شبکه های پیچیده پویا دارا می باشد.

.

مراجع

[۱] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, ۲۰۱۷, ۱۷(٦): ۷۳٤-۷٤۹.

[۲] Brusilovsky, Peter, and David N. Chin. Preface to the special issue on personalization in social web systems. User Modeling and User-Adapted Interaction, ۲۰۱۳, ۲۳(۲-۳): ۸۳-۸۷.

[۳] Jamali M and Ester M. A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks. In Proceedings of the ٤th ACM Conference on Recommender Systems, Barcelona, Spain, September ۲٦-۳۰, ۲۰۱٦: ۱۳٥-۱٤۲.

[٤] Massa P and Avesani P. Trust-aware collaborative filtering for recommender systems. Lecture Notes in Computer Science, ۲۰۱٤(۳۲۹۰): ٤۹۲-٥۰۸.

[٥] Ozsoy M G, Polat F. Trust based recommendation systems. In Proceedings of the ۲۰۱۳ IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM'۱۳, Niagara, ON, Canada, August ۲٥-۲۹, ۲۰۱۳: ۱۲٦۷-۱۲۷٤.

[٦] Massa P and Avesani P. Trust metrics in recommender systems. Computer with Social Trust, ۲۰۱٥-۲٥۹ : ۲۸٥.

[۷] Bedi P, Kaur H and Marwaha S. Trust based recommender system for semantic web. Artificial Intelligence, ۲۰۱۷: ۲٦۷۷-۲٦۸۲.

[۸] Nazemian, A., Gholami, H., and Taghiyareh, F. An improved model of trust-aware recommender systems using distrust metric. In Proceedings of the ۲۰۱۲ IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, Turkey, August , ۲٦-۲۹ .۱۰۷۹-۱۰۸٤ :۲۰۱۲

[۹] Ma H, Yang H, Lyu M R, and King I. SoRec: Social recommendation using probabilistic matrix factorization. In Proceedings of the ۱۷th ACM Conference on Information and Knowledge Management, Napa Valley, California, USA, October ۳۰-۲٦, ۲۰۰۸: ۹۳۱-.۹٤۰

[۱۰] Ma H, Zhou D, and Liu C, et. al. Recommender systems with social regularization. In Proceedings of the ٤th ACM International Conference on Web Search and Data Mining, WSDM , ۲۰۱۱Hong Kong, China, February ۹-۱۲, ۲۰۱۱: .۲۹٦-۲۸۷