



انجمن علمی تجارت الکترونیکی ایران

فصل نامه علمی - آموزشی تجارت الکترونیکی

شماره هفتم، پاییز ۱۳۹۵

گراف کاوی

و کاربرد آن در شبکه‌های اجتماعی

۱- گراف

گراف مجموعه‌ای از گره‌ها و ارتباطات بین آن‌ها است. گراف‌ها ابزار مناسبی برای نمایش داده‌ها و ارتباطات بین آن‌ها به شمار می‌روند. در این ساختار گره‌ها بسته به نوع کاربرد می‌توانند بیان‌گر انواع مختلفی از داده‌ها مانند افراد، صفحات وب، مقالات علمی و یا هر موجودیت مورد مطالعه دیگر باشند؛ ارتباطات نیز می‌توانند شامل ارتباط دوستی، ارجاعات لینک‌ها یا هر نوع ارتباط دیگر بین گره‌ها باشد. ارتباط بین سامانه‌های شبکه‌ای و تئوری گراف ارتباطی ناگسستگی است و گراف‌ها اهمیت زیادی در مدل‌سازی ساختارهای پیچیده‌ای چون ترکیبات شیمیایی، ساختارهای پروتئین‌ها، شبکه‌های زیستی، شبکه‌های اجتماعی، اسناد اکس‌ام‌ال^۱ و بسیاری از موارد دیگر دارند.



۲- گراف کاوی

کشف دانش از ساختار گراف، زمینه‌ی جدیدی را برای کاوش داده‌ها، به نام گراف کاوی^۲ به وجود آورده است. با افزایش نیاز و تقاضا برای تحلیل مقادیر زیادی از داده‌ها، گراف کاوی به یک موضوع پویا و فعال در حوزه‌ی داده کاوی با رویکرد کاوش بر روی داده‌های انبوه ساختاریافته تبدیل شده است. در سال‌های اخیر تحقیقات زیادی روی گراف کاوی با هدف رسیدن به کارایی بالا و ابداع روش‌های جدید، برای مطالعه و کاوش داده‌های ساختار یافته انجام شده است. کاوش بر روی ساختار داده‌ها یکی از چالش‌های جدیدی است که محققان با آن روبه‌رو هستند. کاربردهای گراف کاوی را می‌توان در حوزه‌های مختلفی چون: شبکه‌های اجتماعی، شبکه‌های کامپیوتری، دسته‌بندی ترکیبات شیمیایی، شبکه‌های زیستی، علوم پزشکی، طراحی سیستم‌های توصیه‌گر و بسیاری از موارد دیگر مشاهده کرد که در ادامه برخی از این کاربردها با جزئیات بیشتر ارائه خواهد شد.

• شبکه‌های اجتماعی

یکی از مهمترین کاربردهای گراف کاوی شناسایی انجمن‌ها^۳ یا جوامع در شبکه است که اطلاعات زیادی را درباره‌ی ساختار یک شبکه ارائه می‌دهند و می‌توان آن‌ها را جوامع تشکیل دهنده‌ی شبکه نامید. به عنوان مثال این جوامع نشان می‌دهند که چه افرادی، با چه خصوصیتی تمایل بیشتری به برقراری ارتباط با هم دارند و بالعکس چه افرادی با چه ویژگی‌هایی ارتباطات ضعیف و گسستگی با هم دارند؛ چه گروه‌هایی در یک شبکه می‌توانند سایر افراد را با

^۱ XML

^۲ Graph Mining

^۳Community Detection

سرعت بیشتری تحت تأثیر قرار دهند و بسیاری از موارد دیگر. تاکنون شناسایی انجمن در انواع مختلفی از شبکه‌ها، از جمله شبکه‌های رسانه‌های اجتماعی، شبکه‌های وب، شبکه‌های ترکیبات شیمیایی، شبکه‌های عصبی، شبکه‌های حمل و نقل و ده‌ها شبکه‌ی دیگر با هدف استخراج اطلاعات برای تحلیل و بررسی عناصر این شبکه‌ها صورت گرفته است. این حوزه از کاربردها جزء مسائل خوشه‌بندی قرار می‌گیرند.



• تحلیل سبد خرید مشتریان

یکی از مسائلی که اخیراً فروشگاه‌ها به آن توجه زیادی کرده‌اند یافتن فهرست کالاهایی است که احتمال خریدشان با هم توسط مشتریان زیاد است. این مسأله در بهبود رضایت مشتری و افزایش سود حاصل از فروش محصولات در فروشگاه‌ها تأثیر عمده‌ای دارد. با استفاده از روش‌های مؤثر و کارآمد در علم گراف‌کاوی می‌توان در هر فروشگاه کالاها و محصولات را که احتمال خریدشان با هم زیاد است، خوشه‌بندی کرد و برای طراحی سیستم‌های توصیه‌گر از آن‌ها استفاده نمود.



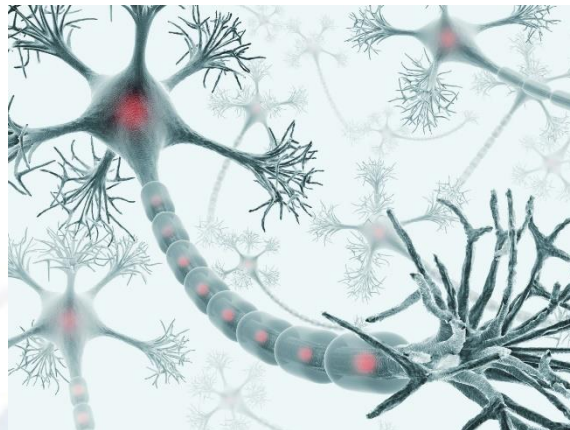
†Social Media

‡Clustering



- **شناسایی سرطان و تشخیص بیماری**

در علوم پزشکی و دارویی با استفاده از مجموعه داده مربوط به ترکیبات شیمیایی و استفاده از روش‌های دسته‌بندی، گراف و کاوش زیرگراف‌های مکرر می‌توان ترکیبات شیمیایی سرطان‌زا و ناقل بیماری را شناسایی کرد و از آنها برای تولید دارو و درمان بیماری استفاده کرد.



- **شناسایی ارتباط‌های هزرنامه در خروجی موتورهای جست‌وجو**

در سطح شبکه‌ی وب، با مجموعه‌ای گسترده، متنوع و پویا از داده‌ها روبه‌رو هستیم. موتورهای جست‌وجو یک راه مؤثر برای دسترسی به اطلاعات مفید از میان این حجم وسیع از داده‌ها به شمار می‌روند. از طرفی وجود هزرنامه‌ها در وب، موتورهای جست‌وجو را گمراه می‌کنند و مانع از دسترسی کاربران به اطلاعات مفید و مرتبط خواهند شد. شناسایی و فیلتر این هزرنامه‌ها با استفاده از روش‌های دسته‌بندی درگراف‌کاوی می‌تواند صورت بگیرد.



کاربردهای ارائه شده تنها چند مورد از کاربردهای بی‌شمار گراف‌کاوی بر روی شبکه‌های موجود در دنیای واقع است که هر کدام ذره‌ای از ضرورت و اهمیت توجه به این علم را نشان می‌دهد.

۲-۱- اهمیت و کاربرد روش‌های گراف کاوی

تحلیل شبکه‌های اجتماعی (SNA)؛ شامل مطالعه‌ی روابط بین افراد، تحلیل ساختارهای اجتماعی، موقعیت اجتماعی و تحلیل نقش افراد و بسیاری از موارد دیگر می‌باشد. در این شبکه‌ها روابط بین افراد از جمله دوستی، خویشاوندی، همکاری و غیره را می‌توان در قالب یک گراف نشان داد، که در آن افراد گره‌ها و ارتباطات بین آن‌ها یال‌ها را نشان می‌دهند. مطالعه و بررسی گراف مربوط به شبکه‌های اجتماعی اطلاعات زیادی را در درباره مفاهیم، علل اصلی شکل‌گیری این شبکه‌ها، نحوه برقراری ارتباط و بسیاری از موارد دیگر را فراهم می‌کند. مهمترین سوالی که در حال حاضر برای شبکه‌های اجتماعی مطرح است، این است که چگونه می‌توان جوامع را در این شبکه‌ها پیدا کرد به گونه‌ای که این جوامع به درستی عضویت افراد را در گروه‌های مختلف نشان بدهد. از مهمترین دلایل دیگری که استفاده از روش‌های گراف کاوی را برای تحلیل شبکه‌های بزرگ مقیاس و از جمله شبکه‌های اجتماعی ضروری کرده است قابلیت این روش‌ها در شناسایی الگوها و هنجارهای شبکه، مدیریت، کشف ساختارهای پنهان و استخراج اطلاعات مفید از میان حجم انبوهی از داده‌هایی است که در قالب گراف نمایش داده می‌شود. اطلاعاتی که بدون استفاده از روش‌های گراف از میان این حجم انبوه داده‌ها قابل استخراج نیست.

۲-۲- دامنه‌های گراف کاوی

با توجه به رشد سریع شبکه‌ها، علم گراف کاوی به سرعت در حال توسعه و پیشرفت است. به گونه‌ای که هر جا نامی از شبکه به میان می‌آید نمودی از کاربردهای این علم را می‌توان دید، برخی از زیرحوزه‌هایی که گراف کاوی آن‌ها را پوشش می‌دهد عبارت است از:

- مسائل شناسایی زیرگراف‌های همسان
- شاخص‌گذاری و ویدئو
- کاوش الگوهای همبسته‌ی گراف
- کاوش الگوهای بهینه‌ی گراف
- کاوش جریان کار
- کاوش شبکه‌های زیستی
- کاوش ساختار وب
- هسته‌های گراف
- کاوش ارتباطات
- جست‌وجو گراف
- دسته‌بندی گراف
- خوشه‌بندی گراف
- خلاصه‌سازی الگوهای گراف
- و... بسیاری از حوزه‌های دیگر.

۶ Social Network Analysis

۷Video Indexing

همه‌ی حوزه‌های گفته شده را می‌توان در قالب سه محور زیر تجمیع و خلاصه نمود:

۱. دسته بندی گراف
۲. خوشه‌بندی گراف
۳. کاوش زیرگراف

هدف اصلی مسائل دسته‌بندی گراف آن است که گره‌های مربوط به یک گراف با مقیاس بزرگ و یا مجموعه‌ای از گراف‌های داده شده به عنوان موجودیت‌های مورد بررسی را بر اساس ویژگی‌های مربوط به گره‌ها و یا ساختار گراف به کلاس‌های مختلف دسته‌بندی کرد. مسائل حوزه‌ی داده‌کاوی در دو غالب کلی مسائل یادگیری با ناظر^۱ و یادگیری بدون ناظر^۲ مطرح می‌شوند. دسته‌بندی، به طور کلی جزء مسائل باناظر طبقه‌بندی می‌شود و کاربرد این مسائل را می‌توان در علوم پزشکی، علوم دارویی، دسته‌بندی ترکیبات شیمیایی، دسته‌بندی مشتریان و موارد مشابه مشاهده کرد. در مسائل خوشه‌بندی نیز به دنبال آن هستیم تا گره‌های یک گراف با مقیاس بزرگ و یا مجموعه‌ای از گراف‌ها را به عنوان موجودیت‌های مورد بررسی خوشه‌بندی کنیم، این روش‌ها بر اساس یادگیری بدون ناظر صورت می‌گیرد و از مهم‌ترین کاربردهای آن به شناسایی جوامع در شبکه‌های مختلف می‌توان اشاره کرد و در نهایت در مسائل کاوش زیرگراف‌ها در یک گراف، پیدا کردن مجموعه‌ای از الگوها که هر کدام ممکن است نشان‌دهنده‌ی ویژگی خاصی در گراف باشد، مورد بررسی و مطالعه قرار می‌گیرد.

^۱Supervised Learning

^۲Unsupervised Learning



انجمن علمی تجارت الکترونیکی ایران

شماره‌های تماس: ۸۸۹۹۱۵۶۰ - ۸۸۹۹۱۵۴۰

وبسایت: www.ieca.ir

ایمیل: info@ieca.ir