فصلنامهٔ علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

سال دهم، شماره‌های ۳۷ و ۳۸، پاییز و زمستان ۱۳۹۷
صص: ۴۵- ۷۰

یادگیری رتبه‌بندی محتوای فارسی وب بر مبنای برنامه‌نویسی ژنتیک چند لایه

امیر حسین کیهانی پور
استادیار، گروه مهندسی کامپیوتر، دانشکده مهندسی، پردیس فارابی، دانشگاه تهران

تاریخ دریافت:۱۳۹۸/۱۲/۱۵ تاریخ پذیرش: ۱۳۹۹/۰۳/۲۱

چکیده

یادگیری رتبه‌بندی، یک رویکرد نو ظهور به منظور رفع چالش‌های موجود و بهبود عملکرد جویشگرهای وب، بسیار امید بخش و کارآمد است. در عین حال عدم توجه جدی به سوابق تعاملات کاربران با جویشگر طی فرآیند جستجو و ارزیابی نتایج بدست آمده، یکی از معضلات جدی آن بشمار می‌رود. در عین حال حجم بسیار زیاد ویژگی‌های مورد نیاز از اسناد و پرس‌وجوهای کاربران نیز کاربردی بودن این رویکرد را در شرایط واقعی با ابهام مواجه ساخته است. استفاده از مدل اطلاعات کلیک از گذر داده‌ها و تولید ویژگی‌های کلیک از گذر داده، راهکار نوینی است که بر مبنای آن و با بکارگیری مدل برنامه‌نویسی ژنتیک چند لایه، مدل رتبه‌بندی مناسبی تحت عنوان MGP-Rank برای بازیابی اطلاعات انگلیسی وب، عرضه شده است. در این پژوهش این، با عنایت به ویژگی‌های خاص زبان فارسی، از طریق ارائه سناریوهای مناسب برای ایجاد ویژگی‌های کلیک از گذر داده این الگوریتم، این الگوریتم بومی‌سازی شده است. نتایج حاصل از ارزیابی عملکرد این الگوریتم در حوزه زبان فارسی با استفاده از مجموعه داده dotIR، حاکی از توانمندی قابل ملاحظه آن نسبت به روش‌های مرجع رتبه‌بندی اطلاعات است. این بهبود عملکرد، بخصوص در بخش ابتدایی فهرست نتایج جستجو که غالباً بیشتر مورد مراجعه کاربران است، قابل توجه است.

واژه‌های کلیدی: یادگیری رتبه‌بندی، مدل برنامه‌نویسی ژنتیک چند لایه، ویژگی‌های کلیک از گذر داده، محتوای فارسی وب، مجموعه داده dotIR

۱- مقدمه

استفاده گسترده کاربران محیط وب از جویشگرها به عنوان یکی از مهم‌ترین و پرکاربردترین سامانه‌های موجود در این محیط، موجب گردیده است تا پژوهش‌های زیادی به منظور تقویت و بهبود فرآیند جستجو و رتبه‌بندی اسناد مرتبط با پرس‌وجوهای کاربران، صورت گیرد. گرچه تحقیقات در زمینه انجام رتبه‌بندی مناسب نتایج جستجوهای کاربران جویشگرها، پیشینه قابل توجهی دارد، ولی در سال‌های اخیر با کاربرد گسترده‌تر این سامانه‌ها توسط کاربران، حجم قابل توجهی از اطلاعات مربوط به سوابق تعاملات کاربران با این سامانه‌ها، گردآوری شده است و این امر، امکان استفاده از الگوریتم‌های یادگیری ماشینی را برای ایجاد مدل‌های رتبه‌بندی کارآمد، فراهم نموده است. بطور مشخص،

نویسندهٔ عهده‌دار مکاتبات: امیر حسین کیهانی پور keyhanipour@ut.ac.ir

یادگیری رتبه‌بندی، به مجموعه روش‌های یادگیری ماشینی اطلاق می‌شود که در عملیات رتبه‌بندی، به انجام یادگیری و تشخیص مدل مناسب جهت رتبه‌بندی اطلاعات، می‌پردازند [۲۸]. یادگیری رتبه‌بندی، کاربردهای متعددی در زمینه بازیابی اطلاعات[1]، پردازش زبان طبیعی[2] و نیز داده‌کاوی[3] دارد. در تحقیقات انجام شده در این خصوص، به این نکته توجه شده است که کاربران جویشگرها که بطور مستقیم با این سامانه‌ها درگیر هستند، قادر به ارایه بهترین قضاوت‌ها در مورد کیفیت عملکرد این سامانه‌ها می‌باشند. با این وجود، عدم تمایل اکثریت کاربران به ارایه قضاوت صریح از عملکرد جویشگرهای مورد استفاده، طراحان این سامانه‌ها را بر آن داشته است تا از سوابق تعاملات کاربران طی جستجوهای بعمل آمده، که اصطلاحاً «اطلاعات کلیک از گذر داده[4]»، نامیده می‌شود، به منظور بررسی و بهبود عملکرد این سامانه‌ها بهره گیرند [۲۱]. مطالعات انجام شده، نشان می‌دهد که استفاده از اطلاعات کلیک از گذر داده، می‌تواند منجر به بهبود عملکرد جویشگرهای وب، شود [۳۰] و [۲۲] و [۳۱]. علیرغم این موضوع، فقدان اطلاعات کلیک از گذر داده در اغلب مجموعه‌های داده محک موجود در زمینه یادگیری رتبه‌بندی، موجب شده است تا اکثر الگوریتم‌های ارایه شده در این حوزه، از این اطلاعات ارزشمند، بی‌بهره باشند. در عین حال، استفاده اغلب روش‌های موجود از تعداد زیادی از ویژگی‌های اسناد و پرس‌وجوها، کاربرد آنها را در شرایط واقعی، دشوار می‌سازد.

در ادامه این نوشتار، ابتدا به مرور تحقیقات مرتبط پرداخته خواهد شد و بصورت تحلیلی، ضرورت انجام این پژوهش، مطرح خواهد شد. پس از آن، معرفی الگوریتم MGP-Rank صورت خواهد گرفت و نحوه عملکرد آن بیان خواهد شد. سپس نتایج حاصل از ارزیابی عملکرد الگوریتم یادگیری رتبه‌بندی MGP-Rank در مقایسه با روش‌های رتبه‌بندی مرجع، بر اساس شاخص‌های ارزیابی فرآیند بازیابی اطلاعات، بصورت تحلیل مورد قیاس قرار خواهد گرفت. بخش پایانی این نوشتار نیز به جمع‌بندی این پژوهش و نیز معرفی حوزه‌های تحقیقاتی آتی، اختصاص یافته است.

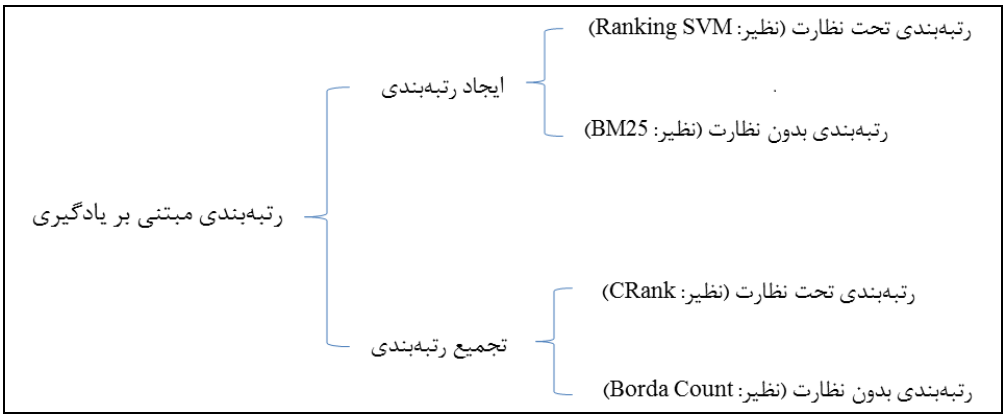
۲- مرور تحقیقات مرتبط

با توجه به نقش محوری رتبه‌بندی نتایج حاصل از جستجو در جویشگرهای وب، در سالهای گذشته تحقیقات گسترده‌ای در این در این زمینه صورت پذیرفته است. بصورت ویژه طی سالیان اخیر، موضوع یادگیری رتبه‌بندی، به عنوان یک روش نوین برای انجام رتبه‌بندی با استفاده از تکنیک‌های یادگیری ماشینی، مطرح شده است و در واقع، یادگیری رتبه‌بندی، به صورت ترکیبی از یادگیری ماشینی، بازیابی اطلاعات و نیز پردازش زبان طبیعی قابل توصیف است. بطور کلی، دو تعریف عمده برای مفهوم یادگیری رتبه‌بندی، ارائه شده است. در نگرش کلان، یادگیری رتبه‌بندی به هر نوع کاربرد تکنیک‌های یادگیری ماشینی در مساله رتبه‌بندی، اطلاق می‌شود. در مقابل و در یک نگرش جزیی‌تر، یادگیری رتبه‌بندی به آن دسته از روش‌های یادگیری ماشینی اشاره دارد که برای ایجاد مدل‌های رتبه‌بندی در مسایل «ایجاد رتبه‌بندی[5]» و «تجمیع رتبه‌بندی[6]»، مورد استفاده قرار می‌گیرند. در [۲۸]، یک دسته‌بندی کلی از انواع روش‌های مطرح در زمینه یادگیری رتبه‌بندی، در شکل شماره یک ارائه شده است.

در هر دو دسته تکنیک‌های یادگیری رتبه‌بندی را می‌توان جزء روش‌های یادگیری ماشینی تحت نظارت، بشمار آورد. بر این اساس، دادگان پایه یادگیری مشتمل بر مجموعه‌ای از اسناد و پرس‌وجوها و نیز قضاوت‌های کاربران خبره در مورد با میزان مرتبط بودن این اسناد و پرس‌وجوها در اختیار

[1] Information retrieval
[2] Natural Language Processing (NLP)
[3] Data mining
[4] Click-through data
[5] Rank Creation
[6] Rank Aggregation

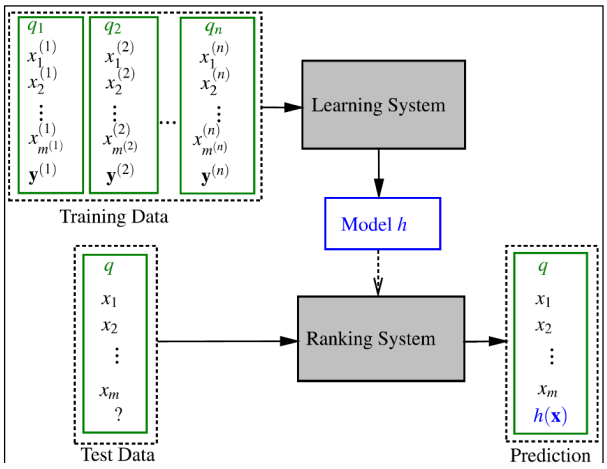
است. تولید مدل بر اساس قضاوت کاربران در رابطه با میزان مرتبط بودن زوج‌های اسناد - پرس‌وجوها صورت می‌گیرد. در روش‌های ایجاد رتبه‌بندی، بر اساس مجموعه داده یادگیری، مدلی به منظور رتبه‌بندی نتایج حاصل از پرس‌وجوهای کاربران، تولید می‌شود. در مقابل، در روش‌های تجمیع رتبه‌بندی، نتایج حاصل از رتبه‌بندی تعدادی از روش‌های رتبه‌بندی پایه، جهت تولید یک مدلی تجمیعی، با یکدیگر ادغام می‌شوند. بصورت خلاصه، می‌توان گفت در فرآیند «ایجاد رتبه‌بندی»، تهیه لیست مرتب، بر اساس ویژگی‌های درخواست و پیشنهادهای ممکن، صورت می‌گیرد، در حالی که در عملیات «تجمیع رتبه‌بندی»، از رتبه‌بندی‌های اولیه صورت گرفته، استفاده می‌شود. معمولاً هرگاه به مساله یادگیری رتبه‌بندی اشاره شود، منظور، ایجاد رتبه‌بندی بر اساس یادگیری تحت نظارت است [۲۸]. الگوریتم ارایه شده در این پژوهش یعنی MGP-Rank نیز از همین دسته روش‌های یادگیری رتبه‌بندی، بشمار می‌رود.



شکل ۱: طبقه‌بندی انواع روش‌های یادگیری رتبه‌بندی (Liu 2011)

شمای کلی عملیات روش‌های یادگیری رتبه‌بندی، در شکل شماره دو نشان داده شده است. بر اساس این ساختار، از آنجا یادگیری رتبه‌بندی، گونه‌ای از یادگیری تحت نظارت می‌باشد، نیاز به مجموعه داده آموزشی دارد. فرآیند تولید مجموعه داده آموزشی، شباهت زیادی با ایجاد داده آموزشی برای ارزیابی دارد. یک مجموعه داده آموزشی، عموماً شامل موارد زیر است:

- تعداد n پرس‌وجوی آموزشی $q_i (i = 1, \dots, n)$،
- مجموعه اسناد متناظر با آنها که به صورت بردار ویژگی‌ها نمایش داده می‌شوند $x^{(i)} = \left\{x_j^{(i)}\right\}_{j=1}^{m^{(i)}}$ (که در آن، $m^{(i)}$ برابر با تعداد اسناد متناظر پرس وجوی q_i است)، و
- قضاوت‌های میزان مرتبط بودن اسناد با پرس‌وجوها.



شکل ۲: شمای کلی چارچوب روش‌های رتبه‌بندی مبتنی بر یادگیری (Liu 2011)

بر این اساس، یک الگوریتم یادگیری به منظور تولید مدل رتبه‌بندی، استفاده می‌شود، به نحوی که این مدل، قادر باشد با توجه به تابع زیان[7] مورد استفاده، میزان مرتبط بودن اسناد به پرس‌وجوها را با دقت بالا و حتی‌المقدور مشابه قضاوت‌های انسانی، پیش‌بینی نماید. در خلال فاز آزمون و با دریافت یک پرس‌وجوی جدید، مدل تولید شده طی فاز آموزش، روی مجموعه اسناد مورد نظر، اِعمال می‌شود تا فهرست رتبه‌بندی شده از آن اسناد، ایجاد شود. بر اساس این رویکرد، روش‌های یادگیری رتبه‌بندی را می‌توان به روش‌های نقطه‌ای[8]، روش‌های جفتی[9] و نیز روش‌های مبتنی بر فهرست[10]، دسته‌بندی نمود. هر دسته از این روش‌ها، فرآیند یادگیری رتبه‌بندی را به اَشکال مختلف، مدل‌سازی می‌کنند. به عبارت دیگر، هر کدام از این انواع روش‌ها، تعاریف متفاوتی از فضاهای ورودی و خروجی بیان می‌کنند یا ممکن است از فرضیه‌های متفاوتی استفاده کنند و نیز توابع زیان متفاوتی را بکار گیرند.

بطور کلی، روش‌های نقطه‌ای، وابستگی بین اسناد را در نظر نمی‌گیرند و بنابراین موقعیت یک سند در لیست مرتب نهایی، در محاسبات، لحاظ نمی‌شود. فضای ورودی در روش‌های جفتی، زوج‌های اسناد می‌باشد که به‌صورت بردارهای ویژگی‌ها، نمایش داده می‌شوند. فضای خروجی نیز شامل اولویت بین این زوج‌های اسناد است که دامنه مقادیر آن از مجموعه $\{+1,-1\}$ می‌باشد. در برخی از روش‌های ارائه شده در این دسته، رتبه‌بندی به مساله رگرسیون، تقلیل می‌یابد [۱۰] و [۱۵] و [۴۵]. یک رویکرد دیگر، مدل‌سازی مساله رتبه‌بندی بصورت یک مساله رده‌بندی است. در تحقیقات مختلف، امکان استفاده از مدل رده‌بندی برای مساله رتبه‌بندی در حوزه بازیابی اطلاعات، بررسی شده است که از جمله آنها می‌توان به [۳۳] و [۴۴] و [۲۷]. اشاره نمود. در این پژوهش‌ها، ایده تبدیل رتبه‌بندی به مساله رده‌بندی دودویی مطرح شده است. در تحقیقات دیگری نظیر [۲۵] و [۳]. بکارگیری رده‌بندی چند کلاسه برای حل مساله رتبه‌بندی، پیشنهاد شده است. بکارگیری مدل رگرسیون ترتیبی، در دسته سوم روش‌های نقطه‌ای، دنبال شده است که به عنوان نمونه می‌توان به [۸] و [۹] اشاره نمود. با توجه به اینکه اغلب شاخص‌های ارزیابی در حوزه بازیابی اطلاعات مبتنی بر موقعیت اسناد و در سطح پرس‌وجوها محاسبه می‌شوند، روش‌های نقطه‌ای در شرایط واقعی، دچار محدودیت‌های جدی هستند.

فضای فرضیه در روش‌های جفتی یادگیری رتبه‌بندی، شامل توابع دو متغیره‌ای است که یک جفت سند را به عنوان ورودی، دریافت می‌کنند و ترتیب نسبی آن دو را در قالب خروجی، ارائه می‌دهند. ضمناً تابع زیان مورد استفاده در روش‌های جفتی، صرفاً ترتیب نسبی بین زوج‌های اسناد را در نظر می‌گیرد. در نتیجه، موقعیت نهایی اسناد در فهرست مرتب نهایی، به سختی قابل استحصال خواهد بود. از جمله ایده‌های مطرح شده در روش‌های جفتی رتبه بندی یادگیری، می‌توان به استفاده از شبکه‌های عصبی مصنوعی [۳۸] و [۱۶] و [۵۲] و [۲۴]. استفاده از مدل ماشین بردار پشتیبان[11] [۲۱] و [۱۷]. و نیز بکارگیری الگوریتم‌های یادگیری ماشینی [۴] و [۵۱] و [۱۱] و [۲۰] و [۲۹]. اشاره نمود. ذکر این نکته در خصوص روش‌های جفتی، ضرورت دارد که با توجه به اینکه اغلب شاخص‌های ارزیابی در حوزه بازیابی اطلاعات، در سطح پرس‌وجوها و بر مبنای موقعیت نهایی اسناد در فهرست مرتب‌شده عمل می‌کنند، فاصله قابل توجهی، بین مدل عملکرد روش‌های جفتی و رتبه‌بندی مطلوب در مقوله بازیابی اطلاعات، مشاهده می‌شود.

فضای ورودی در روش‌های مبتنی بر فهرست، شامل مجموعه اسناد متناظر با پرس‌وجوی داده شده است. فضای

[7] Loss Function
[8] Pointwise
[9] Pairwise
[10] Listwise
[11] Support Vector Machine (SVM)

خروجی نیز عبارتست از فهرست مرتب‌شده اسناد یا به تعبیر دیگر، جایگشتی از مجموعه اسناد داده شده است. این دسته از الگوریتم‌های یادگیری رتبه‌بندی را بر اساس نوع تابع زیان مورد استفاده، به دو دسته، تقسیم نمود. دسته نخست، روش‌های بهینه‌سازی مستقیم[12] نامیده می‌شوند و شامل روش‌های رتبه‌بندی هستند که توابع زیان آنها، ارتباط مستقیمی با شاخص‌های استاندارد ارزیابی در حوزه بازیابی اطلاعات دارد [۴۰] و [۵۰] و [۴۶] و [۵۴] و [۵] و [۳۶] و [۷] و [۴۹]. در برخی دیگر از روش‌های اخیر، از رویکرد یادگیری عمیق[13] برای تخمین شاخص‌های ارزیابی استفاده شده است [۱] و [۳۷] و [۳۴] و [۳۹]. همچنین ایده استفاده از قابلیت‌های الگوریتم ژنتیک، به منظور بهینه‌سازی شاخص‌های ارزیابی، در برخی تحقیقات مورد توجه بوده است که از جمله آنها می‌توان به این موارد اشاره کرد: [۱۳] و [۴۱] و [۱۴] و [۴۸] و [۲۳]. در مقابل، در دسته دوم روش‌های فهرستی، الگوریتم‌های رتبه‌بندی قرار می‌گیرند که در آنها توابع زیان مورد استفاده، ارتباط مستقیمی با شاخص‌های استاندارد ارزیابی جستجو، ندارد. از جمله این روش‌ها می‌توان به [۴] و [۴۳] و [۳۴] و [۵۳] و [۴۴] اشاره نمود.

علیرغم مطالعات گسترده صورت گرفته در حوزه یادگیری رتبه‌بندی، این حوزه تحقیقاتی، با نواقص و محدودیت‌های جدی، همراه است. از مهم‌ترین این معضلات، عدم استفاده از اطلاعات مربوط به سوابق رفتارهای کاربران در حین انجام جستجوها، در فرآیندهای بازیابی و رتبه‌بندی نتایج است. در واقع، عدم اقبال کاربران برای ارایه بازخورد صریح از نحوه عملکرد جویشگر مورد استفاده، محققان را بر آن داشته است تا با پردازش سوابق رفتار کاربران، بازخورد ضمنی رفتار آنها را استخراج کنند. بنا بر نتایج تحقیقات انجام شده، اطلاعات حاصل از پردازش سوابق تعاملات و رفتارهای کاربران، که اصطلاحاً اطلاعات کلیک از گذر داده، نامیده می‌شود، بسیار ارزشمند بوده و بکارگیری آنها می‌تواند منجر به بهبود کیفیت فرآیند بازیابی اطلاعات، به میزان قابل توجهی شود [۲۲] و [۳۰] و [۳۱]. فقدان این قبیل اطلاعات، در اغلب مجموعه‌های داده محک[14] موجود در زمینه یادگیری رتبه‌بندی، باعث گردیده است تا اغلب روش‌های یادگیری رتبه‌بندی موجود، امکان استفاده از این قبیل اطلاعات را نداشته باشند که این امر، یکی از معضلات و محدودیت‌های اصلی پیش‌روی حوزه یادگیری رتبه‌بندی، بشمار می‌رود. در عین حال، عرضه انبوهی از ویژگی‌های مرتبط با پرس‌وجوها و نتایج و بکارگیری آنها در الگوریتم‌های موجود، کاربری آنها را در شرایط واقعی، دشوار می‌سازد. علت این امر آن است که در شرایط واقعی، تهیه داده‌های مربوط به این تعداد زیاد از ویژگی‌ها، فرآیندی دشوار و وقت‌گیر می‌باشد. بر این اساس و با الهام از پایه مفهوم «اطلاعات کلیک از گذر داده»، ایده جدیدی در سال‌های اخیر مطرح شده است که در آن، به تولید مجموعه‌ای از ویژگی‌های ثانویه بر پایه ویژگی‌های پایه مجموعه داده محک مورد استفاده پرداخته می‌شود. این مجموعه را اصطلاحاً «ویژگی‌های کلیک از گذر داده» می‌نامند. در [۲۳]، با اِعمال مدل برنامه‌نویسی ژنتیک چند لایه روی ویژگی‌های کلیک از گذر داده، الگوریتم جدیدی موسوم به MGP-Rank ارائه شده است که عملکرد قابل توجهی در رتبه‌بندی اطلاعات انگلیسی وب داشته است. توفیق این الگوریتم، الهام بخش این پژوهش بوده است تا با تطبیق ساختار این الگوریتم بر اساس ویژگی‌های زبان فارسی، الگوریتم رتبه‌بندی قدرتمندی برای دادگان فارسی وب، حاصل شود.

۳- معرفی روش پیشنهادی

این بخش به معرفی و بیان نحوه عملکرد الگوریتم MGP-Rank اختصاص یافته است و جزئیات گام‌های اجرای این الگوریتم، مورد بحث و بررسی قرار گرفته است.

[12] Direct optimization
[13] Deep Learning
[14] Benchmark Dataset

۳-۱-کلیات الگوریتم MGP-Rank

الگوریتم MGP-Rank با استفاده از مدل برنامه‌نویسی ژنتیک چند لایه، روی ویژگی‌های کلیک از گذر داده، عمل می‌کند[۲۳]. در این الگوریتم، ویژگی‌های کلیک از گذر داده، در یک چارچوب مدل برنامه‌نویسی ژنتیک چند لایه و چند جمعیتی، بکار گرفته می‌شوند تا بر اساس برخی شاخص‌های بازیابی اطلاعات، بهترین توابع رتبه‌بندی ممکن، شناسایی شوند. مهم‌ترین مزیت مدل برنامه‌نویسی ژنتیک چند لایه، قابلیت جستجوی سریع‌تر و فراگیرتر آن نسبت به مدل برنامه‌نویسی ژنتیک معمول است. بر این اساس، الگوریتم MGP-Rank، از این جهت با روش‌های مشابه، متفاوت است که در آن، از مفهوم اطلاعات کلیک از گذر داده [۲۱] به منظور تولید ویژگی‌های کلیک از گذر داده، استفاده شده است. فرآیند تولید ویژگی‌های کلیک از گذر داده، در هر دو حالت وجود یا عدم وجود اطلاعات مرتبط با کلیک از گذر داده در مجموعه داده محک مورد استفاده، امکان‌پذیر است. سپس از این ویژگی‌ها در قالب معماری LAGEP [۲۶]، به منظور یافتن توابع رتبه‌بندی مناسب، استفاده می‌شود که ترکیبی از ویژگی‌های مرتبط با محتوای اسناد، ویژگی‌های ساختاری صفحات وب و نیز ویژگی‌های مرتبط با نحوه جستجوی کاربران، خواهد بود. در تعیین توابع برتر، از شاخص‌های ارزیابی مرتبط با مقوله بازیابی اطلاعات نظیر دقت[۱۵]، صحت[۱۶]، فراخوانی[۱۷]، F-measure و نیز خاص بودن[۱۸] به عنوان توابع سازگاری، مورد استفاده قرار گرفته است [۳۲]. نتایج حاصل از آزمایش‌های صورت گرفته، نشان‌دهنده کارآمدی روش ارایه شده در مقایسه با الگوریتم‌های مطرح یادگیری رتبه‌بندی، بر اساس شاخص‌های دقت و NDCG است. این بهبود بازیابی، عمدتاً در بخش بالایی فهرست نتایج جستجوها قابل توجه است که غالباً بیشتر توجه کاربران را به خود، معطوف می‌دارد. در عین حال، بهبود حاصل از الگوریتم ارایه شده، در شرایطی که داده محک اولیه، در برگیرنده اطلاعات کلیک از گذر داده کاربران باشد، بیشتر خواهد بود.

۳-۲- مراحل الگوریتم MGP-Rank

الگوریتم MGP-Rank، متشکل از دو مرحله است که در گام نخست آن، بر اساس مفهوم کلیک از گذر داده، از اطلاعات موجود در مجموعه داده محک اولیه، هشت ویژگی ثانویه تولید می‌شود که آنها را ویژگی‌های کلیک از گذر داده می‌نامیم. برای این منظور، از تعدادی سناریوی متفاوت جهت تولید این ویژگی‌ها، استفاده می‌شود. نکته جالب توجه در خصوص این فرآیند، آن است که این کار، حتی در غیاب اطلاعات کلیک از گذر داده در مجموعه داده محک اولیه نیز امکان‌پذیر می‌باشد. در ادامه و طی گام دوم، از این ویژگی‌های ثانویه به منظور تولید توابع رتبه‌بندی، استفاده می‌شود که در ساختار مدل برنامه‌نویسی ژنتیک چند لایه موسوم به LAGEP [۲۵]، بر اساس توابع سازگاری مرتبط با حوزه بازیابی اطلاعات، تکامل می‌یابند تا اینکه نهایتاً توابع رتبه‌بندی مطلوب، شناسایی شوند.

۳-۲-۱- استخراج ویژگی‌های کلیک از گذر داده

هر چند تحقیقات مختلف نشان داده‌اند که بکارگیری اطلاعات کلیک از گذر داده در بهبود نتایج بازیابی، مفید می‌باشد
[۲۲] و [۳۰] و [۳۱]، ولی این قبیل اطلاعات در بسیاری از مجموعه‌های داده ارائه شده برای مساله یادگیری رتبه‌بندی نظیر مجموعه داده LETOR مربوط به مایکروسافت [۳۵] و نیز مجموعه داده تهیه شده توسط شرکت یاهو [۶]، مورد استفاده قرار نگرفته است. از سوی دیگر، این مجموعه‌های داده، شامل تعداد زیادی از ویژگی‌ها هستند و الگوریتم‌های مستخرج از این آنها، بدلیل بکارگیری این حجم بالا از ویژگی‌ها، در شرایط کاربری واقعی، چندان کارآمد نخواهند بود. با توجه به این واقعیات، الگوریتم MGP-Rank در گام نخست و بر اساس مدل مفهومی اطلاعات کلیک از گذر

[15] Precision
[16] Accuracy
[17] Recall
[18] Specificity

داده، اقدام به تهیه هشت ویژگی ثانویه از روی اطلاعات موجود در مجموعه داده محک اولیه داده شده، موسوم به ویژگی‌های کلیک از گذر داده، می‌کند. جزییات مفهومی این ویژگی‌ها در ادامه این بخش، بیان شده است.

در فرآیند استخراج ویژگی‌های کلیک از گذر داده، یک بازنمایی فشرده از مجموعه داده محک اولیه بر اساس مدل مفهومی کلیک از گذر داده ارایه می‌شود [۲۳]. در واقع، مفهوم کلیک از گذر داده، بصورت یک سه‌تایی $\langle Q, R, C \rangle$ شامل پرس‌وجوی کاربر، فهرست نتایج و مجموعه کلیک‌های کاربر، تعریف شده است [۲۱]. با الهام از این مدل، در گام تبدیل، تعداد هشت ویژگی جدید موسوم به ویژگی‌های کلیک از گذر داده، در سه گروه Q، R و C، بشرح زیر تعریف می‌شود:

جدول ۱: ویژگی‌های کلیک از گذر داده معرفی شده مبتنی بر مفهوم اطلاعات کلیک از گذر داده

در این دسته‌بندی، مجموعه Q، شامل ویژگی‌هایی است که تا حدی، مرتبط با ماهیت پرس‌وجوی کاربر هستند. در این دسته، ویژگی $Repetition$، به میزان تکرار واژگان پرس‌وجوی کاربر در بخش‌های مختلف یک سندِ پاسخ، شامل: آدرس (URL)، عنوان و متن مرتبط می‌باشد. ویژگی $QScore$، به میزان مرتبط بودن سند با توجه به پرس‌وجوی کاربر، اشاره می‌کند که در محاسبه آن، ترکیبی از نتایج روش‌های رتبه‌بندی وابسته به پرس‌وجوی کاربر از جمله روش‌های مبتنی بر مدل فضای برداری و روش‌های مدل زبانی، مورد استفاده قرار می‌گیرد. نهایتاً ویژگی $ResultAmount$، معطوف به تعداد پاسخ‌های ارایه شده در قبال پرس‌وجوی مورد نظر می‌باشد.

به همین ترتیب، دسته R، شامل ویژگی‌هایی است که تاکید بیشتری بر مشخصات اسناد، بصورت مستقل از پرس‌وجوی مطرح شده از سوی کاربر دارند. در این دسته، ویژگی $AbsoluteRank$، به رتبه مطلق یک سند صرفنظر از پرس‌وجوی کاربر، می‌پردازد که طبعاً اطلاعاتی نظیر $PageRank$ در آن، نقش مهمی ایفا می‌کنند. ویژگی $StreamLength$ به طول یک سند اشاره دارد و ترکیبی از اطلاعات طول بخش‌های مختلف یک سند از جمله URL، عنوان و متن آن را شامل می‌شود.

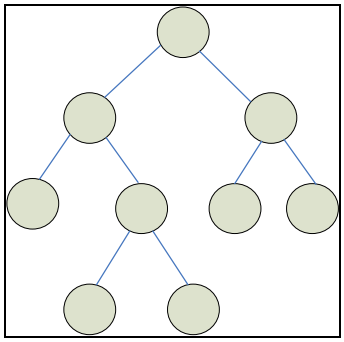
تمرکز عمده ویژگی‌های دسته C، بر اطلاعات قابل حصول از یک زوج سند-پرس‌وجو، در رابطه با کلیک‌های کاربران است. در این دسته، ویژگی $Specificity$، به میزان خاص بودن اطلاعات یک سند، طی تراکنش‌های کاربران در پرس‌وجوهای مختلف، اشاره می‌کند و قصد دارد تا به نوعی، به این نکته بپردازد که یک سند پاسخ، چقدر شامل موضوعات خاص است، بنحوی‌که در تعداد معدودی از پرس‌وجوها، به عنوان یک پاسخ بالقوه، مورد توجه کاربران واقع شده است. ویژگی $Attractiveness$ نیز سعی در نشان دادن میزان جلب توجه کاربران و سامانه‌های بازیابی طی پرس‌وجوهای مختلف به یک سند خاص، دارد. در محاسبه این ویژگی، طبعاً اینکه یک سند، بعنوان اولین یا آخرین پاسخ بالقوه از سوی کاربر، کلیک شده است، حایز اهمیت می‌باشد. نهایتاً ویژگی $ClickRate$، به میزان پتانسیل سند مورد نظر برای دریافت کلیک کاربران طی پرس‌وجوهای مختلف، می‌پردازد.

نکته قابل توجه این است که ویژگی‌های هشت‌گانه مندرج در جدول شماره یک را می‌توان در هر دو صورتِ وجود یا عدم وجودِ اطلاعات صریح کلیک از گذر داده در مجموعه داده محک اولیه، محاسبه نمود. طبعاً روند محاسبه این ویژگی‌ها، وابسته به مفهوم و ساختار ویژگی‌های اصلی موجود در مجموعه‌های داده محک مختلف، متفاوت خواهد بود. بر این اساس، در بخش ۳-۴، سناریوهایی برای محاسبه این ویژگی‌ها روی مجموعه داده فارسی dotIR پیشنهاد می‌شود.

۳-۲-۲- بکارگیری چارچوب برنامه‌نویسی ژنتیک چند لایه

ویژگی‌های کلیک از گذر داده حاصل از اجرای گام نخست الگوریتم MGP-Rank، به عنوان ورودی، جهت بکارگیری در چارچوب مدل برنامه‌نویسی ژنتیک چند لایه و چند جمعیتی موسوم به LAGEP، بکار گرفته می‌شوند [۲۶]. در این ساختار، هر جمعیت، شامل تعدادی افراد است که در اینجا به عنوان توابع رتبه‌بندی بالقوه، مطرح می‌باشند. یک

فرد این جمعیت مثلاً I، بصورت سه‌گانه $< S_v, S_c, S_{op} >$ نمایش داده می‌شود که در آن، S_v، مجموعه متغیرها، S_c، مجموعه مقادیر ثابت و S_{op}، مجموعه عملگرهای ریاضی است. متغیرها، همان ویژگی‌های موجود در داده‌های آموزشی هستند. طی بخش بعدی، جزییات ساختار استفاده شده در آزمایش‌های انجام شده، بیان خواهد شد. بر این اساس، یک جمعیت مانند P که شامل N فرد باشد را می‌توان بصورت زیر تعریف نمود: $P = \{I_1, I_2, \dots, I_N\}$. هر فرد عضو این جمعیت‌ها را می‌توان بصورت یک درخت دودویی نمایش داد که برگ‌های آن، یا متغیرها (ویژگی‌های زوج‌های سند-پرس‌وجو) و یا مقادیر ثابت هستند و گره‌های میانی نیز عملگرهای ریاضی می‌باشند. به عنوان یک نمونه، در شکل شماره سه، بازنمایی فردی از جمعیت را که یک تابع رتبه‌بندی بصورت $\big((F_1 + (F_1 + F_2))/(F_3 - 0.5)\big)$ است، در قالب درخت دودویی، نمایش داده شده است.



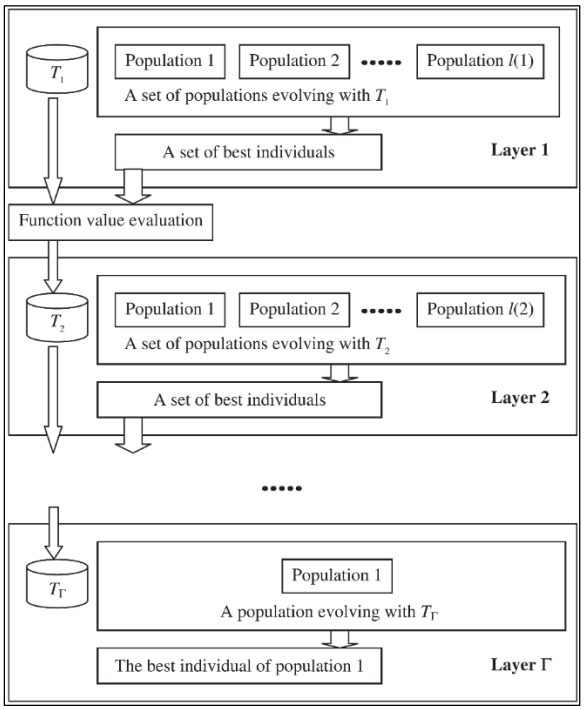
شکل ۳: بازنمایی تابع $\big((F_1 + (F_1 + F_2))/(F_3 - 0.5)\big)$ بصورت درخت دودویی

هر جمعیت با استفاده از سه عملگر ژنتیکی اصلی، یعنی ادغام[19]، جهش[20] و بازتولید[21]، تکامل می‌یابند. نرخ بکارگیری عملگرهای سه‌گانه ادغام، جهش و بازتولید، به ترتیب برابر با R_c، R_m و R_r می‌باشد. در بکارگیری عملگر جهش، نخست بصورت تصادفی، گره‌ای از ساختار دودویی فرد مورد نظر به عنوان نقطه جهش[22] در نظر گرفته می‌شود و با گره دیگری که از مجموعه متناظر یعنی S_v، S_c ویا S_{op}، که بصورت تصادفی انتخاب شده است، جایگزین می‌شود. در واقع، هدف عملگر جهش، اجتناب از بهینه‌های محلی است، زیرا با بکارگیری این عملگر، افراد جدیدی حاصل می‌شوند که ممکن است تا بحال در جمعیت‌های مورد مطالعه، مشاهده نشده باشند. به همین ترتیب، در عملگر ادغام نیز زیردرخت‌هایی از دو فرد که بصورت تصادفی انتخاب شده‌اند، با یکدیگر جایگزین می‌شوند و در نتیجه آن، دو فرد جدید، حاصل می‌شوند. نهایتاً در عملگر بازتولید، قاعده طبیعی بقای قوی‌ترین و مناسب‌ترین افراد، مدل‌سازی می‌شود و بهترین فرد هر نسل از جمعیت، عیناً در نسل بعدی نیز فرصت حضور، خواهد یافت. با توجه به اینکه عملگر جهش، تنوع افراد یک جمعیت را کنترل می‌کند، بدیهی است که نرخ بکارگیری آن (R_m) طی فرآیند تکامل یک جمعیت، بصورت مناسبی تنظیم شود. در LAGEP، این فرآیند از طریق روش تنظیم تطبیقی نرخ جهش (AMRT[23]) صورت می‌پذیرد که جزییات آن در [۲۶] ذکر شده است. بطور خلاصه، در هر نسل از جمعیت، روش AMRT ، اقدام به ارزیابی میزان سازگاری افراد آن جمعیت می‌کند. چنانچه مقادیر میزان سازگاری افراد، یکسان باشد، مقدار R_m افزایش می‌یابد و در غیر این‌صورت، نرخ جهش، برابر با مقدار اولیه آن، خواهد بود. علاوه بر آن، با توجه به اینکه: $R_m + R_c = 1$ می‌باشد، افزایش R_m، به معنی کاهش R_c است. در تحقیق انجام شده، مشخص شده است که استفاده از این تکنیک، منجر به بهبود عملکرد LAGEP خواهد شد [۲۶]. ضمناً بایستی خاطر نشان کرد

[19] Crossover
[20] Mutation
[21] Reproduction
[22] Mutation Point
[23] Adaptive Mutation Rate Tuning

که برای انتخاب افراد جهت اِعمال عملگرهای ادغام و جهش، می‌توان از روش‌های انتخاب مختلف نظیر انتخاب تصادفی[۲۴] یا انتخاب از طریق مسابقه استفاده نمود[۲۵]. در روش انتخاب از طریق مسابقه، ابتدا افرادی از جمعیت هدف به تعداد مشخصی که اصطلاحاً اندازه مسابقه[۲۶] نامیده می‌شود، بصورت تصادفی، انتخاب می‌شوند و سپس بر اساس میزان سازگاری آنها، یک یا دو نفر آنها از به ترتیب، جهت عملگرهای جهش یا ادغام، برگزیده می‌شوند.

در معماری معرفی شده برای LAGEP، جمعیت‌ها در یک ساختار لایه‌ای قرار می‌گیرند که هر لایه، شامل تعدادی جمعیت و مجموعه متغیرها و پارامترهای مختص به خود است. در واقع، هر لایه از مدل LAGEP، خود، یک مدل چند جمعیتی از مدل برنامه‌نویسی ژنتیک است که با استفاده از مجموعه داده آموزشی یکسانی، تکامل می‌یابند و بهترین افراد آنها در یک مجموعه قرار داده می‌شوند و بنا به مدل تشریح شده در [۲۶]، با استفاده از آنها داده‌های آموزشی برای لایه بعدی، تولید می‌شود. در هر لایه و طی تعدادی تکرار، جمعیت‌ها تکامل پیدا می‌کنند و در رقابت افراد آنها با یکدیگر، تنها تعداد معدودی از بهترین آنها با توجه به میزان سازگاری، برای بقا در نسل بعد، باقی می‌مانند. تولید مجموعه متغیرهای لایه‌های بعدی نیز طی روش مشابهی، صورت می‌گیرد [۲۶]. پس از اتمام فرآیند تکامل در همه لایه‌ها، خروجی آخرین لایه، به عنوان خروجی نهایی فرآیند LAGEP، در نظر گرفته می‌شود. شکل شماره چهار، شمای کلی تکنیک برنامه‌نویسی چند لایه و چند جمعیتی را نشان می‌دهد.



شکل ۴: شمای کلی تکنیک برنامه‌نویسی چند لایه و چند جمعیتی (Lin et al. 2008)

۴- آزمایش‌های صورت گرفته

در این بخش، ابتدا به معرفی داده محک dotIR می‌پردازیم و سپس، شاخص‌های مورد استفاده در ارزیابی رتبه‌بندی بیان خواهند شد. ادامه این بخش، به ذکر سناریوهای استخراج ویژگی‌های کلیک از گذر داده روی مجموعه داده محک dotIR، اختصاص یافته است. پس از آن، تنظیمات و پارامترهای آزمایش‌های بعمل آمده، مطرح می‌شود و نهایتاً به طرح نتایج حاصل از آزمایش‌های صورت گرفته و نیز تحلیل نتایج بدست آمده، خواهیم پرداخت.

۴-۱- معرفی داده محک مورد استفاده

در بررسی عملکرد الگوریتم MGP-Rank در محیط وب فارسی، از مجموعه داده dotIR استفاده شده است که مشخصات و ساختار آن در ادامه، بیان شده است. مجموعه داده محک dotIR، تنها داده محک در دسترس برای بازیابی اطلاعات فارسی در محیط وب می‌باشد. این مجموعه داده، توسط دانشگاه تهران و با حمایت مرکز تحقیقات مخابرات ایران، در سال ۱۳۸۸ ایجاد شده است. تولید این مجموعه داده، بر اساس خزش صورت گرفته در دامنه ir. طی سال ۱۳۸۸ و داده‌های جمع‌آوری شده از حدود ۸،۵ میلیون سند

[24] Random Selection
[25] Tournament Selection
[26] Tournament Size

در این دامنه، انجام شده است [۱۲]. این مجموعه داده، شامل ۵۰ پرس‌وجوی مختلف و هزار سند به ازای هر یک از این پرس‌وجوها است که جمعاً پنجاه هزار زوج پرس‌وجو-سند را تشکیل می‌دهند. به ازای هر یک از زوج‌های داده، ۵۶ ویژگی مختلف استخراج شده است که فهرست آنها در جدول پیوست شماره یک، آمده است. این ویژگی‌ها در سه دسته کلی یعنی ویژگی‌های مبتنی بر محتوا، ویژگی‌های مبتنی بر پیوند و نیز ویژگی‌های ترکیبی، ارائه شده‌اند. ضمناً قضاوت انسانی در مورد میزان مرتبط بودن هر زوج پرس‌وجو-سند، در دو سطح مرتبط و نامرتبط، انجام شده است.

به منظور سهولت ارزیابی الگوریتم‌ها، این مجموعه داده به پنج چین مختلف، تقسیم‌بندی شده است. هر دسته، بشرح مندرج در جدول شماره دو، شامل یک بخش آموزش، یک بخش اعتبارسنجی و همچنین یک بخش آزمون می‌باشد.

جدول ۲: دسته‌بندی مجموعه داده dotIR به بخش‌های آموزش، اعتبارسنجی و آزمون

(Darrudi et al. 2009)

چین	داده آموزش	داده اعتبارسنجی	داده آزمون
$Fold_1$	$\{S_1,S_2,S_3\}$	S_4	S_5
$Fold_2$	$\{S_2,S_3,S_4\}$	S_5	S_1
$Fold_3$	$\{S_3,S_4,S_5\}$	S_1	S_2
$Fold_4$	$\{S_4,S_5,S_1\}$	S_2	S_3
$Fold_5$	$\{S_5,S_1,S_2\}$	S_3	S_4

ساختار کلی اطلاعات موجود در این داده محک، در شکل شماره پنج، نشان داده شده است.

Label qid:queryID 1:F_1Value 2:F_2Value 3:F_3Value … 55:F_{55}Value 56:F_{56}Value #docid = docID

شکل ۵: ساختار داده محک dotIR

(Darrudi et al. 2009)

۴-۲- شاخص‌های ارزیابی

به منظور ارزیابی عملکرد الگوریتم MGP-Rank، از شاخص‌های مرجع در ارزیابی رتبه‌بندی یعنی $P@n$، MAP، $NDCG@n$ و $MeanNDCG$، استفاده شده است.

- دقت[27] در نقطه k ($P@k$): اگر قضاوت انسانیِ اسناد، به‌صورت دودویی در اختیار باشد، مثلاً برچسب اسناد مرتبط، یک و برچسب اسناد نامرتبط، صفر باشد، در این صورت، $P@k$ به‌صورت زیر تعریف می‌شود [۳۲]:

$$P = \quad (۱)$$

که $I_{\{.\}}$ تابع شاخص است و $l_{\pi^{-1}(j)}$ به برچسب سند قرار گرفته در موقعیت j لیست مرتب نتایج، π، اشاره می‌کند.

- متوسط میانگین دقت[28] (MAP): بر اساس تعریف ارائه شده از دقت در موقعیت k، میانگین دقت[29] (AP) عبارتست از:

$$A = \quad (۲)$$

که در آن، m تعداد کل اسناد متناظر پرس‌وجوی q است و m_1 تعداد زیر مجموعه‌ای از این اسناد است که میزان مرتبط بودن آنها برابر با یک باشد. مقدار میانگین AP روی همه پرس‌وجوهای آزمون، متوسط میانگین دقت (MAP) نامیده می‌شود [۳۲].

- بهره انباشته کاهشی هنجار شده[30] (NDCG) در نقطه k ($NDCG@k$): این شاخص بر پایه میانگین‌گیری از شاخص بهره انباشته کاهشی[31] (DCG)، محاسبه می‌شود. شاخص بهره انباشته کاهشی (DCG) می‌تواند قضاوت‌های میزان مرتبط بودن را که به‌صورت طبقه‌بندی‌های مرتب چندگانه ارائه شده باشند، در ارزیابی مورد استفاده قرار دهد؛ ضمن آنکه در تعریف آن، یک ضریب کاهشی بر اساس موقعیت اسناد، در نظر گرفته شده است [۱۸] و ۹[۳۲]. برای بیان این موضوع به زبان ریاضی، فرض کنید که لیست

[27] Precision
[28] Mean Average Precision
[29] Average Precision
[30] Normalized Discounted Cumulative Gain
[31] Discounted Cumulative Gain

مرتبی از نتایج نظیر π به ازای پرس‌وجوی q وجود داشته باشد؛ در این صورت، مقدار شاخص DCG در موقعیت k به‌صورت زیر تعریف می‌شود:

$$DCG@k(\pi, l) = \sum_{j=1}^{k} G\left(l_{\pi^{-1}(j)}\right)\eta(j) \quad (۳)$$

که در آن، $G(.)$ رتبه یک سند است که عموماً به‌صورت $G(z) = (2^z - 1)$ می‌باشد و $\eta(j)$ یک ضریب کاهشی بر اساس موقعیت است که غالباً به‌صورت $\eta(j) = 1/log(j+1)$ تعریف می‌شود. با هنجار سازی مقادیر $DCG@k$ با یک مقدار بیشینه (مثلاً Z_k) شاخص بهره انباشته کاهشی هنجار شده ($NDCG$) حاصل می‌گردد که عبارتست از:

$$NDCG@k(\pi, l) = \frac{1}{Z_k}\sum_{j=1}^{k} G\left(l_{\pi^{-1}(j)}\right)\eta(j) \quad (۴)$$

بدیهی است که دامنه مقادیر ممکن برای شاخص $NDCG$، بین صفر و یک قرار می‌گیرد.

- میانگین بهره انباشته کاهشی هنجار شده ($MeanNDCG$): با میانگین‌گیری مقادیر $NDCG@k$ در موقعیت‌های مختلف $k = 1 \dots 10$، شاخص میانگین بهره انباشته کاهشی هنجار شده، محاسبه می‌شود [۳۲].

۴-۳- سناریوهای استخراج ویژگی‌های کلیک از گذر داده

برای تولید ویژگی‌های کلیک از گذر داده در الگوریتم MGP-Rank، بر اساس سناریوهای مندرج در جدول شماره سه، عمل می‌شود و نتایج برترین سناریوها در ادامه، گزارش خواهد شد.

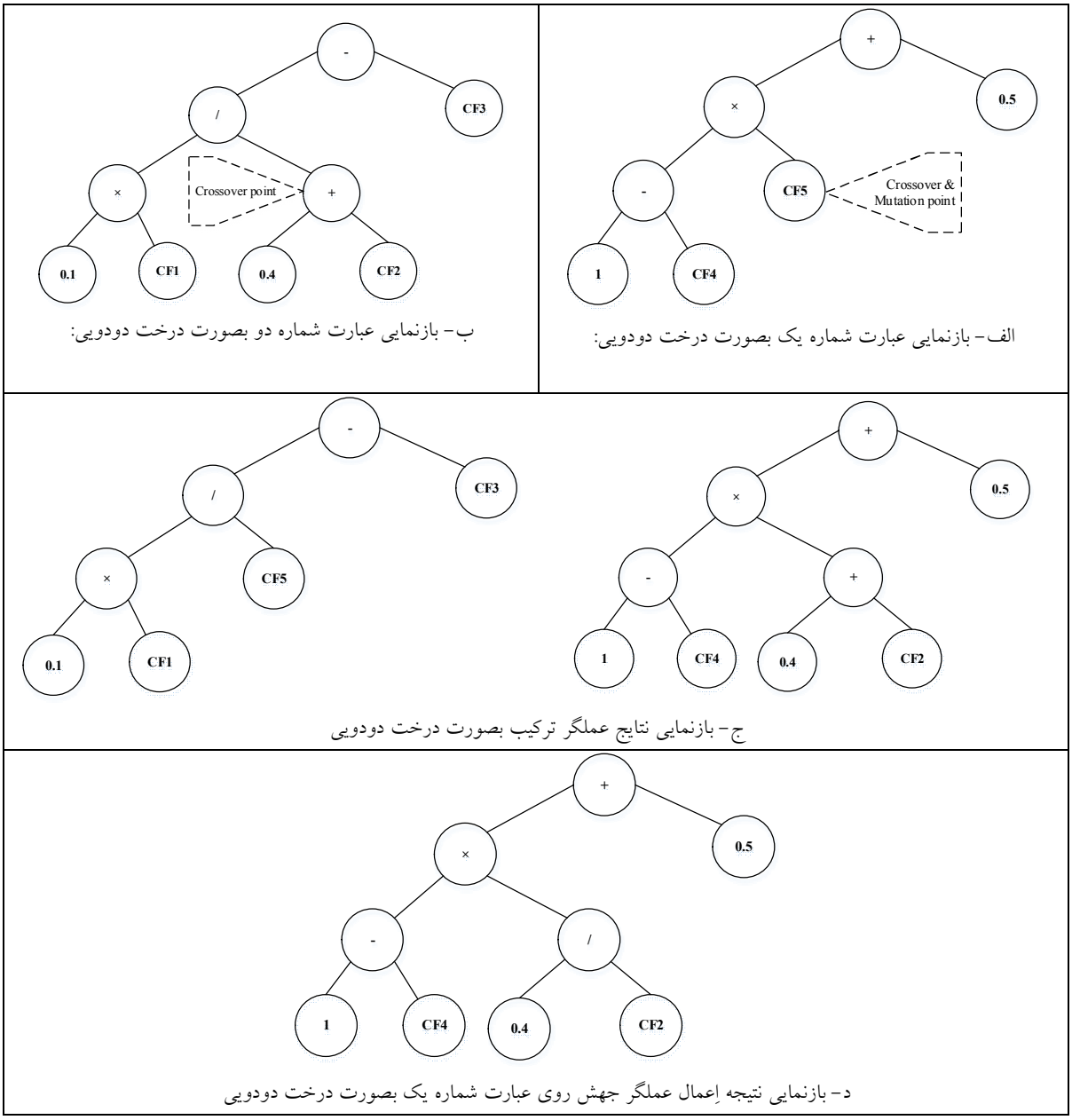
جدول ۳: سناریوهای تولید ویژگی‌های کلیک از گذر داده روی داده محک dotIR

شناسه سناریو	روش محاسبه
IR-DF1	
IR-DF2	

IR-DF3	Q F C
IR-DF4	Q R C
IR-DF5	Q F C
IR-DF6	Q F C
IR-DF7	Q R C

شایان توجه است که در همه این سناریوها، ابتدا خصوصیات پایه داده محک dotIR، به روش Min-Max، هنجارسازی شده‌اند و سپس روش هموار سازی موسوم به Dirichlet prior [۳۲]، روی آنها اِعمال شده است.

ویژگی‌های کلیک از گذر داده بدست آمده، در فرآیند تولید جمعیت، به عنوان مجموعه متغیرها، S_v، مورد استفاده قرار می‌گیرند. هر فرد مانند I بصورت یک درخت دودویی نمایش داده می‌شود. در شکل شماره شش، دو فرد نمونه به همراه نتایج حاصل از اِعمال عملگرهای ترکیب و جهش، به تصویر کشیده شده است. در این مثال، متغیرهای CF1، CF2، CF3، CF4 و CF5، به ترتیب نماینده ویژگی‌های کلیک از گذر داده $QScore$، $Repetition$، $ClickRate$، $AbsoluteRank$ و $Attractiveness$ می‌باشند.



شکل ۶: نمونه افراد جمعیت ژنتیکی و نتایج اِعمال عملگرهای ژنتیک روی آنها

۴-۴- تنظیمات و پارامترهای آزمایش‌ها

در این قسمت، نتایج حاصل از بررسی‌های انجام شده مطابق سناریوهای مندرج در بخش ۳-۴، روی مجموعه داده dotIR ارایه می‌شود. تمامی آزمایش‌ها روی یک رایانه با پردازنده دو هسته‌ای 2.0GHz، شامل 2MB کاشه و 3GB حافظه، صورت پذیرفته است. در ارزیابی عملکرد الگوریتم MGP-Rank از چارچوب LAGEP با تنظیمات مندرج در **Error! Reference source not found.** شماره چهار، استفاده شده است.

جدول ۴: مشخصات چارچوب LAGEP در مطالعه الگوریتم MGP-Rank

پارامتر	مقدار
تعداد لایه‌ها	۳
نرخ بکارگیری عملگرهای ژنتیک	ادغام: ۰,۹۵، جهش: ۰,۰۵
احتمال بکارگیری عملگرهای ریاضی	احتمال مساوی برای $\{ +, -, \times, /, Sin, Cos, Exp, Ln, ... \}$
احتمال بکارگیری ثوابت عددی	احتمال مساوی برای $\{ 0, 0.1, 0.2, ..., 0 \}$ و $\{ e, \pi, \pi/2 \}$
احتمال بکارگیری انواع گره‌ها	متغیرها: ۰,۰۷۵، مقادیر ثابت: ۰,۰۲۵، عملگر ریاضی: ۰,۹
درصد رشد[32]	۵۰

[32] Grow Percentage

چارچوب LAGEP مورد استفاده در این تحقیق، از سه لایه مجزا تشکیل یافته است که جزییات آنها در جدول شماره پنج آمده است. هر لایه، شامل تعدادی جمعیت است که در مقداردهی اولیه آنها از روش رشد[۳۳]، استفاده شده است. در این روش، درخت‌های متناظر با افراد عضو جمعیت‌ها، طی یک فرآیند بازگشتی، ایجاد می‌شوند، به نحوی که عمق درخت حاصله، از مقدار متغیر «عمق درخت»، بیشتر نباشد و حجم درخت نسبت به درخت کامل با همان عمق، از مقدار متغیر «درصد رشد»، تجاوز نکند. جمعیت‌های حاصل از این فرآیند، توسط عملگرهای ژنتیکی با نرخ بکارگیری مشخص شده در جدول شماره چهار، تکامل می‌یابند. سپس از هر جمعیت، فردی با بیشترین میزان تابع برازش، برای بقا در نسل بعد، باقی خواهد ماند. این فرآیند در هر جمعیت تا رسیدن به تعداد مشخصی از نسل‌ها، ادامه پیدا می‌کند.

[33] Grow Method

جدول۵: مشخصات لایه‌های چارچوب LAGEP مورد استفاده

مشخصات لایه	لایه شماره یک	لایه شماره دو	لایه شماره سه
تعداد جمعیت‌ها	۱	۲	۳
تعداد افراد هر جمعیت	۱۰۰	۱۰۰	۱۰
تعداد نسل‌ها	۵۰	۱۰	۱۰
اندازه مسابقه	۵	۷	۱
تعداد افراد باقی مانده از یک جمعیت در نسل بعدی	۱	۱	۱
عمق درخت	۴	۶	۴

هر فرد I از یک جمعیت را می‌توان بصورت یک رده‌بندی کننده، در نظر گرفت و کارآیی آن در فرآیند رده‌بندی را توسط یک ماتریس ابهام[34] با چهار مقدار: True Positive (TP)، True Negative (TN)، False Positive (FP) و False Negative (FN)، مندرج در جدول شماره شش، بیان نمود.

جدول ۶: بیان کارآیی فرد I در رده‌بندی بر اساس مقادیر TP، FP، FN و TN

	نمونه مثبت	نمونه منفی
پاسخ I مثبت است	TP	FP
پاسخ I منفی است	FN	TN

در آزمایش‌های انجام شده، از پنج تابع برازش مختلف و مرتبط با حوزه بازیابی اطلاعات، به شرح مندرج در جدول شماره هفت، به منظور ارزیابی کیفیت افراد جمعیت‌ها، استفاده شده است.

جدول۷: توابع سازگاری استفاده شده جهت تعیین سازگاری افراد جمعیت ا

شناسه تابع برازش	تابع برازش	روش محاسبه
FF1	Accuracy	(TP + TN) / (TP + TN + FP + FN)
FF2	Precision	TP / (TP + FP)
FF3	Recall	TP / (TP + FN)
FF4	F-measure	2 TP / (TP + FP + FN)
FF5	Specificity	TN / (TN + FP)

همه نتایج گزارش شده در قسمت‌های آتی، بر اساس بازه اطمینان ۹۵٪ حاصل شده‌اند. این نتایج، بر اساس میانگین‌گیری از ۱۵۰ بار تکرار هر نسخه از الگوریتم MGP-Rank روی مجموعه داده dotIR حاصل شده است. دلیل این کار، ماهیت تصادفی مدل برنامه‌نویسی تکاملی است که در نتیجه آن، ممکن است خروجی هر بار اجرای الگوریتم، با دفعات دیگر، تفاوت داشته باشد. تنظیمات مورد استفاده در بهترین نسخه‌های مورد مطالعه عملکرد الگوریتم MGP-Rank در زبان فارسی، در جدول شماره هشت ارایه شده است.

جدول ۸: مشخصات نسخه‌های مختلف الگوریتم MGP-Rank جهت ارزیابی روی مجموعه داده dotIR

شناسه ساختار	شناسه سناریو	شناسه تابع برازش
MGP.IR1	IR-DF7	FF4
MGP.IR2	IR-DF4	FF4
MGP.IR3	IR-DF7	FF1
MGP.IR4	IR-DF6	FF1

۴- نتایج بدست آمده

[34] Confusion Matrix

در جدول شماره نه، نتایج حاصل از بررسی کیفیت بازیابی الگوریتم MGP-Rank بر اساس شاخص دقت روی مجموعه داده dotIR، گزارش شده است. لازم به ذکر است که مقایسه عملکرد الگوریتم MGP-Rank با روش‌های مرجع رتبه‌بندی بر اساس شاخص دقت، مطابق نتایج درج شده در جدول شماره ده، انجام می‌شود. به منظور تعیین مبنایی جهت مقایسه عملکرد الگوریتم MGP-Rank نسبت به دیگر روش‌های مرجع رتبه‌بندی، در این پژوهش، عملکرد سه الگوریتم پایه رتبه‌بندی، روی مجموعه دادگان dotIR، مورد بررسی قرار گرفته است و نتایج بدست آمده بر اساس شاخص دقت، در جدول شماره ده، گزارش شده است.

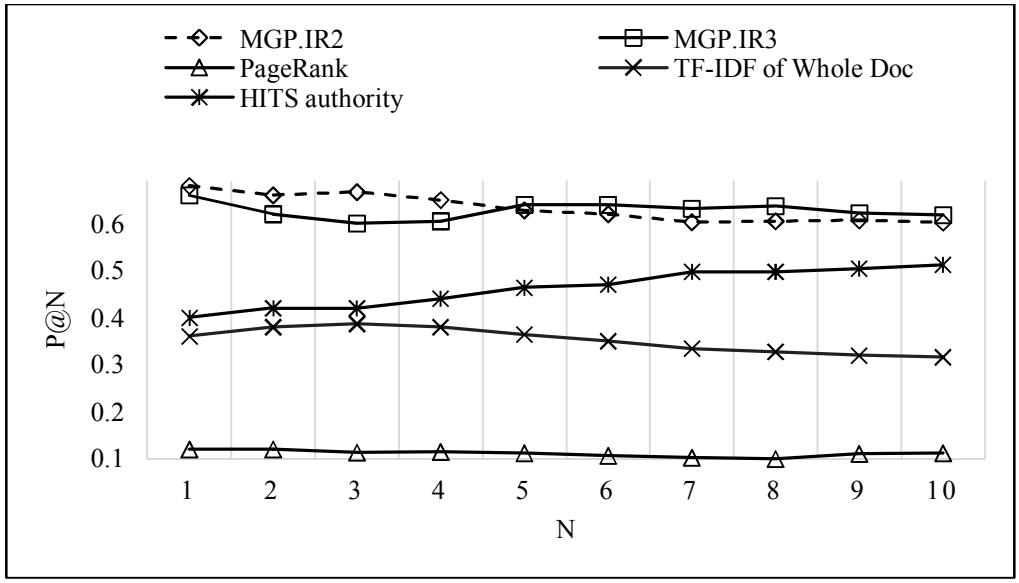
جدول ۹: نتایج حاصل از ارزیابی الگوریتم MGP-Rank بر اساس شاخص دقت روی مجموعه داده dotIR

شناسه ساختار	P@1	P@2	P@3	P@4	P@5	P@6	P@7	P@8	P@9	P@10	MAP
MGP.IR1	0.68	0.61	0.5867	0.57	0.588	0.5867	0.58	0.5775	0.5667	0.56	0.441
MGP.IR2	0.68	0.66	0.6667	0.65	0.628	0.62	0.6029	0.605	0.6067	0.602	0.4408
MGP.IR3	0.66	0.62	0.6	0.605	0.64	0.64	0.6314	0.6375	0.6222	0.618	0.4545
MGP.IR4	0.64	0.61	0.6	0.615	0.64	0.6333	0.6343	0.635	0.62	0.616	0.454

جدول۱۰: نتایج ارزیابی روش‌های استاندارد رتبه‌بندی طبق شاخص دقت روی مجموعه داده dotIR

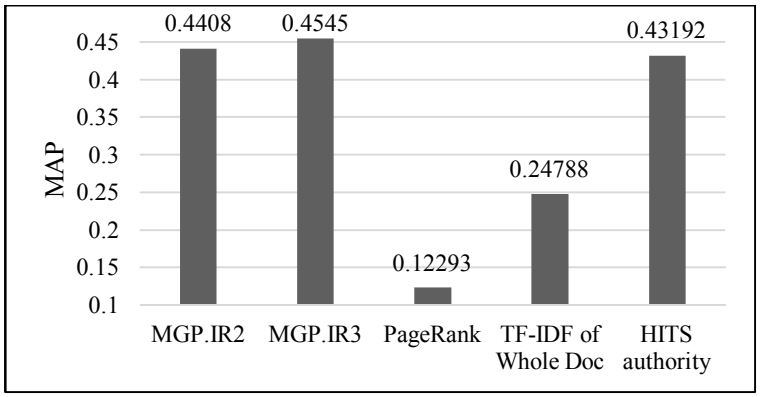
الگوریتم	P@1	P@2	P@3			P@4	P@5	P@6	P@7	P@8	P@9	P@10	MAP
PageRank	0.12	0.12	0.1133			0.115	0.112	0.1067	0.1029	0.1	0.1111	0.112	0.1229
TF-IDF of Whole Doc	0.36	0.38	0.3867			0.38	0.364	0.35	0.3343	0.3275	0.32	0.316	0.2479
HITS authority	0.4	0.42	0.42			0.44	0.464	0.47	0.4971	0.4975	0.5044	0.512	0.4319

مقایسه تصویری اطلاعات جدول شماره نه با نتایج عملکرد روش‌های مرجع مندرج در جدول شماره ده، در شکل شماره هفت، نمایش داده شده است.



شکل۷: مقایسه عملکرد الگوریتم MGP-Rank در زبان فارسی با روش‌های مرجع رتبه‌بندی طبق شاخص دقت

در شکل شماره هشت نیز مقایسه الگوریتم MGP-Rank با روش‌های مرجع رتبه‌بندی در زبان فارسی طبق شاخص MAP، نمایش داده شده است.



شکل ۸: مقایسه الگوریتم MGP-Rank با روش‌های مرجع رتبه‌بندی در زبان فارسی طبق شاخص MAP

به همین ترتیب، بررسی عملکرد الگوریتم MGP-Rank روی داده dotIR بر اساس شاخص NDCG نیز انجام شده است که نتایج بدست آمده در جدول شماره یازده، ارایه شده است. ضمن آنکه نتایج بدست آمده به ازای روش‌های رتبه‌بندی مرجع روی دادگان dotIR بر اساس شاخص NDCG، در جدول شماره دوازده، گزارش شده است.

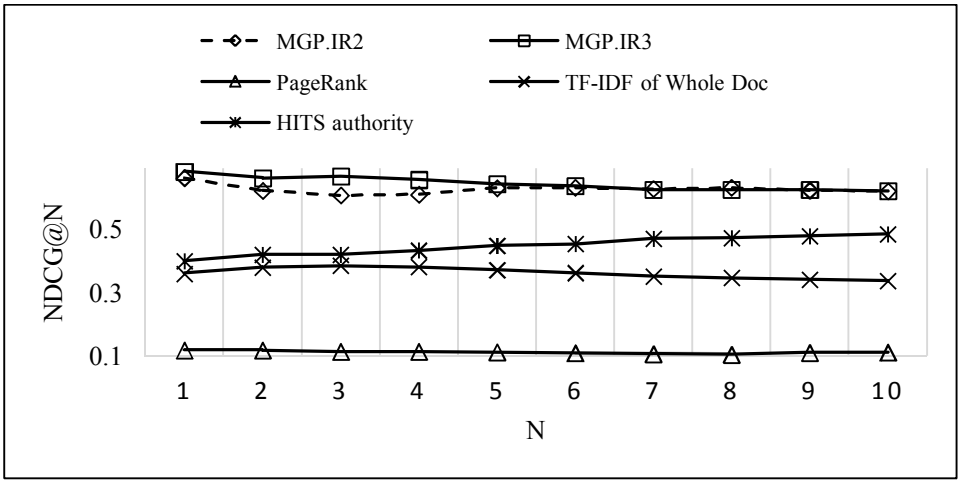
جدول ۱۱: نتایج ارزیابی نسخه‌های الگوریتم MGP-Rank طبق شاخص NDCG روی مجموعه داده dotIR

Mean NDCG	NDCG @10	NDCG @9	NDCG @8	NDCG @7	NDCG @6	NDCG @5	NDCG @4	NDCG @3	NDCG @2	NDCG @1	شناسه ساختار
0.7051	0.5729	0.5773	0.5839	0.5858	0.5899	0.591	0.5815	0.5932	0.61	0.68	MGP.IR1
0.7184	0.619	0.6213	0.6296	0.6257	0.6298	0.6287	0.6079	0.6056	0.62	0.66	MGP.IR2
0.7173	0.6195	0.6231	0.6233	0.6235	0.6347	0.6406	0.6544	0.6648	0.66	0.68	MGP.IR3
0.718	0.6167	0.6189	0.627	0.626	0.6247	0.6274	0.6119	0.6028	0.61	0.64	MGP.IR4

جدول ۱۲: نتایج ارزیابی روش‌های استاندارد رتبه‌بندی بر اساس شاخص NDCG روی مجموعه داده dotIR

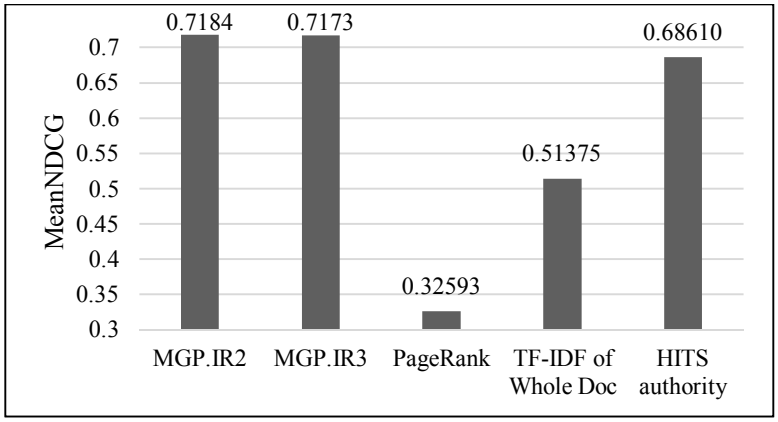
Mean NDCG	NDCG @10	NDCG @9	NDCG @8	NDCG @7	NDCG @6	NDCG @5	NDCG @4	NDCG @3	NDCG @2	NDCG @1	الگوریتم
0.3259	0.1126	0.1121	0.1061	0.1081	0.1107	0.114	0.116	0.1152	0.12	0.12	PageRank
0.5137	0.3378	0.3413	0.3468	0.352	0.3621	0.3711	0.381	0.384	0.38	0.36	TF-IDF of Whole Doc
0.6861	0.4839	0.478	0.4725	0.4703	0.4532	0.4482	0.4328	0.42	0.42	0.4	HITS authority

مقایسه نتایج جدول فوق با عملکرد روش‌های مرجع رتبه‌بندی در شکل شماره نه، انجام شده است که در آن از اطلاعات مربوط به روش‌های مرجع مندرج در جدول شماره دوازده، استفاده شده است.



شکل ۹: مقایسه عملکرد الگوریتم MGP-Rank در زبان فارسی با روش‌های مرجع رتبه‌بندی طبق شاخص NDCG

در شکل شماره ده نیز مقایسه الگوریتم MGP-Rank با روش‌های مرجع رتبه‌بندی در زبان فارسی طبق شاخص MeanNDCG، نمایش داده شده است.



شکل ۱۰: مقایسه الگوریتم MGP-Rank با روش‌های مرجع رتبه‌بندی در زبان فارسی طبق شاخص MeanNDCG

تحلیل نتایج

در این بخش، عملکرد الگوریتم یادگیری رتبه‌بندی MGP-Rank، در زبان فارسی، بر اساس شاخص‌های ارزیابی چهارگانه P@n، MAP، NDCG@n و MeanNDCG، روی مجموعه داده محک dotIR مورد بررسی قرار گرفت. یافته‌های حاصل از نتایج از این ارزیابی‌ها، بطور خلاصه، بشرح زیر است:

- در شاخص دقت نقطه‌ای (P@n)، عملکرد الگوریتم MGP-Rank، کاملاً برتر از روش‌های روش‌های مرجع رتبه‌بندی می‌باشد. به عنوان مثال، الگوریتم MGP-Rank، در شاخص P@1، بهبودی معادل ۷۰٪ در مقایسه با برترین روش مرجع رتبه‌بندی، یعنی HITS authority داشته است. به همین ترتیب، در شاخص P@2، نیز افزایش دقتی به ترتیب معادل ۷۰٪ نسبت به بهترین روش مرجع رتبه‌بندی، بدست آمده است.
- در شاخص متوسط میانگین دقت (MAP)، عملکرد الگوریتم MGP-Rank، از روش‌های مرجع رتبه‌بندی بهتر بوده است. در این خصوص، الگوریتم فوق، افزایش دقتی معادل ۲,۰۵٪ نسبت به برترین روش مرجع یعنی HITS authority داشته است.
- در شاخص NDCG@n، نیز عملکرد الگوریتم MGP-Rank، به طرز چشمگیری بهتر از روش‌های مرجع بوده است. به عنوان نمونه، به ازای شاخص NDCG@1، این الگوریتم، بهبود عملکردی معادل ۶۵٪ نسبت به بهترین روش مرجع رتبه‌بندی، یعنی HITS authority، کسب نموده است.
- در شاخص MeanNDCG نیز روش مورد بحث، موفق‌تر از روش‌های مرجع رتبه‌بندی ظاهر شده‌است. در این زمینه، الگوریتم MGP-Rank، حدود ۴,۷۱٪ نسبت به بهترین روش مرجع رتبه‌بندی، یعنی HITS authority، بهبود عملکرد نشان داده است.

یک نکته جالب توجه در بررسی عملکرد الگوریتم MGP-Rank روی مجموعه داده dotIR، این است که استفاده از شاخص متوسط میانگین دقت (MAP) در اولویت‌بندی ویژگی‌ها برای ادغام آنها بکمک عملگرهای میانگین‌گیری مرتب وزن‌دار، مناسب‌تر از دیگر شاخص‌ها بوده است. همچنین بر اساس آمار فوق، می‌توان به این نکته نیز اشاره نمود که الگوریتم یادگیری رتبه‌بندی MGP-Rank در زبان فارسی نیز همانند زبان انگلیسی، موفق بوده است. مساله شایان توجه دیگر این است که در زبان فارسی نیز بهترین عملکرد این الگوریتم، مربوط به بخش آغازین فهرست نتایج بوده است که همانطور که پیش‌تر نیز ذکر شد، بیشترین میزان توجه کاربران جویشگرهای وب را به خود، معطوف می‌سازد. این عملکرد مناسب الگوریتم MGP-Rank در زبان فارسی، با بکارگیری خصوصیاتی مشابهی همچون TF-IDF، BM25، PageRank، HITS، طول سند و نیز خصوصیات مرتبط با مدل‌های زبانی، محقق شده است.

۵- جمع‌بندی و پیشنهاد پژوهش‌های بعدی

با توجه به نقش تعیین کننده جویشگرهای وب در دسترسی کاربران به اطلاعات و خدمات مورد نیاز، فرآیند رتبه‌بندی در این خصوص، بسیار تاثیر گذار است. با عنایت به کاربر-محور بودن فرآیند رتبه‌بندی نتایج جستجو، بکارگیری اطلاعات مربوط به سوابق رفتار کاربران جویشگرهای وب در حین جستجوی اطلاعات مورد نیاز که اصطلاحاً اطلاعات کلیک از گذر داده نامیده می‌شود، در بهبود عملکرد جویشگرها بسیار مفید می‌باشد. با این وجود، ارایه و در نتیجه، استفاده از این قبیل اطلاعات، در اغلب مجموعه‌های داده محک موجود برای یادگیری رتبه‌بندی و به تبع آن، در اکثر روش‌های مطرح شده در این زمینه، مغفول مانده است. همچنین، تعدد ویژگی‌های ارایه شده در این مجموعه‌های داده، ضمن تحمیل هزینه‌های محاسباتی قابل توجه به روش‌های رتبه‌بندی مطرح شده، کاربرد آنها را در شرایط واقعی، دشوار می‌کند.

به منظور پرداختن به این چالش‌ها، در تحقیقات صورت گرفته، رویکرد نوینی برای حل مساله ایجاد یادگیری رتبه‌بندی، ارایه شده است [۲۳] که طی آن، با نگاه مفهومی به اطلاعات کلیک از گذر داده [۲۱]، از ویژگی‌های موجود در مجموعه‌های داده محک مربوط به یادگیری رتبه‌بندی، تعداد معدودی ویژگی‌های ثانویه موسوم به «ویژگی‌های کلیک از گذر داده» تولید می‌شود. این ویژگی‌ها دارای غنای اطلاعاتی مناسب، به منظور ایجاد روش‌های رتبه‌بندی جدید، می‌باشند. در فرآیند تولید ویژگی‌های کلیک از گذر داده، از مدل مفهومی اطلاعات کلیک از گذر داده، الگو برداری شده است. این ویژگی‌ها که تعداد آنها، هشت عدد می‌باشد، در سه دسته ویژگی‌های مرتبط با پرس‌وجوهای کاربران، ویژگی‌های در رابطه با نتایج جستجوها و نیز ویژگی‌های وابسته به الگوهای کلیک کاربران در حین بررسی نتایج، دسته‌بندی می‌شوند. در تحقیقات پیشین، به منظور محاسبه این ویژگی‌ها از اطلاعات موجود در مجموعه‌های داده محک مورد استفاده، تعدادی سناریوی محاسبه، پیشنهاد و بررسی شده است که عمدتاً محدود به اطلاعات و محتوای انگلیسی وب بوده است. به منظور بکارگیری این ایده در زبان فارسی، در این پژوهش، با عنایت به ویژگی‌های زبان فارسی، سناریوهای مناسبی برای تولید ویژگی‌های کلیک از گذر داده، شناسایی شده است. سپس با بکارگیری نیز مدل برنامه‌نویسی ژنتیک چند لایه، روی ویژگی‌های کلیک از گذر داده، از الگوریتم MGP-Rank، استفاده شده است [۲۳]. بطور مشخص، در این الگوریتم، از قابلیت جستجوی فراگیر مدل برنامه‌نویسی ژنتیک چند لایه و چند جمعیتی، برای یافتن توابع رتبه‌بندی مناسب اسناد از طریق ادغام ویژگی‌های کلیک از گذر داده، بهره گرفته می‌شود. با توجه به گزارش موفقیت قابل توجه این الگوریتم در رتبه‌بندی اطلاعات انگلیسی وب، در این مقاله تلاش شده است با پیشنهاد سناریوهای مناسب جهت استخراج ویژگی‌های کلیک از گذر داده روی دادگان فارسی وب، توانمندی آن در زبان فارسی نیز بکار گرفته شود. ارزیابی عملکرد این الگوریتم در بازیابی اطلاعات فارسی وب، نشان‌دهنده افزایش دقت قابل توجه این الگوریتم، نسبت به الگوریتم‌های پایه رتبه‌بندی می‌باشد. این بهبود عملکرد بخصوص در بخش آغازین فهرست نتایج جستجو که بیشتر مورد توجه کاربران است، قابل ملاحظه‌تر بوده است.

این پژوهش از جنبه‌های مختلفی قابل توسعه است. بصورت خاص، به منظور انجام تحقیقات جامع‌تر پیرامون MGP-Rank، ارایه یک مدل استاندارد و خودکار جهت تولید و سنجش عملکرد سناریوهای محاسبه ویژگی‌های کلیک از گذر داده، بر اساس ساختار و مشخصات و نیز ویژگی‌های عرضه شده در مجموعه‌های داده محک پایه مورد استفاده، حائز اهمیت است. در عین حال، بکارگیری روش‌های مختلف ترکیب اطلاعات در سناریوهای محاسبه ویژگی‌های مبتنی بر اطلاعات کلیک از گذر داده، می‌تواند منجر به افزایش توانمندی این الگوریتم گردد. همچنین با عنایت به این نکته که در فرآیند یادگیری رتبه‌بندی، تولید مدل رتبه‌بندی بر مبنای قضاوت کاربران، صورت می‌گیرد و این قضاوت‌های می‌تواند همراه با ابهام و نایقینی باشد، لذا بررسی روش‌های مواجهه با نایقینی و ابهام ذاتی موجود در قضاوت‌های انسانی، در راستای بهبود عملکرد این الگوریتم‌ها، حائز اهمیت خواهد بود. علاوه بر آن، بکارگیری انواع دیگر توابع سازگاری مرتبط با حوزه بازیابی اطلاعات، بخصوص توابع مرتبط با شاخص NDCG، می‌تواند جالب توجه باشد.

پیوست‌ها

پیوست ۱: فهرست ویژگی‌های موجود در مجموعه داده محک dotIR (Darrudi et al. 2009)

شناسه ویژگی	نام ویژگی	نوع ویژگی
F1	Term frequency (TF) of body	مبتنی بر محتوا
F2	TF of anchor	
F3	TF of title	
F4	TF of URL	
F5	TF of whole document	
F6	Inverse document frequency (IDF) of body	
F7	IDF of anchor	
F8	IDF of title	
F9	IDF of URL	
F10	IDF of whole document	
F11	TF×IDF of body	
F12	TF×IDF of anchor	
F13	TF×IDF of title	
F14	TF×IDF of URL	
F15	TF×IDF of whole document	
F16	Document length (DL) of body	
F17	DL of anchor	
F18	DL of title	
F19	DL of URL	
F20	DL of whole document	
F21	BM25 of body	
F22	BM25 of anchor	
F23	BM25 of title	
F24	BM25 of URL	
F25	BM25 of whole document	
F26	LMIR.ABS of body	
F27	LMIR.ABS of anchor	
F28	LMIR.ABS of title	
F29	LMIR.ABS of URL	
F30	LMIR.ABS of whole document	
F31	LMIR.DIR of body	
F32	LMIR.DIR of anchor	
F33	LMIR.DIR of title	
F34	LMIR.DIR of URL	
F35	LMIR.DIR of whole document	
F36	LMIR.JM of body	
F37	LMIR.JM of anchor	
F38	LMIR.JM of title	
F39	LMIR.JM of URL	
F40	LMIR.JM of whole document	
F41	Sitemap based term propagation	ترکیب محتوا و پیوند
F42	Sitemap based score propagation	
F43	Hyperlink base score propagation weighted in-link	
F44	Hyperlink base score propagation weighted out-link	
F45	Hyperlink base score propagation uniform out-link	
F46	Hyperlink base feature propagation weighted in-link	
F47	Hyperlink base feature propagation weighted out-link	
F48	Hyperlink base feature propagation uniform out-link	
F49	HITS authority	مبتنی بر پیوند

	HITS hub	F50
	PageRank	F51
	In-link number	F52
	Out-link number	F53
	Number of slash in URL	F54
	Length of URL	F55
	Number of child page	F56

منابع

1. Q. Ai, K. Bi, J. Guo, W.B. Croft Croft, "*Learning a Deep Listwise Context Model for Ranking Refinement*", In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 135-144, 2018.

2. Q. Ai, X. Wang, S. Bruch, N. Golbandi, M. Bendersky, and M. Najork, "*Learning Groupwise Multivariate Scoring Functions Using Deep Neural Networks*", In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, pp. 85-92, 2019.

3. C.C. Alves, M.A. Gonçalves, D. Sousa, and T. Salles, "*Generalized BROOF-L2R: A General Framework for Learning to Rank Based on Boosting and Random Forests*", In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, pp. 95-104, 2016.

4. Z. Cao, T. Qin, T.Y. Liu, M.F. Tsai, and H. Li, "*Learning to rank: from pairwise approach to listwise approach*", In Proceedings of the 24th International Conference on Machine Learning, pp. 129-136, 2007.

5. S. Chakrabarti, R. Khanna, U. Sawant, and C. Bhattacharyya, "*Structured learning for nonsmooth ranking losses*", In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 88-96, 2008.

6. O. Chapelle & Y. Chang, "*Yahoo! Learning to Rank Challenge Overview*", Journal of Machine Learning Research, pp. 14, 1-24, 2011.

7. O. Chapelle & M. Wu, "*Gradient descent optimization of smoothed information retrieval metrics*", Information Retrieval, Vol. 13, No. 3, pp. 216-235, 2010.

8. W. Chu & Z. Ghahramani, "*Preference learning with Gaussian processes*", In Proceedings of the 22nd International Conference on Machine Learning, pp. 137-144, 2005.

9. W. Chu & S.S. Keerthi, "*New approaches to support vector ordinal regression*", In Proceedings of the 22nd International Conference on Machine Learning, pp. 145-152, 2005.

10. D. Cossock & T. Zhang, "*Subset ranking using regression*", In Proceedings of the 19th annual conference on Learning Theory, pp. 605-619, 2006.

11. F. Dammak, H. Kammoun, and A.B. Hamadou, "*Improving pairwise learning to rank algorithms for document retrieval*", In of IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1-8, 2017.

12. E. Darrudi, H.B. Hashemi, A. Aleahmad, A. Habibian, A. ZarehBidoki, A. Shakery, and M. Rahgozar, "*A standard web test collection for IR domain*", Technical Reoprt, Iran Telecommunication Research Center, 2009.

13. W. Fan, M.D. Gordon, and P. Pathak, "*Ranking function optimization for effective web search by genetic programming: an empirical study*", In Proceedings of the 37th Hawaii International Conference on System Sciences, pp. 1-8, 2004.

14. W. Fan, M.D. Gordon, and P. Pathak, "*On linear mixture of expert approaches to information retrieval*", Decision Support System, Vol. 42, No. 2, pp. 975-987, 2006.

15. N. Fuhr, "*Optimum polynomial retrieval functions based on the probability ranking principle*", ACM Transactions on Information Systems, 183-204, 1989.
16. J. Gao, H. Qi, X. Xia, and J.Y. Nie, "*Linear discriminant model for information retrieval*", In Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 290-297, 2005.
17. Z. Hu, Y. Wang, Q. Peng, and H. Li, "*Unbiased LambdaMART: An Unbiased Pairwise Learning-to-Rank Algorithm*", In Proceedings of the World Wide Web Conference, pp. 2830-2836, 2019.
18. K. Järvelin & J. Kekäläinen, "*IR evaluation methods for retrieving highly relevant documents*", In Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 41-48, 2000.
19. K. Järvelin & J. Kekäläinen, "*Cumulated gain-based evaluation of IR techniques*", ACM Transactions on Information Systems, Vol. 20, No. 4, pp. 422-446, 2002.
20. X.B. Jin, G.G. Geng, G.S. Xie, and K. Huang, "*Approximately optimizing NDCG using pair-wise loss*", Information Sciences, Vol. 453, pp. 50-65, 2018.
21. T. Joachims, "*Optimizing search engines using clickthrough data*", In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 133-142, 2002.
22. T. Joachims, L.A. Granka, B. Pan, H.A. Hembrooke, and G .Gay, "*Accurately Interpreting Clickthrough Data as Implicit Feedback*", In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 154-161, 2005.
23. A.H. Keyhanipour, B. Moshiri, F. Oroumchian, M. Rahgozar, and K. Badie, "*Learning to rank: new approach with the layered multi-population genetic programming on click-through features*", Genetic Programming and Evolvable Machines, Vol. 17, pp. 203-230, 2016.
24. M. Köppel, A. Segner, M. Wagener, L. Pensel, A. Karwath, and S. Kramer, "*Pairwise Learning to Rank by Neural Networks Revisited: Reconstruction, Theoretical Analysis and Practical Performance*", arXiv:1909.02768v1, 2019.
25. P. Li, Q. Wu, and C.J. Burges, "*McRank: Learning to rank using multiple classification and gradient boosting*", Advances in Neural Information Processing Systems 20, pp. 845-852, 2008.
26. J.Y. Lin, H.R. Ke, B.C. Chien, and W.P. Yang, "*Classifier design with feature selection and feature extraction using layered genetic programming*", Expert Systems with Applications, Vol. 34, 1384-1393, 2008.
27. Y. Lin, J. Wu, B. Xu, K. Xu, and H. Lin, "*Learning to rank using multiple loss functions*", International Journal of Machine Learning and Cybernetics, Vol. 10, pp. 485-494, 2019.
28. T.Y. Liu, *Learning to Rank for Information Retrieval*, Berlin: Springer-Verlag, 2011.
29. H. Liu, Z. Wu, X. Zhang, "*CPLR: Collaborative pairwise learning to rank for personalized recommendation, Knowledge-Based Systems*", Vol. 148, pp. 31-40, 2018.
30. C. Macdonald, I. Ounis, "*Usefulness of Quality Click-through Data for Training*", In Proceedings of the 2009 workshop on Web Search Click Data, pp. 75-79, 2009.
31. C. Macdonald, R.L. Santos, I. Ounis, "*On the Usefulness of Query Features for Learning to Rank*", In Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 2559-2562, 2012.
32. C.D. Manning, P. Raghavan, and H. Schütze, *An Introduction to Information*

41. A. Trotman, "*Learning to rank*", Information Retrieval, Vol. 8, No. 3, pp. 359-381, 2005.
42. M.F. Tsai, T.Y. Liu, T. Qin, H.H. Chen, and W.Y. Ma, "*Frank: a ranking method with fidelity loss*", In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 383-390, 2007.
43. M.N. Volkovs & R.S. Zemel, "*Boltzrank: learning to maximize expected ranking gain*", In Proceedings of the 26th International Conference on Machine Learning, pp. 1089-1096, 2009.
44. J. Wang, L. Yu, W. Zhang, Y. Gong, Y. Xu, and B. Wang, "*IRGAN: A Minimax Game for Unifying Generative and Discriminative Information Retrieval Models*", In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 515-524, 2017.
45. T. Xia, S. Zhai, and S. Wang, "*Analysis of Regression Tree Fitting Algorithms in Learning to Rank*", arXiv:1909.05965v1, 2019.
46. J. Xu, T.Y. Liu, M. Lu, H. Li, and W.Y. Ma, "*Directly optimizing IR evaluation measures in learning to rank*", In Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 107-114, 2008.
47. Q. Xu, M. Li, and M. Yu, "*Learning to rank with relational graph and pointwise constraint for cross-modal retrieval*", Soft Computing, Vol. 23, pp. 9413-9427, 2019.
48. J.Y. Yeh, J.Y. Lin, H.R. Ke, and W.P. Yang, "*Learning to Rank for Information Retrieval Using Genetic Programming*", In Proceedings of the SIGIR 2007 Workshop on Learning to Rank for Information Retrieval, pp. 1-8, 2007.
49. J.Y. Yeh, J.Y. Lin, H.R. Ke, and W.P. Yang, "*Learning to Rank for Information Retrieval Using Layered Multi-Population Retrieval*, Cambridge University Press, 2008.
33. R. Nallapati, "*Discriminative models for information retrieval*", In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 64-71, 2004.
34. L. Pang, J. Xu, Q. Ai, Y. Lan, X. Cheng, and J. Wen, "*SetRank: Learning a Permutation-Invariant Ranking Model for Information Retrieval*", arXiv:1912.05891v1, 2019.
35. T. Qin, T.Y. Liu, J. Xu, and H. Li, "*LETOR: Benchmark dataset for research on learning to rank for information retrieval*", In Proceedings of the LR4IR 2007, in conjunction with SIGIR 2007, pp. 3-10, 2007.
36. T. Qin, T.Y. Liu, and H. Li, "*A general approximation framework for direct optimization of information retrieval measures*", Information Retrieval, Vol. 13, No. 4, pp. 375-397, 2009.
37. R. Rahimi, A. Montazeralghaem, and J. Allan, "*Listwise Neural Ranking Models*", In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, pp. 101-104, 2019.
38. E. Renshaw, A. Lazier, C. Burges, T. Shaked, M. Deeds, N. Hamilton, and G. Hullender, "*Learning to rank using gradient descent*", In Proceedings of the 22nd International Conference on Machine Learning, pp. 89-96, 2005.
39. S. Tan, Z. Zhou, and P. Li, "*Fast Item Ranking under Neural Network based Measures*", In Proceedings of the 13th International Conference on Web Search and Data Mining, pp. 591-599, 2020.
40. M. Taylor, J. Guiver, S. Robertson, and T. Minka, "*Softrank: optimising non-smooth rank metrics*", In Proceedings of the 1st International Conference on Web Search and Web Data Mining, pp. 77-86, 2008.

Genetic Programming", In Proceedings of the 2012 IEEE International Conference on Computational Intelligence and Cybernetics, pp. 45-49, 2012.

50. Y. Yue, T. Finley, F. Radlinski and T. Joachims, "*A support vector method for optimizing average precision*", In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 271-278, 2007.

51. Z. Zheng, H. Zha, and G. Sun, "*Query-level learning to rank using isotonic regression*", In Proceedings of the SIGIR 2008 Workshop on Learning to Rank for Information Retrieval, pp. 9-14, 2008.

52. W. Zhou, J. Li, Y. Zhou, M.H. Momen, "*Bayesian pairwise learning to rank via one-class collaborative filtering*", Neurocomputing, Vol. 367, pp. 176-187, 2019.

53. X. Zhu & D. Klabjan, "*Listwise Learning to Rank by Exploring Unique Ratings*", In Proceedings of the 13th International Conference on Web Search and Data Mining, pp. 798-806, 2020.

54. O. Zoeter, M. Taylor, E. Snelson, J. Guiver, N. Craswell, N., and M. Szummer, "*A decision theoretic framework for ranking using implicit feedback*", In Proceedings of the SIGIR 2008 Workshop on Learning to Rank for Information Retrieval, pp. 24-31, 2008.