



بهینه‌سازی اجرای پرس‌وجوها در پایگاه داده‌های رابطه‌ای با الگوریتم تکاملی ترکیبی

کیون اصغری^۱، علی صفری‌مقانی^۱، فریبرز محمودی^{۱*}، محمدرضا میدی^۲

^۱ دانشکده مهندسی برق، رایانه و فناوری اطلاعات دانشگاه آزاد اسلامی قزوین

^۲ دانشکده مهندسی کامپیوتر دانشگاه صنعتی امیر کبیر

دریافت مهر ۱۳۸۶، تجدید نظر بهمن ۱۳۸۶، پذیرش فوری ۱۳۸۷

چکیده

بهینه‌سازی پرس‌وجوهای پایگاه‌داده‌ای، یکی از مسائل تحقیقاتی مشکل است. تکنیک‌های جستجوی جامع مانند: برنامه‌نویسی پویا برای پرس‌وجوهای با تعداد روابط کم، مناسب هستند ولی با افزایش تعداد روابط موجود در پرس‌وجو، بدلیل نیاز به مصرف زیاد حافظه و پردازش، استفاده از این روش‌ها مناسب نخواهد بود، بنابراین مجبوریم از روش‌های تصادفی و تکاملی استفاده کنیم. استفاده از روش‌های تکاملی بدلیل کارایی و قدرتمندی آنها، تبدیل به ناحیه تحقیقاتی مناسبی در زمینه بهینه‌سازی پرس‌وجو گردیده است. در این مقاله یک الگوریتم تکاملی ترکیبی برای حل مساله بهینه‌سازی ترتیب اجرای عملگرهای پیوند در پرس‌وجوهای پایگاه داده‌ای پیشنهاد شده است. این الگوریتم از دو روش الگوریتم‌های زنگنه‌کی و آتاماتاهای یادگیر بطور همزمان برای جستجو در فضای حالات مستله استفاده می‌نماید. در این مقاله، نشان داده شده است که با استفاده همزمان از آتاماتاهای یادگیر و الگوریتم‌های زنگنه‌کی در فرایند جستجو، سرعت رسیدن به جواب افزایش پیدا کرده و از بدام افتادن الگوریتم در مینیمم‌های محلی جلوگیری می‌شود. نتایج آزمایش‌ها، برتری الگوریتم ترکیبی را نسبت به روش‌های مبتنی بر الگوریتم زنگنه‌کی و آتاماتای یادگیر نشان می‌دهد.

واژه‌های کلیدی: بهینه‌سازی پرس‌وجو، عملگر پیوند، آتاماتای یادگیر، الگوریتم زنگنه‌کی

۱. مقدمه

تکنیک‌های جستجوی جامع، از لحاظ حافظه و زمان پرهزینه خواهند بود. پرس‌وجوهای با تعداد پیوند زیاد در سیستم‌های جدید مانند سیستم‌های مدیریت پایگاه داده استنتاجی^۱، سیستم‌های خبره^۲، سیستم‌های مدیریت پایگاه داده‌ای مهندسی

مدل داده‌ای رابطه‌ای توسط کاد [۱]، معرفی شده است و در سالهای اخیر، سیستم‌های پایگاه‌داده رابطه‌ای بعنوان استانداردی در انواع کاربردهای علمی و تجاری، شناخته می‌شوند. کار بر روی عملگر پیوند، بدلیل هزینه‌های ارزیابی بالای این عملگر، هدف اولیه بهینه‌سازهای پرس‌وجوی رابطه‌ای است. اگر پرس‌جودا در حالت محاوره‌ای باشند، شامل تعداد کمی رابطه خواهد بود که بهینه‌سازی این عبارات را می‌توان بوسیله یک جستجوی جامع انجام داد.

اما در صورتی که تعداد روابط بیش از ۵ یا ۶ رابطه باشد،

*: نویسنده مکاتبه‌گذار

۱ Deductive

2 System Expert

System-R در اکثر موقع دارای هزینه کمتری می‌باشد. از ویژگی‌های دیگر این الگوریتم، قابلیت بکارگیری آن در معماری موازی می‌باشد. از دیگر کارهای صورت گرفته مبتنی بر الگوریتم زنتیکی، روش ارایه شده توسط استینبران و همکارانش بوده است [۷] که روش‌های کدگذاری و عملگرهای زنتیکی متفاوتی را بکار گرفته‌اند. نمونه دیگر از الگوریتم‌های تکاملی بکار گرفته شده برای مسئله بهینه‌سازی پیوندها، روش برنامه‌نویسی زنتیکی می‌باشد که توسط استیلگر و اسپیلیوپولو [۱۶] مطرح شده است. بهینه‌ساز زنتیکی CGO نیز توسط مولرو و همکارانش در [۱۷] ارایه شده است. در این مقاله یک الگوریتم تکاملی ترکیبی برای حل مساله بهینه‌سازی ترتیب اجرای عملگرهای پیوند در پرس‌و‌جواب‌های پایگاه داده‌ای پیشنهاد کرده‌ایم. این الگوریتم از دو روش الگوریتم‌های زنتیکی و آتماتاها‌ی یادگیر بطور همزمان برای جستجو در فضای حالات مسئله استفاده می‌نماید. نشان داده شده است که با استفاده همزمان از آتماتاها‌ی یادگیر و الگوریتم‌های زنتیکی در فرایند جستجو، سرعت رسیدن به جواب افزایش پیدا کرده و از بدام افتادن الگوریتم در مینیمم‌های محلی جلوگیری می‌شود. نتایج آزمایش‌ها، برتری الگوریتم ترکیبی را نسبت به روش‌های مبتنی بر الگوریتم زنتیکی و آتماتا‌ی یادگیر نشان می‌دهد.

ادامه مقاله بدین صورت سازماندهی شده است که بخش دوم، مساله ترتیب عملگرهای پیوند در پرس‌و‌جواب‌های پیوندی را تعریف می‌کند. توضیح مختصری درباره آتماتاها‌ی یادگیر و الگوریتم‌های زنتیکی در بخش ۳ آمده است. در بخش ۴، الگوریتم ترکیبی پیشنهادی در این مقاله شرح داده می‌شود و در بخش ۵، نتایج آزمایش‌ها ارایه می‌گردد. بخش ۶ نیز شامل نتیجه‌گیری می‌باشد.

۲. تعریف مسئله

بهینه‌سازی پرس‌و‌جواب، فعالیتی است که طی آن، یک طرح کارا برای اجرای پرس‌و‌جواب (qep)^۳، تولید می‌گردد و یکی از مراحل اساسی در پردازش پرس‌و‌جواب می‌باشد. در این مرحله، سیستم مدیریت پایگاه داده از بین تعدادی طرح اجرا، بهترین طرح را بر می‌گزیند. بگونه‌ای که اجرای پرس‌و-

CAD/CAM)، سیستم‌های پشتیبانی تصمیم، داده‌کاوی^۴ و سیستم‌های مدیریت پایگاه‌داده‌های علمی و ... دیده می‌شوند. به هر حال سیستم‌های مدیریت پایگاه‌داده^۵، نیازمند بکارگیری تکنیک‌های بهینه‌سازی پرس‌و‌جواب با هزینه کم برای مقابله با چنین پرس‌و‌جواب‌های پیچیده‌ای هستند. دسته‌ای از الگوریتم‌های جستجوی ترتیب مناسب برای اجرای عملگرهای پیوند، الگوریتم‌های قطعی می‌باشد که کل فضای حالات را بصورت کامل جستجو می‌کنند و بعضاً با بکارگیری روش‌های مکاشفه‌ای^۶، این فضا را کاهش می‌دهند [۷]. یکی از این الگوریتم‌ها، روش برنامه‌نویسی پیویاست^۷ که اولین بار برای بهینه‌سازی ترتیب پیوند در System-R توسط سلینگر و همکارانش مطرح گردید [۹، ۲۰]. مهمترین ایراد الگوریتم این است که با افزایش تعداد روابط موجود در پرس‌و‌جواب، نیازمند مصرف زیاد حافظه و پردازنده می‌باشد. از دیگر الگوریتم‌های قطعی می‌توان به الگوریتم حداقل قابلیت انتخاب^۸ [۷]، الگوریتم KBZ [۱۰] و الگوریتم AB [۱۱] اشاره کرد. جهت نشان دادن ناتوانی و ضعف الگوریتم‌های قطعی در مقابله با پرس‌و‌جواب‌های بزرگ، الگوریتم‌های دیگری به نام الگوریتم‌های تصادفی معرفی شده‌اند. الگوریتم‌های مطرح شده در این زمینه، الگوریتم بهمود مکرر^۹ [۵، ۶، ۱۲]، الگوریتم نرم کردن شبیه‌سازی شده^{۱۰} [۵، ۱۲، ۱۳، ۱۴]، بهینه‌سازی دو مرحله‌ای^{۱۱} [۱۲]، نرم کردن شبیه‌سازی شده گردشی^{۱۲} [۸] و نمونه‌گیری تصادفی^{۱۳} [۱۵] می‌باشد.

با توجه به طبیعت الگوریتم‌های تکاملی و اینکه در اکثر موقع مقاوم و کارتر می‌باشند و با درنظر گرفتن کارهای صورت گرفته در این زمینه، مناسب‌ترین گزینه برای حل این مسئله، استفاده از الگوریتم‌های تکاملی می‌باشد. اولین کار انجام شده بر روی مسئله بهینه‌سازی ترتیب پیوندها با استفاده از الگوریتم زنتیکی، توسط بنت و همکارانش ارایه گردید [۲]. در حالت کلی الگوریتم بکار رفته توسط آنها در مقایسه با الگوریتم برنامه‌نویسی پیویاسی بکار رفته برای

3 Decision Support system- DSS

4 Data Mining

5 Database Management System-DBMS

6 Heuristic

7 Dynamic Programming

8 Minimum Selectivity

9 Iterative Improvement -II

10 Simulated Annealing

11 Two -Phase Optimization

12 Tabu Search Simulated Annealing

13 Random Sampling