A Novel Approach to Distributed Rule Matching and Multiple Firing Based on MapReduce

Tianyang Dong, College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China

Qiang Cheng, College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China

Bin Cao, College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China

Jianwei Shi, College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China

ABSTRACT

In order to solve the poor performance problem of massive rules reasoning, as well as the inconsistency problem of working memory in distributed rule matching, this article presents the formal definition of interference relations between rules, and proposes a novel approach to distributed rule matching and multiple firing based on MapReduce. This approach adopts the way of access request control to detect and exclude interference rules, then selects several rule instantiations to perform multiple firing and concurrent execution, thus reducing the number of inference cycles effectively. By detecting the interferences between rules, this method selects and executes compatible rule sets, and avoids the inconsistency problem of system working memory. In order to verify the validity of the authors' approach, this article develops a production system based on MapReduce, and applied this approach in the master server of a distributed production system. The experimental results show that their method can promote the performance of massive rules reasoning effectively.

KEYWORDS

Access Request Control, Big Data, Cloud Computing, Mapreduce, Multiple Rule Firing, Rule Matching

INTRODUCTION

Production system, as known as rule engine, is a common way to build expert systems (Giarratano & Riley, 2005). It has been widely used in some fields such as business, science, engineering, manufacturing, and medicine. The production system can adapt to the changing requirements of enterprise information systems, and reduce almost 10% cost of the information systems for enterprises and organizations (Garnter, 2002a; Garnter, 2002b). The agility and easy-to-use of systems promote the development of business rule engine (Kaul, Storey, & Woo, 2017; Batra, 2017). The market penetration ratio of business rule technology is just 20% in 2000, and now it have reached about 80%. However, compared with other types of information systems, production system usually has the problem of poor performances. It is even worse for the massive rules reasoning. Grid and cloud computing have

changed the IT landscape in the way we access and manage IT infrastructures (Li et al.,2015; Lin et al., 2015; Jianwei et al., 2015; Yuyu et al., 2016). Nowadays, both the cloud computing paradigm and MapReduce programming framework have become key enablers for running big data analytics and large-scale compute- and data intensive applications (Palanisamy et al., 2015; Lee et al., 2016; Qi et al., 2015; Eldawy et al., 2016). The way of using cluster or cloud to construct production systems (Petcu, 2005a) can flexibly expand system processing capability by increasing the cluster scale. That can effectively respond to the challenges of massive rule processing.

In order to improve the performance of production system for massive rules reasoning, as well as keeping the consistency of working memory in distributed rule matching, this paper presents the formal definitions of interference relations between rules, and proposes a novel approach to distributed rule matching and multiple firing based on MapReduce. This approach adopts the way of access request control to detect and exclude interference rules, then selects several rule instantiations to perform multiple firing and concurrent execution, thus reducing the number of inference cycle effectively. Through detecting interferences between rules, this approach can select and execute compatible rule set, and finely avoid the inconsistency problem of the system working memory. In order to verify the validity of this approach, this paper developed a distributed production system based on MapReduce, which is known as a rule engine for traffic information service. The approach to distributed rule matching and multiple firing is deployed in the master server of this distributed production system. The experimental results and successful applications in the public-travel traffic information service system show that the approach can effectively improve the performance of massive rules reasoning.

RELATED WORKS

The researches on multiple rule firings are mostly devoted to centralized production system. They try to select as many rules as possible to execute, in order to break the constraints of selecting only one rule caused by the traditional conflict resolution strategies. But there are all kinds of dependencies among the rules, which will lead to the inconsistency problem of working memory after multiple rule execution. So it needs to find out the interferences to be excluded according to the inter-rule dependencies, select compatible rule set to execute, and guarantee the consistency of working memory in the case of multiple rule firings. Ishida (1991) used data dependency graph to represent the referring and processing relations between rules and working memory elements, identified and excluded interference rules by using the corresponding selection algorithm. Schmolze et al. (1992) proposed A2 algorithm to exclude rule instantiations with interference characteristic of disabling and clashing. Kuo et al. (1991) selected compatible rules based on the RTC interference matrix.

The key of parallel rule firing in centralized production system is how to mine the association rules. Therefore, Schmolze (1991) presented a formal solution to the problem of guaranteeing serializable behavior in synchronous parallel production systems that execute many rules simultaneously. Rand et al. (1996) presented an algorithms to explore a spectrum of trade-offs between computation, communication, memory usage, synchronization, and the use of problem specific information. This method solves the problem of mining association rules on a shared nothing multiprocessor. Han et al. (2000) proposed two parallel algorithms for mining association rules: intelligent data distribution algorithm and hybrid distribution algorithm. The hybrid distribution algorithm further improves upon the intelligent data distribution algorithm by dynamically partitioning the candidate set to maintain good load balance. Thabtah et al. (2006) proposed a new associative classification method called Multi-class Classification based on Association Rules (MCAR), which takes advantage of vertical format representation and uses an efficient technique for discovering frequent items. Bădică et al. (2011) surveyed several major approaches using rules in multi-agent systems and distributed agent architectures which run rule engines at their core. Gupta et al. (1988) exploited very fine-grained parallelism to achieve significant speed-ups, which is also distinct from other parallel implementations in that they parallelize a highly optimized C-based implementation of OPS5. It is 10-20 times faster

Related Content

Interactive Indexing of Documents with a Multilingual Thesaurus

Ulrich Schieland Ianna M.S.F. de Sousa (2003). *Effective Databases for Text & Document Management (pp. 24-35).*

www.irma-international.org/chapter/interactive-indexing-documents-multilingual-thesaurus/9203

OOXKSearch: A Search Engine for Answering XML Keyword and Loosely Structured Queries Using OO Techniques

Kamal Tahaand Ramez Elmasri (2009). *Journal of Database Management (pp. 18-50).*

www.irma-international.org/article/ooxksearch-search-engine-answering-xml/4123

Performance Implications of Knowledge Discovery Techniques in Databases

Balaji Rajagopalanand Ravindra Krovi (2003). *Advanced Topics in Database Research, Volume 2 (pp. 191-212).*

www.irma-international.org/chapter/performance-implications-knowledge-discovery-techniques/4346

A Scalable Middleware for Web Databases

Athman Bouguettaya, Zaki Malik, Abdelmounaam Rezguiand Lori Korff (2006). *Journal of Database Management (pp. 20-46).*

www.irma-international.org/article/scalable-middleware-web-databases/3361

An Efficient NoSQL-Based Storage Schema for Large-Scale Time Series Data

Ruizhe Ma, Weiwei Zhouand Zongmin Ma (2024). *Journal of Database Management (pp. 1-21).*

www.irma-international.org/article/an-efficient-nosql-based-storage-schema-for-large-scale-time-series-data/339915