ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

Survey on Large Scale Ontologies Based on Map Reduce by Using IDIM

Tejashree B. Shinde, Prof.G.S.Deokate

Master of Engineering Student, Dept. of Computer Engineering, Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Maharastra, India

Assistant Professor and HOD, Dept. of C.E., Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Maharastra, India

ABSTRACT: With the upcoming data deluge of semantic data, the fast growth of ontology bases has brought significant challenges in performing efficient and scalable reasoning. Traditional centralized reasoning methods are not sufficient to process large ontologies. Distributed reasoning methods are thus required to improve the scalability and performance of inferences. This paper proposes an incremental and distributed inference method for large-scale ontologies by using Map Reduce, which realizes high-performance reasoning and runtime searching, especially for incremental knowledge base. By constructing transfer inference forest and effective assertional triples, the storage is largely reduced and the reasoning process is simplified and accelerated. Finally, a prototype system is implemented on a Hadoop framework and the experimental results validate the usability and effectiveness of the proposed approach.

KEYWORDS: Big data, MapReduce, ontology reasoning, RDF,Semantic Web.

I.INTRODUCTION

In this paper, we propose an incremental and distributedinference method (IDIM) for large-scale RDF datasets via Map Reduce. The choice of Map Reduce is motivated by the fact that it can limit data exchange and alleviate load balancing problems by dynamically scheduling jobs on computing nodes. In order to store the incremental RDF triples more efficiently, we present two novel concepts, i.e., transfer inference forest (TIF) and effective assertional triples (EAT). Their use can largely reduce the storage and simplify the reasoning process. Based on TIF/EAT, we need not compute and store RDF closure, and the reasoning time so significantly decreases that a user's online query can be answered timely, which is more efficient than existing methods to our best knowledge. More importantly, the update of TIF/EAT needs only minimum computation since the relationship between new triples and existing ones is fully used, which is not found in the existing literature. The main contributions of this paper are summarized as follows.

1) We propose a novel representation method TIF/EAT to support incremental inference over large-scale RDF datasets which can efficiently reduce the storage requirement and simplify the reasoning process.

2) An efficient and scalable reasoning method called IDIM is presented based on TIF/EAT, and the corresponding searching strategy is given to satisfy end-users' online query needs.

3) We have implemented a prototype by using the Hadoop platform. It allows one to perform experiments of different methods on billion triples challenge (BTC) benchmark data. A real-world application on healthcare domain is also presented to validate the effectiveness of our method.

II.RELATED WORK

Semantic inference has attracted much attention from both academia and industry nowadays. Many inference engines have been developed to support the reasoning over Semantic Web. For example, Anagnostopoulos and Hadjiefthymiades proposed two fuzzy inference engines based on the knowledge-representation model to enhance the context inference and classification for the well-specified information in Semantic Web. Guo *et al.* introduced a novel Rule XPM approach that consisted of a concept separation strategy and a semantic inference engine on a multiphase forward-chaining algorithm to solve the semantic inference problem in heterogeneous e-marketplace activities. Paulheim and Bizer studied the problem of inference on noisy data and presented the SD Type method based on statistical distribution of types in RDF datasets to deal with noisy data. Milea *et al.* presented a temporal extension of

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

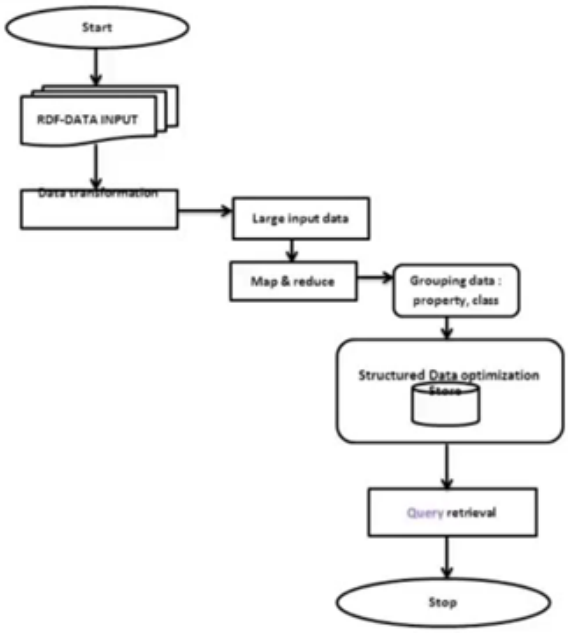
Vol. 3, Issue 12, December 2015

the web ontology language (OWL) for expressing time-dependent information. These ontology reasoning methods are conducted on a single machine or local cluster. The reasoning speed is directly dependent on the scale of the ontology, which is not suitable to a large ontology base. To deal with such large base, some researchers turn to distributed reasoning methods. Weaver and Hendler presented a method for materializing the complete finite RDF closure in a scalable manner and evaluated it on hundreds of millions of triples. Urbani *et al.* proposed a scalable distributed reasoning method for computing the closure of an RDF graph based on Map Reduce and implemented it on top of Hadoop. Schlicht and Stuckenschmidt highlighted the main drawback of the Map Reduce-based reasoning and then introduced Mapresolve method for more expressive logics.

III.PROPOSED ALGORITHM

In this paper, our inference method is based on Map reduce and hadoop platform. Map reduce Is a programming model for parallel and distributed processing Of batch jobs. Each job contains a map and a reduce, in whichThe map phase assigns a key to each element and then partitions. The input data, while the reduce phase processes each partitionIn parallel and merges all intermediate values with the sameKey into final results. By writing the map and reduce functions, A mapreduce programming model provides a more convenient Way for regular programmers to process large datasets. In this we are using following algorithm:

1) K Means Clustering
2) Reasoning Over PTIF
3) Reasoning Over DRTF
4) Reasoning Over CTIF



IV.ACKNOWLEDGEMENT

We would like to take this opportunity to express my sincere gratitude to my Project Guide Prof.G.S.Deokate (Assistant Professor and HOD, Computer Engineering Department, SPCOE) for his encouragement, guidance, and insight throughout the research and in the preparation of this dissertation. He truly exemplifies the merit of technical excellence and academic wisdom.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

V. CONCLUSION AND FUTURE WORK

In the big data era, reasoning on a Web scale becomesincreasingly challenging because of the large volume of data involved and the complexity of the task. Full reasoning over the entire dataset at every update is too time-consuming to be practical. This paper for the first time proposes an IDIM to deal with large-scale incremental RDF datasets to our best knowledge. The construction of TIF and EAT significantly reduces the recomputation time for the incremental inference as well as the storage for RDF triples. Meanwhile, users can execute their query more efficiently without computing and searching over the entire RDF closure used in the prior work. Our method is implemented based on Map Reduce and Hadoop by using a cluster of up to eight nodes. We have evaluated our system on the BTC benchmark and the results show that our method outperforms related ones in nearly all aspects.

REFERENCES

[1] M. S. Marshall *et al.*, "Emerging practices for mapping and linking lifesciences data using RDF—A case series," *J. Web Semantics*, vol. 14, pp. 2–13, Jul. 2012.
[2] M. J. Ibá˜nez, J. Fabra, P. Álvarez, and J. Ezpeleta, "Model checking analysis of semantically annotated business processes," *IEEE Trans. Syst.,Man, Cybern. A, Syst., Humans*, vol. 42, no. 4, pp. 854–867, Jul. 2012.
[3] V. R. L. Shen, "Correctness in hierarchical knowledge-based requirements," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 30, no. 4, pp. 625–631, Aug. 2000.
[4] J. Guo, L. Xu, Z. Gong, C.-P. Che, and S. S. Chaudhry, "Semantic inference on heterogeneous e-marketplace activities," *IEEE Trans. Syst.,Man, Cybern. A, Syst., Humans*, vol. 42, no. 2, pp. 316–330, Mar. 2012.
[5] J. Cheng, C. Liu, M. C. Zhou, Q. Zeng, and A. Ylä-Jääski, "Automatic composition of Semantic Web services based on fuzzy predicate Petri nets," *IEEE Trans. Autom. Sci. Eng.*, Nov. 2013, to be published.
[6] D. Kourtesis, J. M. Alvarez-Rodriguez, and I. Paraskakis, "Semanticbased QoS management in cloud systems: Current status and future challenges,"*Future Gener.Comput. Syst.*, vol. 32, pp. 307–323, Mar. 2014.