MAPREDUCE FRAME WORK FOR LARGE-SCALE DATA SETS BY INCREMENTAL AND DISTRIBUTED INFERENCE METHOD

Shaik Nasrath[1], Alladi Suresh Babu[2]
[1]Gates college, Gooty; CSE
Email:Nasrath.shaik8@gmail.com[1], sureshalladi.cse@jntu.ac.in[2]

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. Distributed reasoning methods on computing RDF closure for reasoning, which takes much time and space (generally the ontology size is more than original data size). Moreover, each time when new RDF arrives, full rereasoning over the entire dataset is needed to compute the new RDF closure. WebPIE newly-arrived RDF triples and old ones but fails consider the relations between them, thus resulting in a number of duplicated triples during the reasoning thereby its performance.

Here proposing an incremental and distributed inference method (IDIM) for large-scale RDF datasets via MapReduce. The choice of MapReduce 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, Two novel concepts will be applied, 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.In order to store the incremental RDF triples more efficiently.

CONCLUSION:

In the big data era, reasoning on a Web scale becomes increasingly challenging because of the large volume of data involved and the complexity of the task. Full rereasoning over the entire dataset at every update is too time-consuming to be practical. This approach 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. This method is implemented based on MapReduce 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.

REFERENCE:

[1] Bo Liu, Member, IEEE, Keman Huang, Jianqiang Li and MengChu Zhou, Fellow, IEEE, " An Incremental and Distributed Inference Method for Large-scale Ontologies based on Mapreduce Paradigm", IEEE TRANSACTIONS ON CYBERNETICS, VOL.45 NO.1, JANUARY 2015.