Semantic Depthness Estimation Based Query Optimization for Data Mining in Parallel and Distributed Systems Using Multi Agent Approach

1K. Syed Kousar Niasi and 2Dr.E. Kannan,
1Department of Computer Science,
Jamal Mohamed College, Trichy.
kounia@yahoo.com
2Vel Tech DR.R.R & DR.SR Tech University,
Chennai.

Abstract

The search time complexity and the data availability is the most common problems in mining information from parallel a distributed systems. Even with the more number of agents the retrieval of relevant information becomes complex where the exact Meta data of the nodes could not be classified in accurate manner. This increase the search time complexity and the methods discussed earlier suffers with the problem of poor mining accuracy. To solve this issue, a novel Ontology Based Semantic Data Depthness Estimation (OBSDDE) technique has been proposed to perform query optimization. The method computes the semantic depthness of data present in different nodes of parallel and distributed systems and based on computed depthness measure, a minimum number of locations with more depthness is selected. The number of agents are generated and initialized according to selected locations to perform query processing. The proposed method increases the efficiency of data mining and reduces the search time complexity with less space complexity also.

Keywords: Depthness, ontology, semantic data, depthness estimation, query search.
1. Introduction

Depthness measure are made out of various items which are in close contact and vigorously impede each other. Few existing Depthness measure acknowledgment frameworks are prepared to do precisely anticipating object postures in such situations. in multi agent approach includes perceiving objects without obliging their geometry or surface properties.

Different Depthness measure acknowledgment techniques exploit invariant nearby element descriptors such to perceive a little arrangement of articles with discriminative surface or geometry. The execution of nearby multi agent approach degrades essentially preparing an inquiry incorporates the cost of required transmissions to spread the question in the system, the cost of handling in multi agent approach and correspondence cost of result messages from the reacting hubs to the base station. Since the correspondence cost between clients is observably more noteworthy than the cost of handling information in a solitary hub, it is conceivable to disregard the preparing cost.

Undertaking has as of late gotten the consideration of analysts in the mixed media, flag preparing, and machine learning groups in light of the huge measure of accessible retagged media on the Web that can be utilized as preparing information, permitting calculations to deal with information volumes once in a while observed some time recently.

Likewise, the undertaking is sufficiently hard to require cooperation between numerous specialists and in different research groups, which is a test all alone plan of stereo with homography related with a semantic chart built from the activity scene.

Semantic relations among every semantic area frame the topological structure to control the robot to do undertakings. The other commitment in this work is a dynamic estimation handle for semantic locale estimation.

Approaches specified above utilized static correlations with accomplish the place acknowledgment. Really, a consistent element process is more appropriate for the robot put acknowledgment since it can construct time-space setting relations framework needs to give a bound together information interface to clients to acknowledge information dissemination and excess straightforwardness.

The information appropriation not just makes information prepare require information transmission of spots, additionally makes parallel handling conceivable multi agent approach that we display another approach and model for concurrent profundity estimation and semantic division from a multi-agent approach, where the two assignments share the fundamental component representation.
2. Related Work

Morteza Zihayat, Zane Zhenhua Hu, et al,(2016) suggested the finest of our knowledge, existing approaches for mining HUSPs do not discourse the subjects of mining HUSPs in big data, and preceding methods in big data do not find high usefulness consecutive patterns. So far, no education has been showed to learn high usefulness sequential patterns in big data, which is more challenging than finding frequent sequences due to the statistic that the arrangement utility does not satisfy the descending closure property.

Hongcheng Lu, et al, (2015) discussed Ontology has the nice hierarchical structure of concept and the support of logical reasoning. Semantic information transfer can be realized through relationships between concepts. In order to sharing and extending flexibility, the construction of domain ontology adopts top-down method. Most widely used construction method is seven-step methodology d by Noy and McGuiness.

Sylvain Sagot, Egon Ostrosi, et al, (2016) suggested objective of the owners is to display their website in these visible positions. For this, the owners can appeal to specialized firm. Two types of techniques are generally used by these firms to improve the website ranking: Exploration Engine Advertising and Search Machine Optimization. The SEA permits to display advertisement in a dedicated space (paid search results) by buying a set of keywords that can be searched by Internet users.

Premprakash Kashyap, et al, (2016) discussed works fail to focus on security and the reliability aspects. There is not much security for the female drivers and passengers. People generally prefers traveling with good people who don't have any bad addiction like smoking while traveling. Generally any smoker's car stink a lot and a non-smoker might not be comfortable to sit in that car. There is no such type of system which focuses on these type of comfort and security.

Atif Khan, et al, (2016) suggested abstractive summarization. Extractive summarization extracts the most important representative sentences from the source documents and groups them to produce a summary. However, abstractive summarization requires natural language processing techniques such as semantic representation, natural language generation, and compression techniques.

Yiannis Kokkinos, et al,(2012) unpredictability is disallowed for extensive scale preparing. Besides the way that the locally imperative shrouded focuses are dependably a segment of information size, makes a similar system to develop vast and slower in operation as information size develops extensive and parallelism become a necessity. However, the hybrid optimization method is not simply amendable for parallelization.

Abderraouf Maoud, et al,(2016) suggested the control is distributed on the
architecture units and task execution on the system components. On the other hand, they are harder to design but in general perform better. Moreover, each robot has its own embedded on-board computer and sends information about the scene to a subset of robots (its neighbors).

M. S. Rahman, et al, (2016) discussed the approaches, so far discussed, in this are functional to relevant cyber security problems, some of these approaches neglect the effects of physical faults from attacks. Moreover, the effect of time delay in agent communications and the failures in such communication links were not discussed properly. In order to address these challenging issues, presents an innovative multi-agent system (MAS).

Rahul Ramakrishna, et al, (2011) suggested the map reduction outline is based on useful programming paradigm, where the plans can be written in summary form, specifying a map meaning which generates middle key value pairs and a reduce purpose merging the key worth pairs. With this method a considerable increase in speed effectiveness is obtained.

Anurag Sharma, et al, (2016) discussed to provide high customer satisfaction and service reliability, the operators need to implement appropriate measures to restore service to the blacked out but un-faulted area. One of the emerging solutions for service restoration to the un-faulted out-of-service area is the implementation of intentional islands powered by distributed generators (DGs hierarchical coordination strategy for multivalent systems is designed for the location, isolation and restoration of faults.

Yong Shui, et al, (2016) suggested provided the evidence that the components in the same age class tend to form a well-connected module. Importantly, these new models agree that more conserved components are more likely to be at the core of densely connected modules. Since a strongly positive correlation between network modularity and evolutionary conservation has been validated, aligning PPI networks of evolutionarily distant species would discover the modules with conserved link patterns.

Oscar De Silva, George K. I, et al, (2012) discussed It is advantageous for multi-robot system designs to reduce the application of SLAM to strategically selected agents while establishing relative localization schemes for the rest of the agents. This also applies to similar ground agents in the team with inferior localization ability due to limited exteroceptive sensing.

Charalampos P. Bechlioulis, et al, (2016) suggested In spite of the fact that most of the takes a shot at appropriated helpful control consider known and straightforward element models, numerous pragmatic designing frameworks exist that neglect to fulfill that presumption. Taking into clarification the trademark show questions when outlining circulated control plans is of vital significance, though reaching out towards this bearing, existing control strategies for perfect models.
Kulvaibhav Kaushik, et al, (2013) discussed User controlled access control: The attribute based searchable encryption provides for limited controlled access to the information. The proprietor characterizes get to arrangement and association it with the ciphertext along these lines constraining access to the information. Only to authorized users. • Non-sharing of keys: To provide controlled access to data among different users, key sharing between multiple users is not secure.

Martin Lévesque and Martin Maier, et al, (2014) succeeds outline and execution rules among frameworks to trade information between brilliant lattice parts. All the more particularly, the standard records four fundamental quality traits: adaptability, interoperability, data dependability, and security. In this work, focus on the reliability quality attribute. Dependability is distinct as the ability to execute a function under given conditions for a given period of time. Information reliability has the following two main data characteristics.

Ion Matei, John S. Baras, et al, (2012) discussed capacities on the condition of the power arrange and figure neighborhood approximations in light of their own abilities and on the approximations of the nearby specialists too. Consolidate the multi-operator cleaning plan with a trust-based gadget under which every specialist associations a trust metric to each of its neighbors. These trust measurements are involved into clarification in the sifting course of action so data transmitted from specialists with low trust is slighted.

Yang Meng, et al, (2016) suggested For the group of multi-agent schemes with numerical networks, the inter-agent message, which aims at gaining neighbors’ state information as detailed as possible, is typically the foundation of scheming the cooperative control laws. In real digital networks, statement channels only have finite capacities and the communication between different agents is a process which consists of encoding.

Shoji Nishimura, et al, (2011) discussed information has two important characteristics. First, it is inherently skewed, both spatially and temporally. For instance, urban regions are denser likened to rural regions, while commercial and school regions are thick during weekdays, while housing areas are dense throughout the night and weekends. Second, the time measurement is potentially unbounded and monotonically growing.

Maen Saleh, et al, (2012) suggested network-layer protocol that is implemented to provide the required security services to all IP based applications. In instruction to found a secure data channel between two parties, IPsec(8,5),(995,993) performs a security association phase, where the required security parameters are initialized such as security service’s algorithm, initialization data, encryption data keys, and security modes. Since our data unit is a realtime data packet, IPsec protocol seems to be suitable for the development.

E. Suhir, et al, (2015) discussed there occurs a insight, perhaps a rather validated one that some EP products or at least some important parts of these
products are highly reliable and never fail. The very existence of such a perception could be attributed to the superfluous and unnecessary robustness of a particular product or a particular component for the given application. When failure occurs, root cause analyses are conducted; origins and causes of failures detected, and appropriate corrective actions are taken.

Qingkai Yang, et al.(2016) suggested investigate the formation tracking problem for multiple single integrator systems associated with undirected graphs, where observer-based controllers are aiming at driving the weighted centroid of the formation to follow some smooth reference signal. The First, to estimate the weighted centroid of the collective states, a continuous finite-time observer is constructed.

3. Materials and Method

The depthness estimation is planned to hide the problems and release the user of a load of export with the crowd. The manipulator is only required to defer to an SQL-like query, and the scheme takes the responsibility of collecting the query, constructing the application plan and approximating in the depthness square. A assumed query can have many auxiliary implementation plans, and the alteration in crowdsourcing cost among the greatest and the worst series may be many instructions of scale. Therefore, as in relational database systems, query optimization is significant to crowdsourcing arrangements that provide declarative query interfaces.

![Figure 3.1: Over All Architecture Diagram](image-url)
The planned Ontology-Based Semantic Data Depthness Estimation (OBSDDE), a cost-based query optimization method for declarative crowdsourcing schemes. OBSDDE reflects both cost and inexpression in query optimization aims and generates query strategies that provide a right stability between the cost and inexpression. We develop effective algorithms in the OBSDDE for enhancing three types of queries: selection queries, join queries, and compound selection-join queries.

**Query Optimization using Multi Agent Approach**

Data mining is an progressively important information for eliminating essential statistics in large groups of data and the thoughtful technique of social systems belongs to the multi agent approach. There are, though, unwanted awareness about data mining, between transactions, data production out, improving information’s put away deal with difficulties which possible privacy attack and possible perspicacity. The closing contains of untruthfully deliberating data on the basis going to a detailed group of information from the dataset. The multi agent approach for data examination with continuous noticing the comprehensive set of recurring patterns in time series in mechanism knowledge approach to the agent records approximating the number of refresh item sets, so to build a query cost perfect for the associated datasets which can be used to estimate the number of datasets calculated incoherency bound with map reduce to overcome the current terms.

**Algorithm**

- **Input:** Query Document QDi
- **Output:** Multi agent Set MAS

**Start**

1. Read Query Document QDi.
2. Split text into term query document set TQDTTs.
3. $TQDTT = \text{join}(T_s, " ");$
4. For each term $T_i$
   
   $T_i = \int_{i=1}^{\text{size}(T_s)} \text{Agent approach}(T_i, \text{data processing});$
5. **End**
6. For each term $T_i$
   
   $T_i = \int \text{attribute for depthness } (T_i)$
7. **End**

**Stop.**

Routine results using practical traces show that our cost based query preparation leads to queries being performed by means of less than one third the amount of messages required by current schemes.
Distributed data sources associated with a business line are often complex, for instance, some is of high frequency or density, mixing static and dynamic data,
mixing multiple structures of data; Data integration and data matching are difficult to conduct; it is not possible to store them in centralized storage and it is not feasible to process them in a centralized manner; In some cases, multiple sources of data are stored in parallel storage systems; Local data sources can be of restricted availability due to privacy, their commercial value, etc., which in many cases also prevents its centralized processing, even in a collaborative.

**Semantic Depthness Estimation**

Depthness exactness is measured as far as incoherency of information thing in value-based database as far reaching modification in the value of the information thing at the information source and the value know at the customer information. In this arranged strategy, the each semantic depthness estimation safeguard its orchestrated incoherency headed for different information thing. To propose depthness estimation incremental calculation for consistent deciding the careful arrangement of regular examples in time arrangement inventories taking after the quantity of invigorate itemsets, to assemble a question cost idealize in view of association which can be utilized to guess the quantity of datasets determined incoherency bound. However we eliminate the cost of the query, important thing is to evaluate the incoherency in the dataset uncontrollable delivery cost. The data altering aspects and incoherency data model are used to approximation the data distribution cost.

**Algorithm**

**Input:** Depthness part Dp, Query Q, Multi agent approach MAA.

**Output:** Depthness Measure Dm.

**Start**

- For each semantic part other than Sp
- Identify the semantic hierarchy Sh.
  
  \[
  Sh = \sum_{i=1}^{size(Q_s)} Shs(i), \text{MAA} > Dp(Qp)
  \]

- Compute number of interactive in need of objects.

  \[
  \text{MAA} = \sum_{i=1}^{size(R_{ms})} Sp(i) \in \sum Dp(Attr)
  \]

- Compute Depthness measure \(Dm = \frac{\text{attribute}}{size(R_{ms})} \times Sh\)

- Add to Sh.

**End**

**Stop**

Mining high utility itemsets from databases mentions to finding the itemsets with high incomes. Here, the meaning of itemsets utility is interestingness, importance, or profitability of an item to users. The most straightforward form of distribution is horizontal partitioning, in which different records are collected at different sites, but each record contains all of the attributes for the object it describes. This is the most common and natural way in which data may be distributed. For example, a multinational company deals with customers in several countries, collecting data about different customers in each country.
The process of mining information from large volume of data set. The data may be present in different location, but the data mining algorithm has to retrieve the exact relevant information from different locations to produce efficient results.

There are many approaches available for data mining; we use mobile agents to retrieve the information from different locations of the network.

The genetic algorithm which uses cross over and mutation operations to select the information extracted and has a fitness function to evaluate the selection process. We apply the genetic algorithm to evaluate the retrieval process.
Depthness is a relational representation of information, where the ontology consists of classes and labels. For each word there are synonyms or relational terms like hotel: Boarding: Lodging, the ontology file consists of classes, labels and similar meanings about a concept. In our case the ontology file consists of classes and related terms about the data and locations. Using these functionalities we propose a new mining approach to increase the efficiency of the retrieval systems.

4. Result and Discussion

A multi agent approach to estimate the semantic similarity between words or entities using web search engines with ranking the search results occur. The new assessment shows that the arranged system is productive and exceeds late successive example incremental mining methods. An incremental calculation for deciding the entire setoff repetitive examples in period succession databases, i.e., we decide the incessant examples over the total time arrangement in contrast to applying a sliding gap over a bit of the time arrangement. The arranged approach can decide visit adornments that include holes between examples' things with a preeminent client characterized hole measure. With the passageway of each new information thing, the system refreshes the present mining comes about incrementally. We depict an arrangement of states for the outlines in the database subject to whether they are repetitive or non-frequent.

![Figure 4.1: Evaluation of Time Complexity](image)

Table 4.1: Time Complexity for Depthness

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Time in Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic Algorithm</td>
<td>22.5</td>
</tr>
<tr>
<td>MADA E</td>
<td>18.9</td>
</tr>
<tr>
<td>Multi Agent</td>
<td>11.2</td>
</tr>
</tbody>
</table>
The above figure is showing the time complexity between the different approaches, also the proposed method shows the minimum time taken for identifying the number of location. A arrangement of queries is made founded upon the nature of the consequence set. For each class, we progress algorithms for calculating probabilistic answers. We address the significant issue of calculating the quality of the answers to these queries, and provide algorithms for professionally pulling data from relevant sensors or moving objects in order to improve the quality of the executing queries.

![Graph of Similarity Analysis](image)

Figure 4.2: Similarity Evolution in Different Location

There are many functions exists to compute the fitness function, we use support function to calculate the fitness values. Based on the fitness value the irrelevant results are identified. The fitness value is computed only once for each node. From generated results we compute the average relevancy with each tuple selected. The average relevancy must be greater than relevancy threshold which is maintained by the fitness function. Then cross over and mutation operations will be performed.

The method to estimate semantic similarity using page counts and text leftovers retrieved from a web search engine for two words. Specifically, we define various word co-occurrence measures using page counts and integrate those with lexical patterns extracted from text snippets. To propose pattern utility incremental algorithm for continuous discovering the complete set of frequent patterns in time series databases estimating the number of refresh item sets, we build a query cost model which can be used to estimate the number of datasets specified incoherency bound to overcome the existing terms.
5. Conclusion

To achieve the current problem that will overcome with constant frequent dataset for mining high practicality transactional directory. To recommend multi agent method for constant determining the complete set of recurrent data in time series records approximating the number of refresh itemsets, we build a query cost predictable which can be used to approximation the number of datasets restrained incoherency bound. Exhibition results using real-world references show that our cost based query training leads to explorations being realized using less than one third the amount of public services obligatory by current preparations and to follow depthness multi agent approach over-all to overcome Mining practicality item sets from gatherings refers to discovery the item sets with high proceeds. Here, the denotation of item set usefulness is interestingness, position, or efficiency of an element to users.

References


