

Study on Systemization of Fuzzy Logic and Engine using Mining Concept

Harish.R¹, Gowtham Prassanaa.N¹, Harshal Golcha¹, Harikrishna Pillutla²

Student, CSE Department, Rajiv Gandhi College of Engineering and Technology, Puducherry, India¹

Assistant Professor, CSE Department, Rajiv Gandhi College of Engineering and Technology, Puducherry, India²

Abstract: In big data on measuring relationships between number of objects in Wikipedia whose pages can be considered as separate objects. Most of the users does not know searching using Google and don't know about Wikipedia website. Previously searched data is used in reducing the real time traffic; thereby it changes traffic situations automatically. In our proposed system we included the distributed scenarios with the global traffic prediction. Using the maximum generalized flow technique [8], the three concepts such as distance, connectivity, and clustering is used in our method. External and internal relationship methods are targeted [5], [4]. Mining elucidatory objects would open a novel way to deeply understand a relationship.

Keywords: Mining Of Wikipedia, separate objects relationship on Mining, Analysis on link, Maximum Generalized Flow.

I. INTRODUCTION

The important aspect of big data is that managed so that it can be able to manage the business requirement a given solution is designed to support. Big data opens both opportunities and challenges for businesses. In order to obtain exact value from big data, it must be efficiently managed and analyzed in a timely manner, and the output that obtained are need to be available in such a way to produce the positive effect change or influence business decisions. The effectiveness also depends on an organization having the right combination of people, process and technology.

The proposed system includes about a detailed and methodological big data survey of related work for measuring relationships or similarities[1],[2]. In connection with real-world data entities, we have found certain characteristics of traffic data, such as spatial focus of accidents, patterns of temporal rush hours, which can be inferred into the data-mining technique to make it much more exact meaning. For example, in the generic time-series, the experimental results that supported the hypothesis are made in the immediate past are usually a good suggest of the short-term future. However, for traffic time series, this is false at the edges of the traffic rush hours. In that case, the historical knowledge domains (perhaps for that same day, time, and location) are the better prediction for the future [3].

By definition, analytics is the introduction and communication of meaningful patterns in data but only for the business, the analytics should be viewed as the wide use of data, the statistics and constituent analysis, using elucidatory and prediction models[3], [4] to take the relation-based business decisions and actions. Analytics helps to get the most out of key processes, functions and roles. It can be accelerated to aggregate both internal and

external data [5]. It enables organizations to meet consumer reporting demands, manage bulk data volumes, to evaluate market advantages, manage risks, efficient controls and, ultimately, enhance organizational performance by turning information into intelligence.

II. EXISTING SYSTEM

In Existing system semantic search engines are being used for searching relationships between two separate objects, using a semantic search knowledge base extracted from network or Wikipedia [4], [7]. However, the semantics in these knowledge bases, such as "is Called," "type" and "sub Class Of," are mainly used to construct a hierarchical knowledge domains and related entities for objects. Such semantic knowledge bases are still not covering relationships existing in Wikipedia, such as "South Africa" is a major "gold" producer. We do not utilize the semantic knowledge bases for measuring relationships in this Concept.

A. DRAWBACKS OF EXISTING SYSTEM:

1. Higher in network traffic
2. Semantic search is used.
3. Low in QOS

III. SYSTEM DESIGN

In our system we focus on Mining Wikipedia Module, Relationship on Wikipedia module, Analysis on link module, Maximum Generalized Flow Based Module.

A. MINING WIKIPEDIA MODULE

Web Page search uses a keyword has grown in this decade, while knowledge search has been introduced to obtain details of a single object and relationships between

other objects [8], such as humans, places and events. Searching knowledge of objects using Wikipedia is one of the top topics in the concept of knowledge search.

In Wikipedia, the knowledge of an object is obtained in a single page updated constantly by many volunteers. Wikipedia also has objects in a number of types, such as people, science, and geography, politic. Therefore, searching using Wikipedia is a better choice for a user to gather knowledge of a single object than typical search engines.

B. RELATIONSHIP ON WIKIPEDIA MODULE

We are using a new method for measuring a relationship on Wikipedia by reflecting all the three concepts: distance, connection, and clustering [10], [11]. Here we are using a new method for measuring the strength of a relationship using the maximum generalized flow. The value of **flow f** is defined as the total amount of f arriving at **destination t**. To obtain the strength of a relationship from object **s** to **object t**, we use the values of a maximum generalized flow from s as the source into t which is the destination; a larger value indicates a stronger relationship. We regard the vertices in the paths forming the maximum generalized flow as the objects constituting the relationship. We can confidently claim that our method can reflect the three representative concepts.

C. ANALYSIS ON LINK MODULE

The two types of relationships between two objects exist in Wikipedia, an external relationship is represented by a link between the two pages for the objects, and an internal relationship is represented by a link structure containing the two pages [12], [14]. User also may need to discover a relationship between two objects. For example, a user may desire to know which countries are strongly based on a field like petroleum, or to know why one country has a stronger relationship with a field like petroleum than other countries. Typical keyword search engines can measure or can show the strength of a relationship. Main problem in measuring relationships starts from the fact that two kinds of relationships: “**internal relationships**” and “**external relationships**”. In Wikipedia, an external relationship is represented by a link. An internal relationship is represented by many links and pages. For example, an internal relationship between petroleum and the USA may be represented by links.

D. MAXIMUM GENERALISED FLOWMODULE

The three concepts, distance, connection, and clustering are very important in measuring relationships; based on cohesion methods underestimate popular objects, even though popular objects might be important for relationships in Wikipedia, Our method can mine elucidatory objects constituting a relationship between the objects by creating paths contributing to the maximum generalized flow [7], [13], [14], that is, paths along which has a large amount of flow in sending data. We propose maximum generalized flow method which shows all the three concepts and does not underestimates the popular objects, in order to measure relationships on Wikipedia.

E. ADVANTAGES OF OUR PROPOSED SYSTEM

1. Improve the knowledge based search technique.
2. Network traffic is reduced.
3. No establishing the relationship between the separate objects in the Wikipedia.

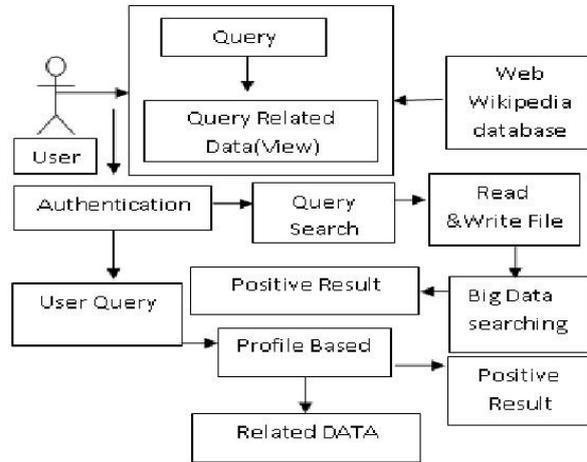


Fig1.1 system architecture

VI. FUZZY LOGIC

This type of logic identifies more than simple true and false values. With fuzzy logic, supposition can be represented with degrees of truth and false [9], [11]. Fuzzy logic has proved that it is particularly useful in expert system and other applications of artificial intelligence. It is also used for some spell checkers to suggest a list of words to replace a misspelled one.

VII. COMPLEXITY

Speed Reduction and the recursive together procedureDo Speed Reduction visits each and every request in the network exactly. For each request visited, Do Speed Reduction is used to find the speed configuration based on the line 2 of Do Speed Reduction which has time cost of given speed configurations.

$$\text{objective function} \leftarrow J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

number of clusters
number of cases
centroid for cluster j

case i
Distance function

Fig1.2 k-means clustering

VIII. CONCLUSION

We have proposed a new method of searching concept in big data for measuring the strength of the relationship between two separate objects on Wikipedia server. By using a maximum generalized flow technique [12], the three representative concepts, distance, connectivity, and clustering, can be inferred in our method.

Our proposed method does not underestimate objects having high degree of relationships [15]. We have included that we can obtain a fairly reasonable priority according to the strength of relationships by our proposed method [6]. Particularly, our method is the only choice for measuring 3- implicit relationships. We have also confirmed that elucidatory objects are helpful to deeply understand a relationship. Some challenges may remain in the future. We are also capable in seeking possibility of the elucidatory objects constituting a relationships mined by our proposed method. We plan to constituents evaluate the elucidatory objects. We are developing a tool for deeply understanding relationships by utilizing elucidatory objects.

REFERENCES

- [1]. E. Anderson, D. Patterson, E. Brewer, "The Magicrouter, an application of fast packet interposing", University of California, Berkeley, May 1996.
- [2]. D. Andresen, T. Yang, V. Holmedahl, O.H. Ibarra, "SWEB: Toward a scalable World Wide Web-server on multicomputer", Proc. of 10th IEEE Int'l. Symp. on Parallel Processing, Honolulu, April 1996.
- [3]. M. M. Chong, A. Abraham, and M. Paprzycki, "Traffic accident analysis using decision trees and neural networks". In IADIS'04, Portugal.
- [4]. M. B. Carterette, J. Allan and R. Sitaraman, "Enhancing the Web's infrastructure: From caching to replication", IEEE Internet Computing, Vol. 1, No. 2, pp. 18 & 27, Mar.-Apr. 1997.
- [5]. P. W. Chu and Z. Ghahramani, "Extensions of Gaussian processes for ranking: Semi supervised and active learning", Proc. of ACM Sigmetrics '98, Madison, WI, pp. 151-160, June 2006.
- [6]. D. Cossock, and T. Zhang "Subset ranking using regression", in Proc 19th annual conference Learning theory, 2006, pp. 605-619.
- [7]. P. Donmez and J. Carbonell, "Active sampling for rank learning via optimizing the area under the ROC curve", in Proc 31th Eur Conf. IR Res Adv Inform Retrieval, 2009, pp. 78-89.
- [8]. V. Cardellini, M. Colajanni, P. S. Yu, "Redirection algorithms for load sharing in distributed Web- server systems", Proc. of 19th IEEE Int'l. Conf. on Distributed Computing Systems (ICDCS'99), Austin, TX., June 1999.
- [9]. V. Fedorov, "Theory of Optimal Experiments", New York, NY, USA Academic 1972.
- [10]. S. Clark, "Traffic prediction using multivariate nonparametric regression", In JTE'03, volume 129..
- [11]. B. Williams, P. Durvasula, and D. Brown, "Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models.", In TRR'98, volume 1644
- [12]. H. M. Zhang, "Recursive prediction of traffic conditions with neural network models," In JTE'00, volume 126.
- [13]. G. Box and G. Jenkins, "Time series analysis: Forecasting and control," San Francisco: Holden-Day, 1970.
- [14]. M. M. Chong, A. Abraham, and M. Paprzycki, "Traffic accident analysis using decision trees and neural networks," In IADIS'04, Portugal.
- [15]. N. Bobroff, A. Kochut, and K. Beaty, "Dynamic Placement of Machines for Managing SLA Violations," Proc. IFIP/IEEE Int'l Symp. Integrated Network Management (IM '07), "TPC-W: Transaction Processing Performance Council," 2007 <http://www.tpc.org/tpcw/2012>,
- [16]. J. S. Chase, D. C. Anderson, P. N. Thakar, A. M. Vahdat, and R. P. Doyle, "Managing Energy and Server Resources in Hosting Centers," Proc. ACM Symp. Operating System Principles (SOSP '01), Oct. 2001.
- [17]. C. Tang, M. Steinder, M. Spreitzer, and G. Pacifici, "A Scalable Application Placement Controller for Enterprise Data Centers," Proc. Int'l World Wide Web Conf. (WWW '07), May 2007.