Subject Features Discovery Method in Semantic Web

Yuan Jia Le[*]

School of mathematics and Computer Engineering, Xi'an University, Xi'an city 710065, P.R. China

Abstract: We have studied Subject Features Discovery methods for discovering resources in Web, but very few handle Web that alters structure. The CMSW (Classification Method for Semantic Web) is the methods that test resources in semantic Web. To make CMSW more result for the Web that are altering their structure, the method DCMSW (CMSW) is proposed which enhances CMSW to renew part structure in little time than running CMSW on the Web. We also enhance CMSW by adjusting settings. DCMSW, worked on network, executes more efficiently than CMSW from one stage to other stage. We have projected a method to optimize process for verifying resources in Semantic Web, which completes efficiently in efficiency and structuring.

Keywords: Classification method, optimization methods, semantic web, subject features.

1. INTRODUCTION

Semantic Web, such as Face-book and Google, has been rapidly developing recently. We can describe such a Web as a diagram, where represents a node as a user and represents an edge their relationship with others. This idea is showed in Fig. (**1**). These relationships can show certain relations as in Face-book, as in Google. Nodes with similar relationship tend to go into weaving resources to form web structures. Moreover, research has discovered three characteristics of semantic network structure. First, we can describe the small world appearance as any two nodes, through only a small number of other two nodes are related to each other nodes. Second, following the pattern of a power function, the power law is the distribution of node degree.
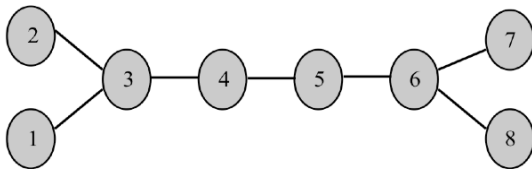

Fig. (1). A diagram with nodes connected to other nodes.

Third is the watched region structure within a network [1]. In the past ten years, the fast growth of semantic Web sites has created a need to analyze its region structures.

In a network, a region is a group of nodes that are connected densely inside and outside sparsely. Subject Features Discovery can show the users information in each region, such as relationships, common interests or opinions. Similarities between persons can be used in advertising within a region, where one person's likes can be suitable for the other persons within the same region; behavioral trends between individuals of a common region.

As Web architectures arise randomly in different shapes and sizes [1], they do not have fixed order or form, so it is difficulties to detect resources accurately. Today, there are many Subject Features Discovery methods in use, changing from brand communication [1] to extent analysis [2]. Experts have designed many methods to discover resources in semantic Webs and do not do well. Today, millions of nodes and billions of edges and are continually changing their architecture include in Webs. The process of combining the region model of a previous stage or snapshot of network architecture, into the discovering of the next stage to improve detecting efficiency of the new region architecture is involved in Subject Features Discovery of dynamic Semantic Web.

And, region architecture's another feature is that some Subject Features Discovery methods discover the most probable region membership in the Web, Also, there are nodes that link two resources, which may be members of the resources they link. We call these linking nodes as center. Some nodes are too distant from other nodes and we should considered abnormal value.

2. RELATED WORK

Ester *et al.* [3] proposed degree-based space gather of applications with disturbance. We proposed a general degree-based gathering method DBCMSW, which forms resources from personal nodes named center points. These center points must meet a number of nearby points user-defined within a given distance. A point with a neighborhood that is very small is named a disturbance point, unless it locates within the nearby points of a center point, which then it is named an edge point. Center points link to other nearby center points to form the center of a region. The ways to group these types of resources is devised by Ester *et al.* [3].

The CMSW method (Classifying Method for Semantic Web) [2], come from DBCMSW, can discover resources, center, and edge points in a Web. A region is developed

from a group of gathered nodes which all meet a given neighbor nodes size. To define the neighbor node of a node, a user-defined threshold is named. Instead of noticing at a node's immediate neighbor nodes, CMSW uses the ε neighbor node of a node and classifies it with those neighbor nodes who share a common set [2]. A structured comparability surveying is used to computer the comparability between two sites. O(n) is time complexity on account of one time that go through of the set of nodes in the web.

In the EOF-GRAPH-IO method [4], an increasing changing process was explained. This method uses variable of DBCMSW, called EOF-GRAPH, to renew the present region architecture of a web from a discovered architecture and its changing cross time. An added characteristic to EOF-GRAPH, compared to CMSW, is that it can find overlay resources, by permitting each node to take over many region marks instead of one. Also, to name a degree-based neighbor nodes of a node, EOF-GRAPH, uses the interval between two web nodes, while CMSN uses comparability of neighbor nodes.

The variation in edges and nodes are showed in the edges node. Accordingly, each web node' variation, whether removed or added, is linked with the previous architecture of the web, and renew to the region architecture is run linked with the edge changing. This method have defined seven possible results that an edge node alter in the web would produce to region architecture.

The study of edge degree in Subject Features Discovery was provided and introduced insight into its accomplishment with other Subject Features Discovery method by [4]. They recommended the degree measured data of edge nodes divided by the possible data of edge nodes, where a region receiving nodes that allow a given edge degree start to maintain satisfactory.

Similar to Subject Features Discovery by node degree [5] is the idea of edge degree. This degree computers the proportion of the number of inter region edges to the number of intra region edges [1].

A common basic method is applied in m Tags Transfer method (TTM) [6] at the beginning of this method, each node has a unique label, or region. Through many repeating process, each node that gained label rom each neighbor can randomly select a label, repeating, until find the convergence or reach a specified number of repeating.

TTM is a Simple and effective method that can only find nearby resources, with no repetition. So a change of the TTM method was put forward that helped to integrate region intersect. TTM (Tag Transfer Method) [7] integrate region with each node to hold more than one Tag to find integrate resources.

Best Choice Method (BCM) is a technology that simulates the process of succession process, combination, selection mutations and to optimizing the consequence to a problem. To use BCM with Subject Features Discovery, the factors represent all the personal nodes and other corresponding area. The factor must adjust a given adaptive functions to find a factor that meets good area structure of a web, such as structuring. In finding this adjustment factor, a inheritance is used consisting of many factors, each showing different re-

gion structure. To execute a user-defined number of repeating in each factor variants into one between two or more factor, and produce the highest due to the fitness value. At the end of the repeating, the highest value factor of fitness function is as the last of the regional structure. More use of optimization method can be read in the network [8].

3. SUBJECT FEATURES DISCOVERY

Definition on the relationship of nodes in a degree-based structure in their formulation of the DBCMSW method is proposed by Ester *et al.* [3]. Ester *et al.* defined Degree-based Subject Features Discovery, where each node in a region shares a number of connections with other node than outside node. Let G = {V, E} be the web diagram, where E is the set of edges and V is the set of nodes. Also let N(v), and let E(v) be the set of all edges incident to v where v ∈ V, be the set of neighbors of node v, including v. There are several definitions to describe the structure of resources in a degree-based web. Fig. (**2**) shows these points.
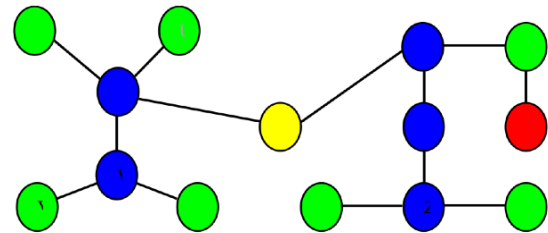

Fig. (2). Definitions of Degree Architecture.

3.1. Neighbor of Nodes

The node's nearby node is not enough to define a linked region, because a linked region must have best comparability with neighbors in its own region. As to deal with those more closely linked nodes, the think of a ε-node was showed. In EOF-GRAPH, Ronhovde, R.K. [5] use a space function to link the tightness between two sites. Their ε-node integrated all nodes that satisfy a defined critical value ε. Given two values u and v:

$$\varepsilon - Nei(v) = \{u \in N(v) \mid dist(v,u) \leq \varepsilon\}$$

The more nearby nodes they have the same point, the bigger their comparability will be; So [4] uses a slightly altering of the ε-nearby nodes defining: in the range of (0, 1).

$$\varepsilon - Nei(v) = \{u \in N(v) \mid sim(v,u) \geq \varepsilon\}$$

CMSW gives a new method when linking nodes tightness. They give a comparability function that discovers a proportion between the numbers of nearby nodes that they both share to the number of nearby nodes that each node has:

$$sim(v,u) = \frac{|N(v) \cap N(u)|}{\sqrt{|N(v)||N(u)|}}$$

3.2. Center Nodes

We named a node a center node that concludes ε-nearby nodes that has a scale that meets a user-named critical value of μ sites: $|\varepsilon - Nei(v)| \geq \mu$

On the basis of [5], aμis introduced for analyzing.

3.3. Structure Reach-Ability

If u is in the neighbor of v and v is a center node, name node u is attainable from node v, as follow:

$$u \in \varepsilon - Nei(v) \wedge |\varepsilon - Nei(v)| \geq \mu$$

If there is a link of nodes that are all attainable from the prior until u is reached, then name a node u is attainable from node v. Showed by the defining of directly attainable:

$$\{v_1, \cdots, v_n\} \in V; \quad v_1 = v \wedge v_i = u; \quad \forall i \in \{2, \cdots, n\}$$ Directly Reachable (v_{i-1}, v_i).

This is only meeting the conditions if the links of nodes are center nodes.

3.5. Structural Connectivity

The two nodes are edge nodes, and the node that links them is a center: $\exists x \in V$; Architecture attainable $(x, v) \wedge$ Architecture attainable (x, u). If there is a node x from which v and u are structurally attainable, two nodes (v, u) are linked.

3.6. Center and Edge nodes

Through CMSW, if a node v does not be part of any region, it is a center node, but it does have more than two nodes be part of various resources: $\exists(x, y) \in N(v)$; $x \neq y \wedge Region(x) \neq Region(y)$.

If a node v does not belong to any region, it is an edge node and it does not include two nearby node that be part of various resources: $\sim \exists(x, y) \in N(v)$; $x \neq y \wedge$ Region$(x) \neq$Region(y).

4. REGION STRUCTURE INFORMATION

In this part, we will recommend the seven schemes that may arise and the course of replacing an active web.

Border comparability to computer according to an border altering of nodes v and u are $\{e \in E\ e \in E(v)$ or $e \in E(u)\}$.

A altering in the architecture of a web, such as a border being removed or added, would influence the comparability between the two nodes to this border. Instead of examining the comparability of all borders in the web, only those borders that are linked to these two nodes need to be changed.

4.2. Result of a Web Alter

There is the think of an active altering, where a altering will produce a new region, a region will accept a new element, or two resources combine. Seven possible outcomes are recommended by [4], so that a altering in a web would generate. Each of these seven schemes is divided into two sets by how that alters the region construction. Then exists an inactive alter, which can change a region, change an element from a region, or divide a region into two part. These two sets with four schemes within are the basic thinks for changing an active web. A production of a new region would appear when webs change into center and there are no ε-nodes that were centers before; the new center webs can set a new region with its ε-node. For an active altering, a production of a region, Except for a region or a combine of two resources may appear.

Except for a region will appear when a web, which had not been part of a region before, gets a comparability with a center that is bigger than ε, thus it be part of this center region now. So that the sites of these sites can become one, when a center of one region gets comparability with a center of other region that is bigger than ε. the combining of two sites can happen.

A region is divided into several parts, shifting from a region, or shifting of a region may appear in an inactive altering. A separation will occur and it produces an interval between the other centers of the region when a web, that was a center before, now is not a web. This interval will produce a separation in more links of centers leading each set of centers to constitute independent webs. A shifting from a region appears, or adequate comparability, to a center of its region and is not marked as an element of that region when a web not has a strong relation. a shifting of a region takes place, When all centers of a region are decreased from their center state and the region is shifted.

Any alter in a web will include a changing of a side. To decide if an altering will generate active or inactive alter to a web is not an effortless thing. Then according to two rules, [4] mentions two sets of schemes with side altering as: a new side generates an active scheme, and the change of a side generates an inactive scheme. This think may perform for the interval-based method, on the basis of our estimation on a comparability-based method. The next part explains our method, which deals with the side altering of a web. But this is not always such the situation.

5. DCMSW: DEVELOPING TO CMSW

Our altering to CMSW makes it to develop webs without the predefined starting point of field and can update a changing web. Our research is on account of the method CMSW [2] and uses reinforce to its degree for testing. We put forward the method DCMSW (Dynamic CMSW), which can deal with changing Web.

CMSW's start for ε, in the scope of (0, 0.95), names the smallest comparability between the neighbor nodes that can be the two webs to be ε nearby node. Depending on user explaining for ε can reduce running if a false ε is used, and carrying out multitask with various methods is inefficient. This ε-node is that names the region architecture, we have found that better performing for any web can be in the scope of (0.35, 0.7) in our study with measuring all kinds of Webs.

A big value ε will produce many small webs, while a ε value too small will generate several big webs. In our research, to computer the ε-sites of a web, we carry out a check ε the scope (0.38, 0.82) and get a mean value of the outcomes. We have found a better number of μ to be 2, which is the number that we employed. Way 1 explains the changing replacing of a web with DCMSW. This way generates measurable outcomes and relieves the demand to execute multi-times to examine ε numbers.

The web is renewed from the webs of another side altering. With a set of phases of a web, carry out CMSW on the

first stage. Then for all consistent stages, get the various types in sides among the two sides. This updating on the web deals with a site that can turn into a center. When a web is not part of the set, one final altering to the web is finished. An available set id or a new set id is transmitted by all linked webs to constitute a new region when an altering in the web is tested and can to be renewed.

6. OPTIMIZATION METHODS USED TO CHANGING WEB

A predefined a certain amount of repeating is executed on the species group which changes each factor into an intersection between more factors that generate the maximum values on account of the adaptive relationship. In discovering this optimized factor, a species group is applied composing of many factors, every factor expressing various region structures. Except for CMSW, we have studied the idea of optimization methods and their application with changing Web. Best Choice Method (BCM) is a skill that simulates the course of transmissibility, changing, selecting, and intersect to improve the result to a question. By using BCM with Subject Features Discovery, the factors express all personal webs and their region. To discover a factor that meets better region construction of a web, the factor must optimize an offered adaptive relationship, for example, structuring. The factor with the optimized adaptive relationship value it applied as the end region structure at the repeating. Contents on the application of optimization methods in Web showed at [8].

If only a little altering to the web has happened, for the use of developing methods in changing Web, then the former region architecture will generate a better value for the adaptive relationship. Developing methods go with a serial architecture for each factor of the species group, and slowly finds the better factors to become into proper ones. When going with the beginning changing species group, the think generates requiring several repeating over the species group to discover a better architecture than is met. This data permits for a good end of optimization methods, when a positive time is curtained in the repeating that no more factors could be found instead of finishing in the process of an order of good factors being found. To utilize this think of several repeating, we recommend a predefined repeating number, not for the numbers of species group changing, but the consistent repeating number where no good factor is discovered.

We present the method DBCM (Dynamic Best Choice Method) for the developing methods with changing Web. DBCM uses the formerly discovered "best" factor of the former stage as a factor in the species group of the later. If this factor occurs to still be the better factor for the present period, then the developing method will not change the species group amount of times. This produces in the next stage beginning with a factor in the species group with a better adaptive.

7. CHANGING WEB SETS

We conduct on real circumstances Web of changing scale in computing the executing of our methods.

HTTP-TH practical web. The web's information contains a stage from February 1995 to June 2006 and the stage for each time on when it was increased to the Intelligence Agent [9]. This data made it able to produce periods for the web's altering later. We transformed this data into 640 periods by what continuous work days the diagram had altered. This high level practical web concludes up to 28,680 nodes and 375,206 sides. Each web matches to a specific content in a set, and each side express a content from one set to another.

ITM Semantic Improvement SMTP web. This web explains the shift terminal applying between pupils in a region. Shift terminal applying includes interviewing or SMTP (Simple Mail Transfer Protocol) data between pupils in the same region. Our phases for this web are period that demonstrate new link between two persons. This is a quite petty diagram made of up to 70 nodes and 280 sides. This developing web represents the connections that set up from June 2006 until July 2010.

ITM region web. We produced phases in the same method as the SMTP network, at present that illustrates an altering in exchanging. This information was produced to research the making policy course of persons in answer to web exchanging. We concentrated on the SMTP web of this information describing the exchanging between persons, resemble to the Call-SMTP web. The web forms of up to 70 nodes and 160 sides.

8. RESULTS

Two methods have been presented to deal with Subject Features Discovery in Semantic Web, DCMSW and DBCM. Both methods test resources more efficient than their beginning Subject Features Discovery hardly to reduce in comprehensive structuring, which we could estimate the effect of the Subject Features Discovery of our methods. DBCM illustrates a certain degree improvement in efficiency and structuring on BCM.

8.1. DCMSW on Semantic Web

Our method DCMSW executes Subject Features Discovery on changing Web effective than CMSW, with few to no executing reduce. DCMSW can change a web in a more high level way than CMSW can on the entire web. This is finished by re-computing the new region architecture only in the fields of the web that demonstrated an altering in architecture. Our carrying out mean over the scope of 0.3 to 0.7 for a better region value generates consequence very resemblance to the best region value selected from the identical scope on CMSW.

The altering causes DCMSW to update the whole web and not be limited to a particular region. It is obvious that DCMSW can execute renewing effective than CMSW finishes the whole web on some stages. On mean, using a computer with a 2.8 GHz Inter CPU with 16GB of storage, CMSW computed all stages in 84 minutes and DCMSW carried out in only 52 minutes. This shows that DCMSW is good executed on little alters to improve runtime. Structuring is the most common way in calculating the efficiency of resources in Subject Features Discovery methods. The other two webs were too little to see a clearly improving in effi-

ciency in the two methods. Structuring is the criterion for computing the advantage of the region architecture of a web, but structuring executes better when all webs are composed into their perfect region. Structuring [10, 11] is confirmed as:

$$Q = \frac{1}{2M} \sum_{u,v} (A_{uv} - \frac{k_u k_v}{2M}) s_{uv}$$

where A is the nearby array of the web, k_i is the level of site i, and s_{ij} is 1 if sites i and j have the identical region element, and 0 or else.

CMSW, and DCMSW, not all webs are elements of a region. Center and outside web do not be part of any region, which reduces the structuring. With this idea, it is hard to contrast CMSW and DCMSW with other methods that do not include the think of center and outside web.

The DCMSW conducts Subject Features Discovery on changing Web almost equally to what CMSW would generate by conducting the whole web. This changing region architecture is a main content when handling with Web that develops to a large-scale. Contrast CMSW and DCMSW by ways with structuring will still be a good way. Our effort of CMSW using a small data conducts the better data from CMSW. There is unimportant changing between the stages on both Webs, which are good contrast with running CMSW to find which data is the better to use.

8.2. DBCM on Semantic Web

To make a comparison from DBCM to normal Best Choice Method (BCM), we conducted DBCM as standard over all stages and ran BCM on each phase respective. DBCM generates a steady performance time of optimization methods by every stage, while decreasing the frequency of repeating required to find a better region element.

BCM does not all the way generate a better region structure every time, except for the frequency of repeating was improved more, improving the performance period. DBCM reduces the repeating number of the Call-SMTP web by about 45% and the SMTP web by about 40%. Figs. (**3** and **4**) show the structuring of webs in the end stage contrasted between BCM and DBCM on the SMTP Web. As showed from the diagrams, DBCM generated steady consequences by structuring for each running. With the operation consistency and the repeating reduce, it is obvious that the technology of DBCM enhance optimization methods enormously for changing Web.
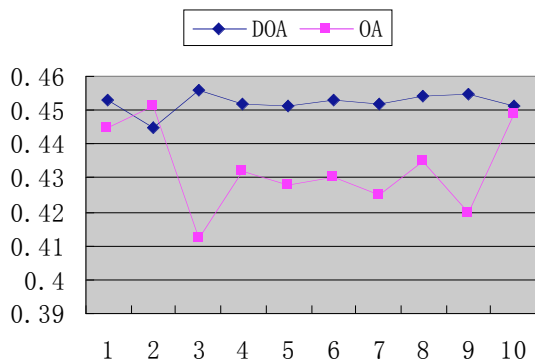

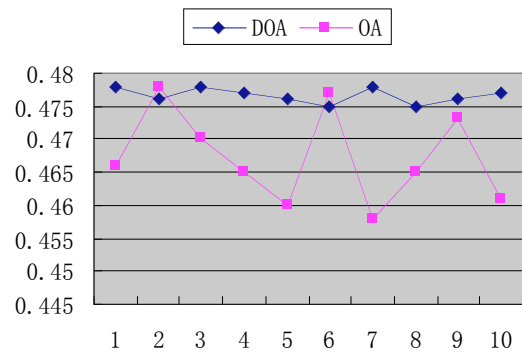**Fig. (3).** Structuring of SMTP web of ten runs of OA and DOA

Fig. (4). Structuring of Call-SMTP web of ten runs of OA and DOA.

CONCLUSION

We have expressed that DCMSW and DBCM are effective in process mode, if not good, in structuring to their Subject Features Discovery contents. This efficiency improve effective in structuring is perfect when big Semantic Webs are included, which are continually altering their architecture. Because of we have expressed with our tests on the web, the amount of alters from one phase to another is an important factor in the performance period of a DCMSW. With this, it is possible to conduct DCMSW after little alters have appeared in the web, maintaining the region architecture at present frequently.

Skills provided in DCMSW and DBCM, may be applied in Subject Features Discovery methods for transforming them to changing Subject Features Discovery methods. With the frequent improvement in the scale of semantic Web, it is even impossible to use static Subject Features Discovery ways to evaluate these Webs. Subject Features Discovery must to turn to the direction of dynamic network Subject Features Discovery to fit with current Web. This is a significant field of approaching study in the area of Subject Features Discovery.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   S. Fortunato, "Community detection in graphs," *Physics Reports 486*, 2010. Available at: http://arxiv.org/abs/0906.0612

[2]   R. Hu, "Channel Access Controlling in Wireless integrated information Network using Smart Grid System," *Applied Mathematics & Information Sciences*, vol. 6, pp. 813-820, 2012.

[3]   M. Ester, H.P. Kriegel, J. Sander, and X. Xu, "A Degree-Based Method for Discovering Resources in Large Spatial Databases with Noise," In: *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996.

[4]   T. Falkowski, "Region Analysis in Semantic Web. PhD. Otto-von-Guericke-University," *Magdeburg*, 2009.

[5]   R. K. Ronhovde, R. Peter, and Z. Nussinov, "An Edge Degree Definition of Overlapping and Weighted Graph Resources," Available at: http://arxiv.org/abs/1301.3120, 2013.

[6]   R. Hu, "Stability Analysis of Wireless integrated information Network Service via Data Stream Methods," *Applied Mathematics & Information Sciences*, vol. 6, pp. 793-798, 2012.

[7]   J. Xie, B.K. Szymanski, and X. Liu, "STTM: Uncovering Overlapping Resources in Semantic Web via a Speaker-Listener Interaction Dynamic Process," In: *Proc. ICDM Workshop*, pp. 344-349, 2011.

[8]   C. Pizzuti, "BCM-Net: An Optimization method for Subject Features Discovery in Semantic Web. PPSN," Volume 5199 of Lecture Notes in Computer Science, Springer, pp. 1081-1090, 2008.

[9]   R. Hu, and J. H. Guo, "New Network Access Control Method Using Intelligence Agent Technology," *Applied Mathematics & Information Sciences*, vol. 7, pp. 43-48, 2013.

[10]   W. Dong, B. Lepri, and A.S. Pentland, "Modeling the Co-Improvement of Behaviors and Semantic Relationships Using Mobile Phone Data," *Media*, pp. 134-143, 2011.

[11]   M.E.J. Newman, "Structuring and Region Structure in Web.," In: *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 23, pp. 8577-8858. http://dx.doi.org/10.1073/pnas.0601602103, 2006.