Expert knowledge similarity measurement using network graph edit distance

Qingyun	Liao	419598668@qq.com	China	Beijing Institute of Technology
Qi	Lv	performerlv@163.com	China	Beijing Institute of Technology
Ying	Wang	wangying318@gmail.com	China	Beijing Institute of Technology
Xuefeng	Wang	wxf5122@bit.edu.cn	China	Beijing Institute of Technology
Dong	Wan	wandong1018@foxmail.com	China	Beijing Institute of Technology

The 21st century is the era of knowledge economy. For those who need expertise service, finding the right people to do the right things has always been a problem. For those expertise service provider, how to locate competitors who owned similar expertise or partners who acquired complementary knowledge accurately and quickly is of the same importance.

In that case, we proposed 2 research questions: (1) Who is the right expert? (2)How to find competitors and partners?

The one that acquired knowledge just matches problems is the right expert. Different problem requires different expertise to solve. Working in same domain, competitors have similar knowledge structure and they are facing almost the same problems. Thus ,expert knowledge similarity help to find competitors in a domain. Furthermore ， if one wants to find a partner who has complementary expertise, expert knowledge similarity measurement makes it much more efficiently.

In a word, the study of expert knowledge similarity measurement is meaningful for expert finding, expert identification and competitors or partners locating.

Expert knowledge similarity measurement is a relative new domain and related work was few. Usually, scientists measured similarity of scientific and technical literature to reflect expert's similarity, including attribute based way and semantic based way [1，2]. Qian Pan analyses the relationship between textual attributes and the attribute barycenter coordinate model, and establish text attribute barycenter coordinate model to compute the similarity between texts and queries [1]. Some measured expert similarity based on social network. But McDonald pointed out that the SNA-based way may could not satisfy concrete needs of customers [3].What's more, Boeva comes up with a method for comparing expert profiles within the context of expert networks by measuring expertise similarity between experts [4]. Xuan Huang established a new method of generating expert's scientific labels using topic model [5].

This article focus on expert knowledge measurement, using edit distance method on knowledge-network-based graph. It is an exploration for expert knowledge similarity measurement. Also, it is a relatively new and creative method to apply graph edit distance on expert knowledge similarity measurement.

After extracting expertise features "author's keywords" for experts from the text of papers published by them, we generate a author-keywords matrix． Then the method constructs expert knowledge network maps to represent expert knowledge, containing signification and knowledge structure.

The framework of this study is showed in Figure 1.

Academic papers are rigorous and contain lots of information. These standard documents are outcomes of experts' major studies. The field “author’s keywords” was composed of words or phrases. It summarizes the theme of paper and was though concise. We choose expert’s published paper in web of science and field "author's keywords" to do the following work.

Via Vantage-Point, we generate “keywords” to “keywords” matrix and an “authors” to “keywords” matrix. Then we import “keywords” to “keywords matrix in Gephi to draw an all-domain knowledge base map. In the same way, we get expert’s knowledge maps for each author. Furthermore, we use graph edit distance between expert knowledge network maps to measure expert knowledge similarity. What’s more, combing all-domain knowledge base map with expert knowledge network map, we can draw an expert knowledge network map for a domain.

In order to measure the similarity of expert knowledge, we creatively propose graph edit distance to calculate the value of similarity between expert knowledge network map. In order to compute the graph edit distance dλmin (g1, g2) often A*-based search techniques using some heuristics are employed. A* is a best-first search algorithm which is complete and admissible, i.e. it always finds a solution if there is one and it never overestimates the cost of reaching the goal[14].

The key and difficult point was how to define cost function. We take node-similarity and node-weight into account. Firstly, this article measured similarity among 24257 keywords, based on keywords co-occurrence matrix. Then a list recorded the standard similarity of every two keywords was acquired. In that case, the cost of node substitution for every two node was clear.

Next, for every expert knowledge network map, different keywords (nodes) were given different weights according to their frequency in this expert’s records. So in expert knowledge network map, every nodes also had weights. So the real cost should multiply node-weight.

The following work was recode A*-based search methods to apply on expert knowledge network map.

We chose “big-data” domain to do case study. Import data in Vantage-point, after data cleaning, we get 24257 keywords. Figure 2 shows an example of expert knowledge network map.

Finally ,we get an excel of experts knowledge similarity matrix in big data domain. For every expert, we can find several similar experts who share knowledge. Also, here we get an big data domain expert knowledge distribution map. It shows critical and basic knowledge in big-data domain, such as “parallel computing”. And graph edit distance method is proved an efficient way on expert knowledge similarity measuring

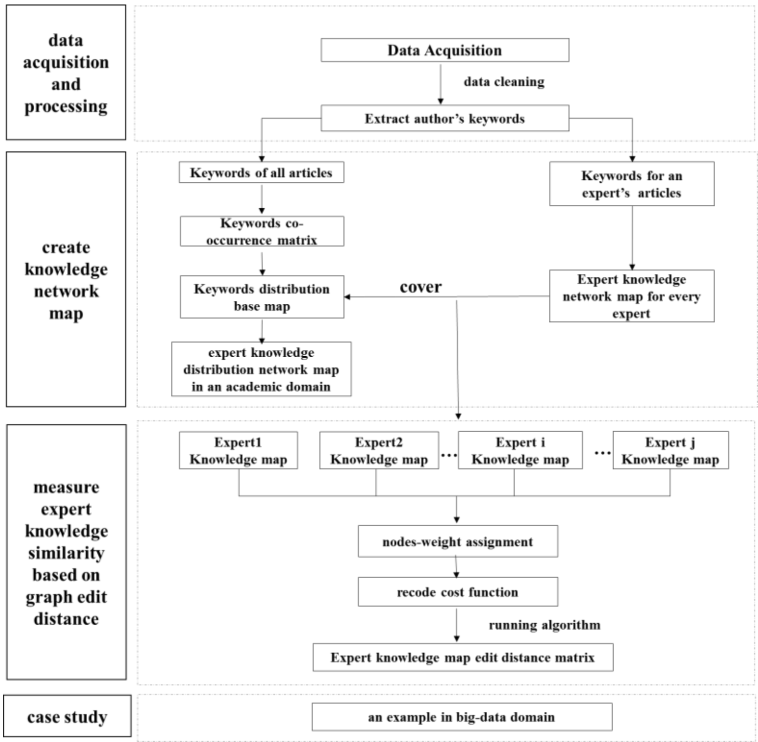

Figure 1 Expert knowledge measurement study frame diagram

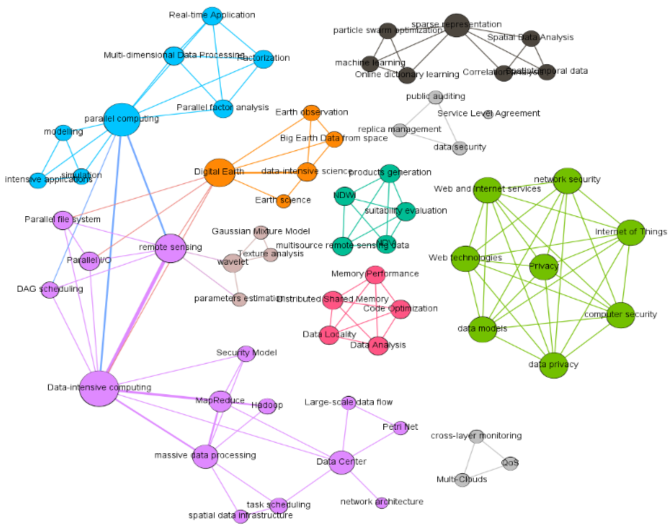

Figure 2 An example for expert knowledge map

Reference

[1]PAN Qian-Hong, WANG Ju, SHI Zhong-zhi. Text similarity computing based on attribute theory[J]. Chinese Journal of Computers ,1999, 22(6):651-655.

[2] JIN Bo ,SHI Yan-jun ,TENG Hong-fei,Similarity algorithm of text based on semantic understanding[J]. Journal of Dalian University of Technology, 2005, 45(2):291-297.

[3] Mcdonald D W. Recommending collaboration with social networks: a comparative evaluation[C]. Sigchi Conference on Human Factors in Computing Systems. ACM, 2003:593-600.

[4]Boeva V, Krusheva M, Tsiporkova E. Measuring expertise similarity in expert

networks[C].Intelligent Systems. IEEE, 2012:053-057.
[5]xuanhuang. Scientific data acquiring and experts similarity researching based on open access[D].Zhejiang University, 2016.
[6] V. Levenshtein, Binary codes capable of correcting deletions, insertions and reversals. Sov. Phys. Dokl. 10(8), 707–710 (1966).
[7] R.A. Wagner, M.J. Fischer, The string-to-string correction problem. J. Assoc. Comput. Mach. 21(1), 168–173 (1974).
[8] S.M. Selkow, The tree-to-tree editing problem. Inf. Process. Lett. 6(6), 184–186 (1977).
[9] M.A. Eshera, K.S. Fu, A graph distance measure for image analysis. IEEE Trans. Syst. Man Cybern. (Part B) 14(3), 398–408 (1984).
[10]W.H. Tsai, K.S. Fu, Error-correcting isomorphism of attributed relational graphs for pattern analysis. IEEE Trans. Syst. Man Cybern. (Part B) 9(12), 757–768 (1979).
[11]W.H. Tsai, K.S. Fu, Subgraph error-correcting isomorphisms for syntactic pattern recognition. IEEE Trans. Syst. Man Cybern. (Part B) 13, 48–61 (1983).
[12]H. Bunke, G. Allermann, Inexact graph matching for structural pattern recognition. Pattern Recognit. Lett. 1, 245–253 (1983).
[13]A. Sanfeliu, K.S. Fu, A distance measure between attributed relational graphs for pattern recognition. IEEE Trans. Syst. Man Cybern. (Part B) 13(3), 353–363 (1983).
[14] Riesen K. Structural Pattern Recognition with Graph Edit Distance[M].Structural Pattern Recognition with Graph Edit Distance: Approximation Algorithms and Applications. Springer Publishing Company, Incorporated, 2016.