Research on Information Fusion for Multi-sensor Expert Features

Ye Guanghui & Li Gang

Information Management School of Wuhan University

Nov.26th.2014

Introduction

In view of most suggestive several techniques for Internet's innovation, I believe expert feature recognition could be carried out based on three kinds of intelligence.

（1）The first is knowledge intelligence, which mainly takes advantage of knowledge resources in various academic database ,which reflects experts ' research areas.

（2）The second is web intelligence, which related with various Web resources, such as expert affiliations web site, Baidu encyclopedia, Wikipedia, and so on, with the help of which we may obtain expert profiles using text analysis methods.

Introduction

In view of most suggestive several techniques for Internet's innovation, I believe expert feature recognition could be carried out based on three kinds of intelligence.

（3）The last is social network intelligence through scientific research groups, social network platforms (such as Sina weibo), academic forums (such as small insects), and technology communities (CSDN), we can get users' tag,post,comment and so on, which reveals the concerns and interests of experts.

Literature Review-1

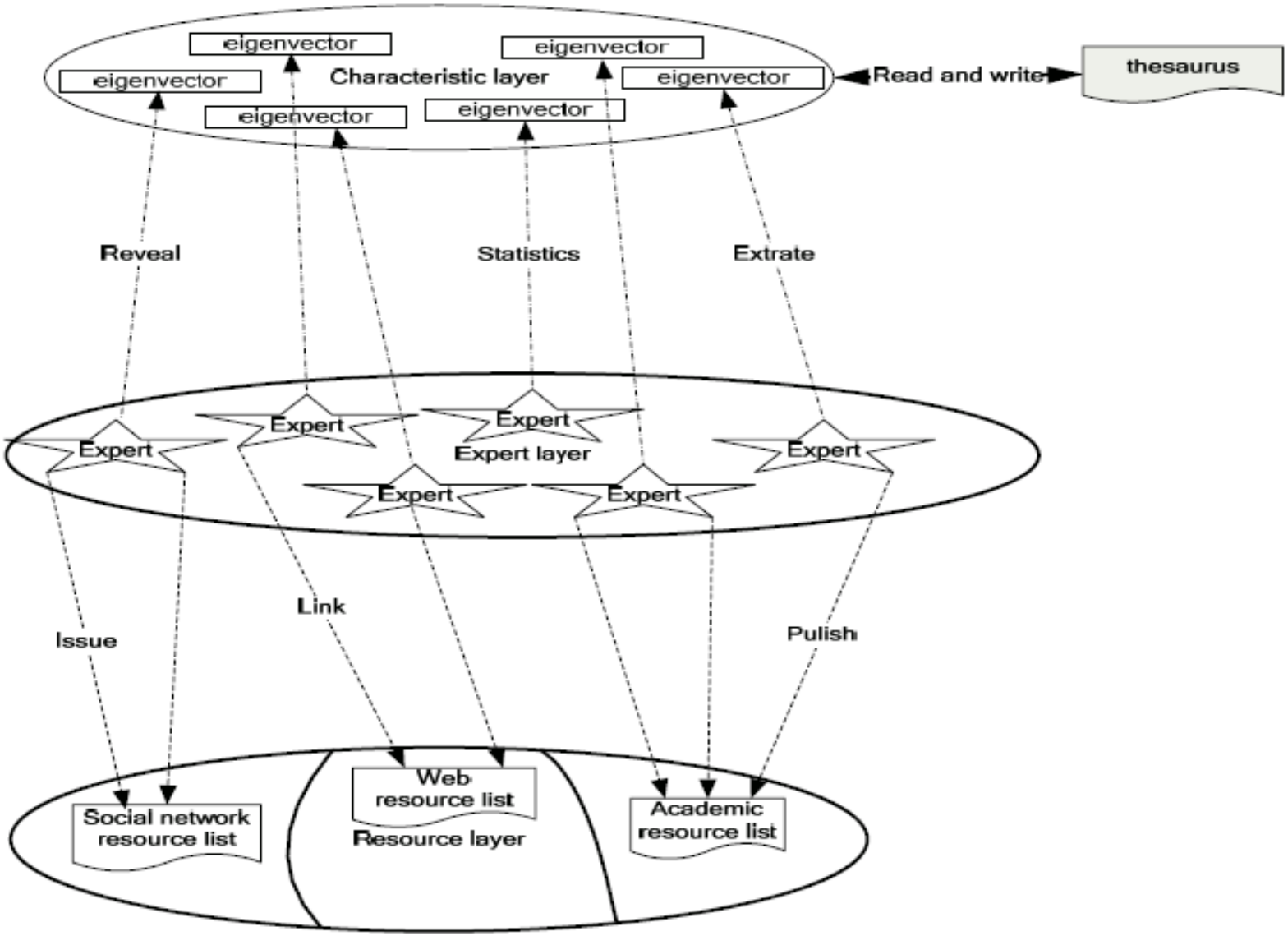

Fig 1.The abstract process of expert feature recognition

Literature Review-2

Focusing on Fig 1, relevant researchers have done some work, which can be summarized into four parts:

(1) Recognition method based on knowledge resources

(2) Recognition method based on Web resources

(3) Recognition method based on social network resources

(4) Integrated recognition method of multiple resources

Methodology-1

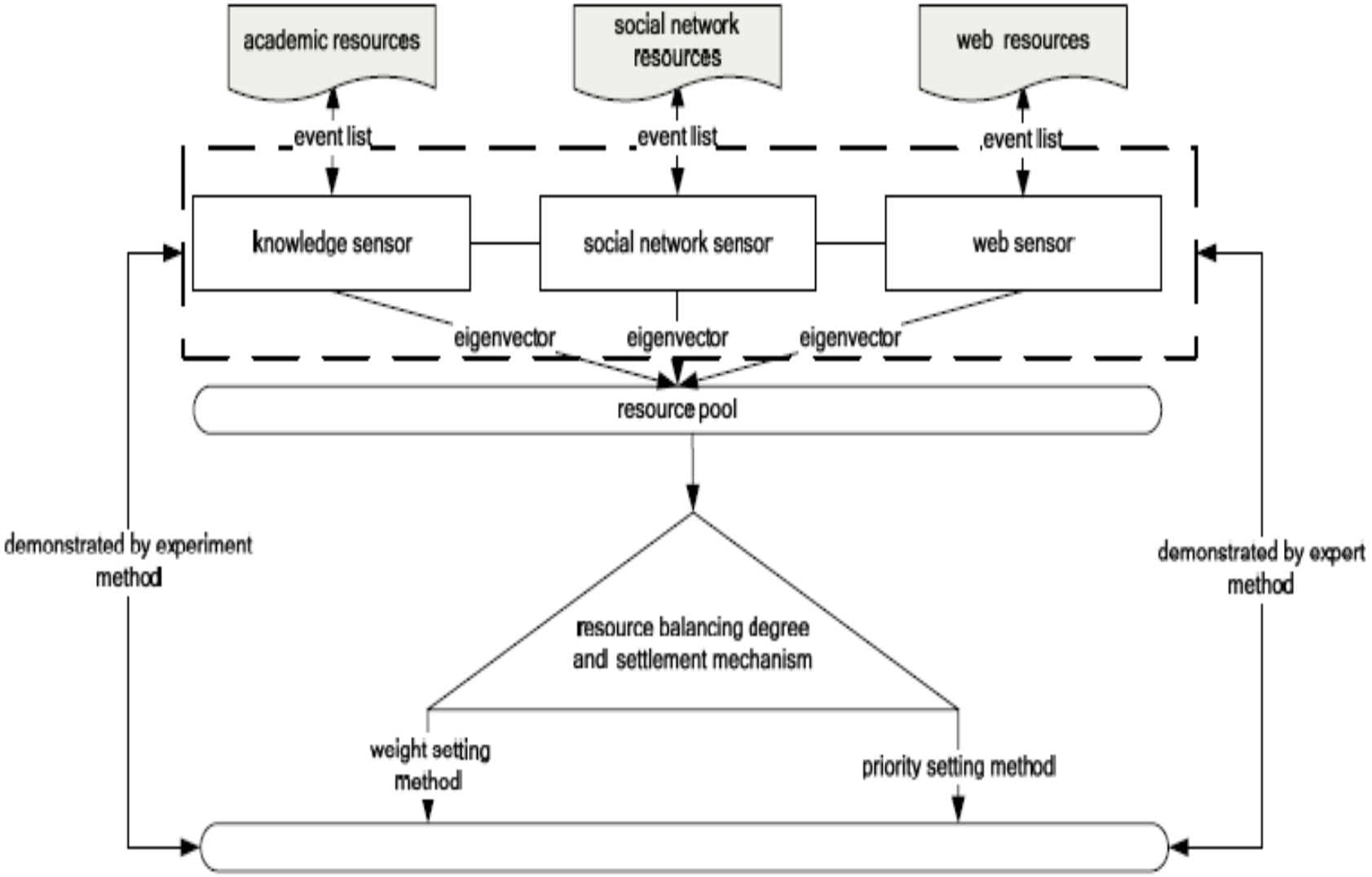

Fig 2. The research framework of expert feather recognition

Methodology -2

Table.1.The event list of knowledge sensor

sensor	event list	
knowledge sensor	attribution recognition of expert	expert name recognition
		expert affiliation recognition
	attribution recognition of academic resources	resource name recognition
		author recognition
		publishing date recognition
		journal name recognition
		keywords recognition or extraction
	relationship recognition between expert and academic resources	co-authors
		translation
		sole author

Methodology -3

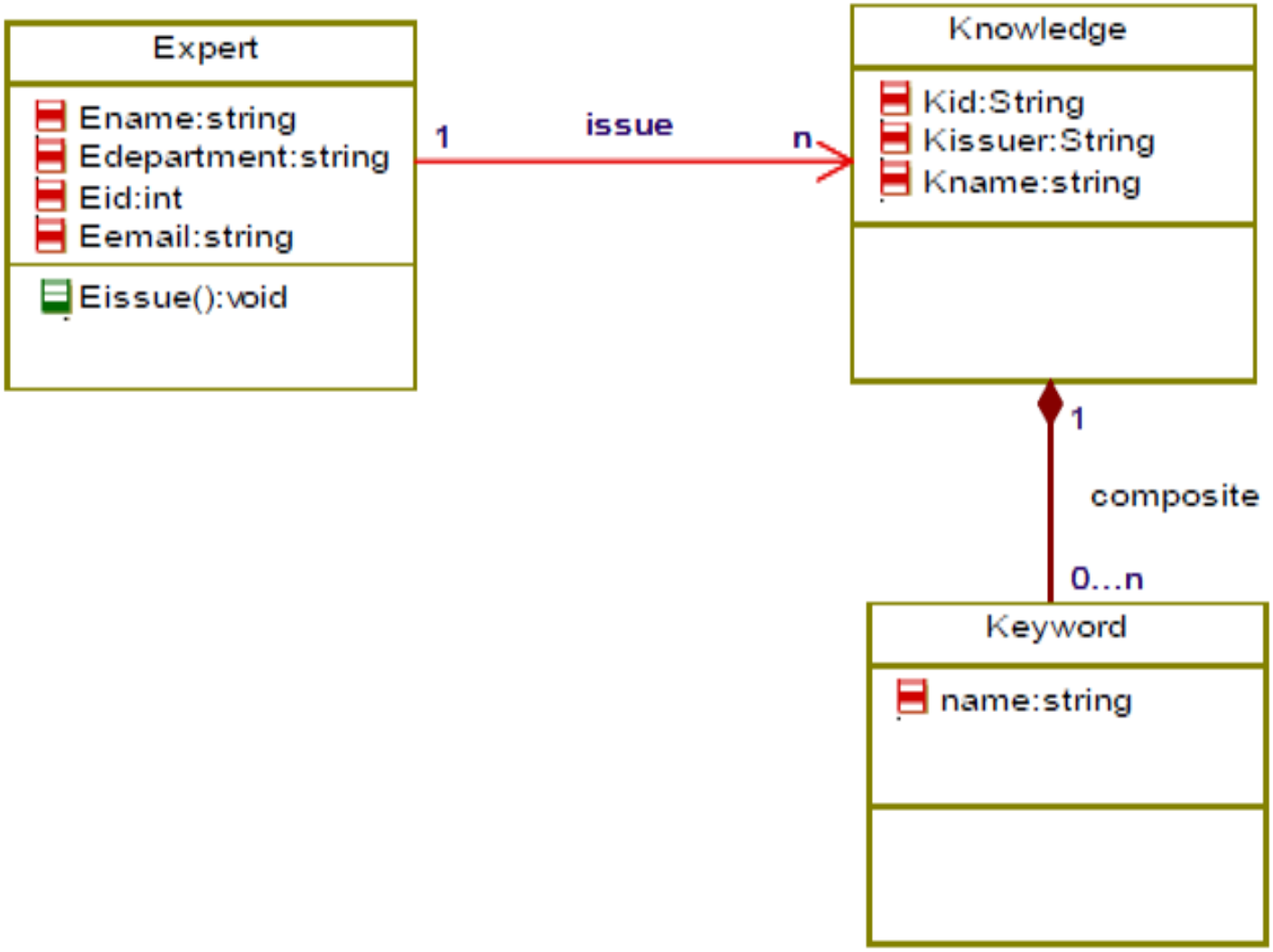

Fig 3. Experts-academic resource structure

Methodology -4

Three entities in Fig.3 have gone through two multiplicity passes, so knowledge expert feature vectors is equal to accumulation of keywords vector in his(her) all academic resources, showed in eq.(1), which takes time span and expert research direction evolution into account.

$$V_e = \sum_{1}^{n} (V_k) \qquad (1)$$

Methodology -5

Table 2.The event list of Web sensor

sensor	event list		
web sensor	recognition of external features	document number	
		document length	
		object described recognition	
	recognition of content features	information recognition based on models	model recognition
			keywords extraction
		information recognition based on rules	rules recognition
			keywords extraction
		information recognition based on statistics	term frequency statistics
			inverse document frequency statistics

Methodology -6

Table 3. The event list of social network sensor

sensor	event list			
social network sensor	attribution recognition of node	node name recognition		
		node size recognition		
	attribution recognition of relationship	associated object recognition		
		relationship category recognition	relationship between expert and resource	
			relationship between expert and entity	
		tag recognition	resource tag	
			entity tag	
		relationship strength recognition		

Methodology-7

As there exists noise or conflict among three expert eigenvectors obtained by single baseline method, so we have designed multi-sensor information fusion algorithm, in which “resource balancing degree” is the key factor.

Resource balancing degree is to describe the distribution of the resource acquired by three kinds of sensors. If r1,r2, r3 respectively represents amount of resources knowledge sensor, Web sensor, social network sensor have gotten.p1=r1/ (r1+r2+r3) represents knowledge resources proportion, so are p2 and p3, Resource balancing degree can be calculated with eq. (2)

$$\sigma=\sqrt{(p_1-p_2)^2+(p_1-p_3)^2+(p_2-p_3)^2} \qquad (2)$$

Methodology-8

++++++++++++++++++++++++++

1 Input: $v_1, v_2, v_3, p_1, p_2, p_3, \sigma, \sigma_t$

2 Output

3 Process Read v_1, v_2, v_3, σ

4 If $\sigma \in [0, \sigma_t]$ Then $v=p_1*v_1+p_2*v_2+p_3*v_3$

5 Else if $\sigma \in (\sigma_t, \sqrt{2}]$ and $p_1=Max(p_1, p_2, p_3)$ then $v=v_1$

6 Else if $\sigma \in (\sigma_t, \sqrt{2}]$ and $p_2=Max(p_1, p_2, p_3)$ then $v=v_2$

7 Else $v=v_3$

8 End if

++++++++++++++++++++++++++

Empirical analyses-1

(1) Analysis object

We select CNKI as knowledge base, top 11 journals in computer categories as knowledge resources. Based on search results, we randomly select 10 computer experts from top 20 experts in high yield.

(2) The control group

We choose expert research interests provided in the site of C-DBLP (http://www.cdblp.cn) as the control group.

Empirical analyses-2

(3) Test procedure

①Using knowledge sensor to obtain expert eigenvector v1 and amount of resource r1

②Using social network Sensor to obtain expert eigenvector v2 and amount of resource r2.We choose ①"knowledge resources" as social network nodes in the paper, that is to say social network is citation network. In order to simplify calculations, this article only analyzes the citation of highly cited papers and publishing source from 11 kinds of periodicals mentioned . We obtain highly cited papers with the help of Ziff's second law, the formation of which is showed in eq. (3)

$$r_i' = \mathrm{Round}(\mathrm{SQRT}(r_i)) \quad (i=1,2\cdots,10) \quad (3)$$

Empirical analyses-3

(3) Test procedure

③Using Web sensor to obtain expert eigenvector v3 and amount of resource r3.

In view of web resources with large quantities and uneven quality, so we set our qualified experts from high credibility Web sites, including expert affiliations Web site, http://www.sciencenet.cn (Web of science),http://baike.baidu.com (Baidu encyclopedia) and http://zh.Wikipedia.org (Wikipedia), as shown in Table 4.

Table 4. Statistical analysis result of Web resources

index \ web site	affiliations web site of expert	web of science	Baidu encyclopedia	Wikipedia
total of resource	8	6	2503	39
total of effective resources	8	2	9	3
precision of resource	100%	33%	0.4%	8%
vacant percentage of resources	20%	80%	10%	70%

Empirical analyses-4

（4）Set σ_t =1.0, follow multi-sensor information fusion algorithms to generate comprehensive expert eigenvector v, as shown in Table 5.

Table 5. Result of multisource information fusion

expert name	r_1	r_2	r_3	v
ZWM	83	30	3	parallel processing（3.58）message passing（3.58）cloud computing（3.36）interconnection network（2.28）cluster of workstations（2.28）
YJY	141	51	3	face recognition（21.69）feature extraction*（15.18）pattern recognition（11.57）feature extraction（5.78）image segmentation（4.34）
XXF	80	36	3	genetic algorithm（4.71）cloud manufacturing（3.33）service engineering（2.69）relational database（2.69）virtual enterprise（2.69）
XL	81	40	2	Wireless sensor networks（3.90）distributed system（2.63）event sequencing constraints（1.98）operating system（1.98）routing algorithm（1.98）
XYG	113	37	2	predictive control（14.13）nonlinear system（7.43）rolling plan（4.46）generalized predictive control（3.72）mobile robots（3.72）
SZZ	105	87	3	rough set（9.37）dynamic description logic（5.92）Chinese information processing（3.57）description logic（3.23）text categorization（3.12）
SBL	89	36	1	XML（4.94）data mining（3.53）multimedia database（3.53）content-based image retrieval（2.83）ontology（2.83）
PQS	109	23	2	computer animation（12）virtual reality（7）cartoon animation（5）implicit surface（4）multiresolution representation（3）
LC	98	161	2	Wireless sensor networks（53.05）sensor network（9.25）congestion control（6.19） stochastic Petri net（5.26）performance analysis（4.88）
SCY	84	34	1	Agent（15.53）multi-agent system（9.88）intention（9.18）mental state（5.65）belief（5.65）

Empirical analyses-5

（5）Compared to expert feature provided by C-DBLP and using vector cosine formula to calculate similarity, we find that this method is more objective than expert evaluation.

Table.6. Comparison of two experimental results

Similarity / Control of relationship / Expert	without control of relationship	under the control of equivalence relationship and hierarchical relationship
ZWM	0.1980	0.4951
YJY	0.6428	0.6055
XXF	0.1010	0.2315
XL	0.0000	0.0981
XYG	0.5774	0.7127
SZZ	0.2372	0.2372
SBL	0.2988	0.4939
PQS	0.1961	0.2224
LC	0.1818	0.4671
SCY	0.3333	0.3333
Average value	0.2766	0.3897

Empirical analyses-6

（6）Table 6 shows that the similarity between them is 0.2766 without control of thesaurus. When we do manual control over synonymous and hierarchical relationships among words, such as consolidation "computer animation“ and "cartoon", Petri and Petri nets, the similarity is upgrade to 0.3879. Likewise, if we bring in outside thesaurus, and join the control word correlation, the similarity may be further enhanced.

Table.6. Comparison of two experimental results

Control of relationship / Similarity / Expert	without control of relationship	under the control of equivalence relationship and hierarchical relationship
ZWM	0.1980	0.4951
YJY	0.6428	0.6055
XXF	0.1010	0.2315
XL	0.0000	0.0981
XYG	0.5774	0.7127
SZZ	0.2372	0.2372
SBL	0.2988	0.4939
PQS	0.1961	0.2224
LC	0.1818	0.4671
SCY	0.3333	0.3333
Average value	0.2766	0.3897

Conclusion

In view of definite difference of journal sources and other multi-sources information fusion methods, for example MSDFS (Moreira et al., 2013), the result is within the acceptable range. Through table 6, we discover the similarity varies widely, our multi-source information fusion method also needs to make improvements by increasing the number of experts feature, adding vocabulary to standardize key words selection, and so on.

However it should be noted that academic resources associated with expert feature recognition take great amount, for example, Moreira et al. (2013) use DBLP Computer Science Bibliography Dataset as data source, Lu et al. (2008) choose teachers of Wuhan University for test object. The research on "niche experts", "Web experts" also need to be further done.

Thanks for your attention！