# Recommendation System of Grants-in-Aid for Researchers by using JSPS Keyword

Shin Kamada

Dept. of Intelligent Systems Fa Graduate School of Information Sciences Hiroshima City University

3-4-1, Ozuka-Higashi, Asa-Minami-ku Hiroshima, 731-3194, Japan da65002@e.hiroshima-cu.ac.jp Takumi Ichimura Faculty of Management and Information Systems Prefectural University of Hiroshima

> 1-1-71, Ujina-Higashi, Minami-ku Hiroshima, 734-8559, Japan ichimura@pu-hiroshima.ac.jp

Takanobu Watanabe Community Liaison Center Prefectural University of Hiroshima

1-1-71, Ujina-Higashi, Minami-ku Hiroshima, 734-8559, Japan t-watanabe@pu-hiroshima.ac.jp

Abstract-An acquisition of a research grant is important for the researchers to conduct a research. The university will build up the organization and reinforce the acquirement of external funds. The researcher becomes aware of grant information and should investigate what kinds of grant it is. Therefore, the staff at the support center for the Industry-Academia collaboration will classify the grant into some categories according to the research fields. However, the task is difficult to realize the matching of the research fields, because the expert knowledge is required to completely classify them. We have developed recommendation system of Grant-in-Aid system for researchers by using JSPS (Japan Society for the Promotion of Science) keywords. The characteristic keywords are extracted from web sites and then the association rules between researchers and grants are determined in the IF-THEN rule format. This paper discusses the experimental results by using the developed system.

*Index Terms*—Recommendation system, Grants-in-Aid for researchers, Japan Society for the Promotion of Science (JSPS) Keyword, Association Analysis

#### I. INTRODUCTION

Advances in recent information technology enable to collect various database not only numerical values but also comments, numerical evaluation, and binary data such as pictures. The technical methods to discover knowledge in such databases are known to be a field of data mining and developed in various research fields. Association Analysis [1] is one of famous data mining methods which find interesting knowledge in large database.

An acquisition of a research grant is important for the researchers to conduct a research. The university will build up the organization and reinforce the acquirement of external funds. The researcher becomes aware of grant information and should investigate what kinds of grant it is. Therefore, the staffs at the support center for the Industry-Academia collaboration will classify the grant into some categories according to the research fields, then recommend them each department. However, the task is difficult to realize the matching of the research fields, because the expert knowledge is required to completely classify them. In addition to this, there is a case that some grant information don't have enough amount of information to classify because it is paper based poster or email. In this case, the staffs must refer additional information such as Web site. Moreover, it takes much time for classification that is suitable for an individual because of self-classification. But, researcher fields may be different even if they belong to same department. In this way, it is very important problem that the matching between grants information and researchers in university don't well work.

In this paper, we developed matching system between grants information and researchers by using JSPS (Japan Society for the Promotion of Science) [2] keyword [3]. The characteristic keywords with high frequency are extracted from Web site of a grant organization and then JSPS keyword table is used to determine the researcher field of a grant information. Morover, association rules with strong relationships between the characteristic keywords can be extracted by Association Analysis. This paper explains about our developed recommendation system and discusses some experimental results.

The remainder of this paper is organized as follows. In Section II, the basic concept of Association Analysis and Apriori algorithm which is the well-known algorithm of Association Analysis will be explained briefly. Section III describes our developed recommendation system of Grants-in-Aid by using JSPS keyword. Section IV describes some experimental results. In Section V, we give some discussions to conclude this paper.

### **II. ASSOCIATION ANALYSIS**

This section describes the basic concept of Association Analysis [1] and Apriori Algorithm [4] which is fast algorithm for them.

Association Analysis [1] is one of famous data mining methods, which is often used to discover hidden relationships in large database. Table I shows an example of database in As-

<sup>©2015</sup> IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Transaction ID	ItemSets
1	{Item 1, Item 2}
2	{Item 2, Item 3, Item 5}
3	${\text{Item 2, Item 3}}$
4	{Item 4}
5	{Item 2, Item 3}

sociation Analysis. A database consists of some transactions, each transaction has some items (the database of Table I has 4 transaction and 5 kinds of items). Association Analysis is try to find frequent association rules between items with strong relationships, each rule is represented  $(X \Rightarrow Y)$  such as IF-THEN rule (it means that if item X appear in a transaction, then Y will be also appear in same transaction.).

In Association Analysis, the strength of relationships of association rule is defined by several way, "Support" and "Confidence" are well-known one. "Support" shows the frequency in database, calculated by Eq. (1).

$$\operatorname{supp}(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{M},$$
 (1)

where, X and Y are items in database, M is the total number of transactions.  $\sigma(X \cup Y)$  is the number of transactions that contain both X and Y (for example, supp({Item2}  $\Rightarrow$ {Item3}) in Table I is 3/5).

"Confidence" measures how often items in Y appear in transactions that contain X, calculated by Eq. (2).

$$\operatorname{conf}(X \Rightarrow Y) = \frac{\operatorname{supp}(X \Rightarrow Y)}{\operatorname{supp}(X)},$$
 (2)

where,  $\operatorname{supp}(X)$  is the proportion of transactions that contain X in database (for example,  $\operatorname{conf}({\operatorname{Item2}} \Rightarrow {\operatorname{Item3}})$  in Table I is 3/4). The value of "Support" and "Confidence" is higher, it means that the association rule is more frequent.

However, there is computational complexity in large database because the number of possible configurations of association rules is large according to the database size. The fast algorithm to solve this problem was Apriori Algorithm, introduced by Agrawal et.al [4]. Apriori Algorithm has 2 basic principle: one is to define minimum threshold for "Support" and "Confidence", the other is that if an item set is frequent, then all of its subsets must also be frequent. Fig. 1 shows the procedure of Apriori Algorithm

Step 1)Generate candidate item set  $C_1$  that contains one item.

- Step 2)Generate frequent item set  $L_1$  with min\_support in  $C_1$ . min\_support is the minimum threshold for "Support".
- Step 3)Generate candidate item set  $C_k$  from  $L_{k-1}$ . Initial value of k is 2.
- Step 4)Generate frequent item set  $L_k$  with min\_support



Fig. 2. The overview of our developed system

in  $C_k$ . Step 5)Add 1 to k and the procedure from Step 3) to Step 4) is executed until candidate item set  $C_k$  cannot be generated.



# III. DEVELOPMENT OF RECOMMENDATION SYSTEM OF GRANTS-IN-AID BY USING JSPS KEYWORD

#### A. JSPS Keyword [2]

In this section, we explain about our developed recommendation system of Grants-in-Aid by using JSPS Keyword. JSPS is Japan Society for the Promotion of Science, the keywords are defined for each researcher fields. In our system, JSPS keywords are used to estimate researcher fields of grants information and researchers.

Table II shows an example of JSPS Keyword. It consists of 4 columns: Area, Discipline, Researcher Field and Keyword. Area, Discipline and Researcher Field means the research category, Keyword is the word related to the researcher fields. The total number of Area, Discipline, Researcher Field and Keyword is 14, 80, 322 and 3674, respectively (about 11 keywords are defined in each researcher field). Our developed system can calculate how appear JSPS keyword with strong relationships in one Web site by using Association Analysis.

#### B. Developed Recommendation System [3]

Our developed recommendation system of Grants-in-Aid for researchers will be operated as following procedure. Fig. 2 shows the overview of this system.

1) Download grant information

Download PDF and HTML files from Web site in each grant organization by using "wget" (Linux shell command).

# TABLE II AN EXAMPLE OF JSPS KEYWORDS [2]

Area	Discipline	Research Field	Keyword
Informatics	Principles of Informatics	Theory of informatics	Theory of computation, Automata theory, · · ·
		Mathematical informatics	Optimization theory, Mathematical finance, · · ·
	Human informatics	Intelligent informatics	Machine learning, Knowledge acquisition, · · ·
Complex systems	Human life science	Eating habits	Cooking and processing, Food storage, ···
		Clothing life/Dwelling life	Dwelling life, Clothing culture, · · ·
Engineering	Mechanical engineering	Materials/Mechanics of materials	Continuum mechanics, Structural mechanics, ···
		Thermal engineering	Thermophysical property, Convection, · · ·
	Integrated engineering	Aerospace engineering	Aerodynamics, Structure/Material, ···
	• • •		•••

#### 2) Convert to text files

HTML files and PDF files that obtained in Step 1) are converted to text files. Following regular expression is used to remove HTML tags.

```
< ( " [ ^ " ] * " | ' [ ^ ' ] * ' | [ ^ ' ">] ) *>
```

PDF files are converted to text files by using "pdftotext" (Linux shell command).

3) Create Transaction

Converted text files that obtained in Step 2) are divided into sub-text according to paragraph as shown in Fig. 3. Each sub-text can be seen as a transaction in association analysis.

4) Create Item Sets

Some words are extracted from a transaction by using MeCab. MeCab is well-known Japanese Morphological Analysis Engine [5]. Extracted words can be seen as items in the transaction.

5) Extract Association Rules

"Support" and "Confidence" are calculated by Eq. (1) and Eq. (2). Then, frequent association rules with min\_supp and min\_conf are extracted by using Apriori algorithm. min\_supp and min\_conf are the minimum threshold for "Support" and "Confidence". In this system, we used the R library called "arule [6]" to calculate Apriori algorithm. Finally, researcher fields are estimated to calculate the most frequent JSPS keywords in acquired association rules.

## C. Estimation of researcher fields for teachers

Our developed system also can estimate researcher fields for researchers in university as well as Section III-B. In Prefectural University of Hiroshima, each researcher data consists of research fields, research title, related keywords and so on. However, they don't have enough amount of information to extract rules by Association Analysis. Therefore, we calculate the most frequent JSPS keywords and its researcher field in researcher data of Prefectural University of Hiroshima.

### **IV. EXPERIMENTAL RESULTS**

In this section, we describe some experimental results. 18 grants information and 42 teachers were used in this



Fig. 3. How to create transaction

experiment. The following parameters were used for Apriori Algorithm:  $\min\_supp = average$  of the total support,  $\min\_conf = 0.8$ .

Table III shows the estimated researcher fields for 18 grans information. Second column of the table shows the example of acquired association rules. Third column shows the estimated researcher field.

In Artificial Intelligence Research Promotion Foundation [16], the number of acquired transactions was 2071. Table IV shows an example of transactions, following words were more frequent: "Artificial", "Intelligence", "Robot", "Information" and so on. Table V shows acquired association rules with high frequency in same organization. 248 rules were acquired as frequent rules with min\_supp and min\_conf. As a result, the research field "Informatics, Human informatics, Intelligent robotics" was determined because JSPS keywords in this research field consist of "Intelligence", "Robot" and so on.

Table VI shows the summary of estimated researcher fields for 43 researchers in Prefectural University of Hiroshima. For each department, the characterestic researcher field were acquired.

Finally, matching result between grants information and researchers was acquired as show in Table VII. Second column of the table shows the department which matched researchers belong to. Third columns shows the number of researchers matched with each grant information. About 2 researchers were matched for each grans information in "Discipline".

# V. CONCLUSION

This paper presented the our developed Grants-in-Aid system by using JSPS Keyword. By using our developed system, researcher fields of grants information and researchers in university are estimated according to JSPS keyword, matching results can be acquired instead of self-classification. As the experimental result, the matching between 18 grants information and 42 teachers was performed, roughly classification was acquired and then it enable to recommend a grant information that is suitable for a each researcher.

In future, we will improve the performance of the developed system and develop the system operating as Web system.

#### ACKNOWLEDGMENT

This work was supported by 2014 Priority Research Program assigned by University President of Prefectural University of Hiroshima.

#### REFERENCES

- T.Watanabe, A Study on Fast Algorithm for Extracting Fuzzy Association Rules, In Proc. of 26th Fuzzy System Symposium, pp.349-350, 2010.
- [2] Japan Society for the Promotion of Science (JSPS) Keywords, https://www-kaken.jsps.go.jp/kaken1/keywordList.do#top, [online], 2015.
- [3] S.Kamada, T.Ichimura, T.Watanabe, *Development of Grants-in-Aid system by using JSPS Keyword*, 2015 IEEE SMC Hiroshima Chapter Young Researchers WorkShoppp.121-124, 2015.
- [4] R.Agrawal, T.Imielinski and A.Swami, *Mining association rules between sets of items in large databases*, In Proc. of the 1993 ACM SIGMOD International Conference on Management of Data, pp.207-216, 1993.
- [5] MeCab, http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html, [online], 2006.
- [6] H.Michael, G.Bettina, and H.Kurt, A Computational Environment for Mining Association Rules and Frequent Item Sets, Journal of Statistical Software, Vol.14, Issue 15, 2005.
- [7] The Society of Yeast Sciences, http://www.yeast.umin.jp, [online], 1959.
- [8] The Japan Bifidus Foundation, http://bifidus-fund.jp/index.shtml, [online], 2008.
- [9] Fuji Foundation for Protein Research, http://www.fujifoundation.or.jp, [online], 1979.
- [10] The Public Foundation of Elizabeth Arnold-Fuji, http://www.fujizaidan.or.jp, [online], 1967.
- [11] The Food Science Institute Foundation, http://www.ryouken.or.jp, [online], 2005.
- [12] The Tojuro Iijima Foundation for Food Science and Technology, http://www.iijima-kinenzaidan.or.jp, [online], 2015.
- [13] The Foundation for Dietary Scientific Research, http://www.z-ssk.org/, [online], 2006.
- [14] Tobacco Academic Studies Center, http://www.tasc.or.jp, [online], 2012.
- [15] The Fukuhara Memorial Fund for the Studies of English and American Literature, http://daieikyo.jp/aetp/modules/bulletin/index.php?page=article&storyid=6, [online], 2009.
- [16] Artificial Intelligence Research Promotion Foundation, http://www.airpf.or.jp/, [online], 2012.
- [17] Foundation for the Fusion of Science and Technology, http://www.fost.or.jp/, [online], 1994.
- [18] Trust Companies Association of Japan, http://www.shintaku-kyokai.or.jp/, [online], 1997.
- [19] The Japan Securities Scholarship Foundation, http://www.jssf.or.jp/, [online], 1973.
- [20] Heiwa Nakajima Foundation, http://hnf.jp/, [online], 1959.
- [21] Japan Educational Mutual Aid Association of Welfare Foundation, http://www.nikkyoko.or.jp/, [online], 1959.
- [22] Mazak Foundation, http://www.mazak-f.or.jp, [online], 2003.
- [23] Obayashi Foundation, http://www.obayashifoundation.org, [online], 2001.

Grants Information	Acquired Association Rules	Area, Discipline, Researcher Field
The Society of Yeast Sciences[7]	${Microbiology} \rightarrow {Drug}$	Agricultural sciences, Agricultural chemistry,
	$Agricultural \rightarrow \{Life\}$	Applied microbiology
The Japan Bifidus Foundation[8]	$\{Breeding\} \rightarrow \{Gene\}$	Agricultural sciences, Agricultural chemistry,
	${Division} \rightarrow {Cell}$	Applied biochemistry
Fuji Foundation for Protein Research[9]	${Food} \rightarrow {Health}$	Agricultural sciences, Agricultural chemistry,
	${Health} \rightarrow {Nutrient}$	Food science
The Public Foundation of Elizabeth Arnold-Fuji[10]	${Agricultural} \rightarrow {Food}$	Agricultural sciences, Agricultural chemistry,
	${Industry} \rightarrow {Agriculture}$	Food science
The Food Science Institute Foundation[11]	${Nutrient} \rightarrow {Food}$	Agricultural sciences, Agricultural chemistry,
	${Development} \rightarrow {Food}$	Food science
The Tojuro Iijima Foundation for	$\{\text{Industry}\} \rightarrow \{\text{Food}\}$	Agricultural sciences, Agricultural chemistry,
Food Science and Technology[12]	{Inspection} $\rightarrow$ {Quality}	Food science
The Foundation for Dietary Scientific Research[13]	{Eating habits} $\rightarrow$ {Health}	Agricultural sciences, Agricultural chemistry,
	${Analysis} \rightarrow {Eating habits}$	Food science
Tobacco Academic Studies Center[14]	$\{\text{Reaction}\} \rightarrow \{\text{Stress}\}$	Agricultural, Plant production and environmental
	{Peace of mind} $\rightarrow$ {Environment}	agriculture, Science in genetics and breeding
The Fukuhara Memorial Fund for the	$\{\text{Literature}\} \rightarrow \{\text{English}\}$	Humanities and Social Sciences, Literature,
Studies of English and American Literature[15]	${English} \rightarrow {Trust}$	Literature in English
Artificial Intelligence Research Promotion Foundation[16]	${Artificial} \rightarrow {Intelligence}$	Informatics, Human informatics,
	${Variable} \rightarrow {Information}$	Intelligent robotics
Foundation for the Fusion of Science and Technology[17]	$Software \rightarrow \{Learning\}$	Informatics,
	${Advance} \rightarrow {Simulation}$	Principles of Informatics, Software
Trust Companies Association of Japan[18]	$\{Wealth\} \rightarrow \{Formation\}$	Social Sciences, Law,
	$\{Life\} \rightarrow \{Insurance\}$	New fields of law
The Japan Securities Scholarship Foundation[19]	$\{Law\} \rightarrow \{Economy\}$	Social Sciences, Law,
	$\{Welfare\} \rightarrow \{Social\}$	Social law
Heiwa Nakajima Foundation[20]	${Earth} \rightarrow {International}$	Social Sciences, Law,
	$\{\text{Government}\} \rightarrow \{\text{Law}\}$	International law
Japan Educational Mutual Aid Association of	${Differentiation} \rightarrow {Education}$	Social Sciences, Education,
Welfare Foundation[21]	$\{\text{Development}\} \rightarrow \{\text{Education}\}$	Sociology of education
Mazak Foundation[22]	${Engineering} \rightarrow {Machine}$	Engineering, Mechanical engineering,
	$\{\text{Industry}\} \rightarrow \{\text{System}\}$	Production engineering/ Processing studies
Obayashi Foundation[23]	{Molecular} $\rightarrow$ {Chemistry}	Chemistry, Applied chemistry,
	$\{\text{Heat}\} \rightarrow \{\text{Environment}\}$	Green/ Environmental chemistry

# TABLE III Estimated result for Grans information

 TABLE IV

 An example of acquired transactions in Artificial Intelligence Research Promotion Foundation[16]

Transaction ID	ItamSata
Transaction ID	hemsets
1	{Artificial, Intelligence, Knowledge, Network, Learning, Control, Information }
2	{Character recognition, Image, Authentification}
3	{Automaton, Probability, System, Developement}
4	{Artificial, Intelligence, Relationship, Information, Network}
5	{Human, Interface, Contents}
6	{Robot, Sensor, Effectiveness}
7	{Intelligence, Information, System}
8	{Hybrid, Self-support, Movement, Robot, Vision System}
	•••

# TABLE V Acquired frequent association rules in Artificial Intelligence Research Promotion Foundation[16]

Association Rules	Support	Confidence
${Artificial} \rightarrow {Intelligence}$	0.1033	0.94
${Variables} \rightarrow {Information}$	0.0038	0.88
$\{\text{Older person}\} \rightarrow \{\text{System}\}$	0.0038	0.88
$\{Hybrid\} \rightarrow \{Automaton\}$	0.0024	1.00
$\{Composition\} \rightarrow \{Voice\}$	0.0024	0.83
${\text{Design}} \rightarrow {\text{System}}$	0.0024	0.83
$\{Venture\} \rightarrow \{Robot\}$	0.0010	1.00
$\{Risk\} \rightarrow \{Knowledge\}$	0.0010	1.00
${Advance} \rightarrow {Information}$	0.0009	1.00

 TABLE VI

 Estimated result for researchers in Prefectural University of Hiroshima

Department	Area	Discipline	Researcher Field
Intercultural Studies	Humanities	Literature, Linguistics	Japanese literature, Literature in English
	Social sciences	Psychology, Sociology	Social psychology, Sociology
Health Sciences	Complex systems	Human life science	Eating habits
	Agricultural sciences	Agricultural chemistry	Food science
Management	Social sciences	Management	Management
		Economics	Money/ Finance
Management Information System	Informatics	Human informatics, Principles of Informatics	Intelligent informatics, Information security
	Engineering	Electrical and electronic engineering	Control engineering/ System engineering

# TABLE VII MATCHING RESULT BETWEEN GRANTS INFORMATION AND RESEARCHERS

Grants Information	Matched deparcher	The number of matched researcher
The Society of Yeast Sciences[7]	Health Sciences	1 (matched in Discipline)
The Japan Bifidus Foundation[8]	Health Sciences	1 (matched in Discipline)
Fuji Foundation for Protein Research[9]	Health Sciences	2 (matched in Discipline)
		1 (matched in Researcher Field)
The Public Foundation of Elizabeth Arnold-Fuji[10]	Health Sciences	2 (matched in Discipline)
		1 (matched in Researcher Field)
The Food Science Institute Foundation[11]	Health Sciences	2 (matched in Discipline)
		1 (matched in Researcher Field)
The Tojuro Iijima Foundation for	Health Sciences	2 (matched in Discipline)
Food Science and Technology[12]		1 (matched in Researcher Field)
The Foundation for Dietary Scientific Research[13]	Health Sciences	2 (matched in Discipline)
		1 (matched in Researcher Field)
Tobacco Academic Studies Center[14]	none	none
The Fukuhara Memorial Fund for the	Intercultural Studies	8 (matched in Discipline)
Studies of English and American Literature[15]		2 (matched in Researcher Field)
Artificial Intelligence Research Promotion Foundation[16]	Management Information System	2 (matched in Discipline)
Foundation for the Fusion of Science and Technology[17]	Management Information System	2 (matched in Discipline)
Trust Companies Association of Japan[18]	none	none
The Japan Securities Scholarship Foundation[19]	none	none
Heiwa Nakajima Foundation[20]	none	none
Japan Educational Mutual Aid Association of	none	none
Welfare Foundation[21]		
Mazak Foundation[22]	Management Information System	2 (matched in Discipline)
Obayashi Foundation[23]	Health Sciences	2 (matched in Discipline)