Evaluating automatic medical concepts associations with human judgments

Yan Zhang
School of Information and Library Science, University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
(919)923-7273, 01
yanz@email.unc.edu

Xiangyu Fan
School of Information and Library Science, University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
(919)259-4246, 01
xyfan@email.unc.edu

Javed Mostafa
School of Information and Library Science, University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
(919)962-2182, 01
jm@unc.edu

ABSTRACT
This paper evaluates the associations of a subset of concepts extracted by VCGS (Vocabulary Cluster Generating System), a concept extraction and association tool, based on 6000 titles and abstracts downloaded from EBSCOhost Health Source – Consumer Edition database, against associations decided by 30 participants. The results show that after incorporating LSA (Latent Semantic Analysis) technique, the VCGS system establishes term associations that are not significantly different from users.

Categories and Subject Descriptors
D.2.8 [**Database Management**]: Database Applications – data mining.

General Terms
Algorithms, Performance, Design, Human Factors

Keywords
VCGS, user study, medical term association, LSA

1. INTRODUCTION
Concepts and relations are the foundation of knowledge and thought [6]. Identifying medical terms and relations between medical concepts is a crucial issue for any applications dealing with the analysis, understanding, or presentation of medical information [2, 3, 5]. With large amount of consumer health information readily available on the web, there is an urgent need for efficient tools to analyze the information. VCGS (Vocabulary Cluster Generating System) is a document analysis system that is able to automatically extract and determine associations among medical terms from consumer health literature [1].

It is very common that one concept can be represented by multiple terms, and one term can express multiple concepts. Thus, associations between concepts, represented by terms, are often ambiguous [7]. The judgment of concepts associations is also affected by contextual factors, such as people's current knowledge status and tasks at hand. Given the complexity and subjectivity of human judgment on associations between concepts, it is interesting to examine whether the system is able to make judgments that are close to human beings.

In this paper, we set out to compare term associations generated by VCGS, based on a set of 6000 EBSCOhost titles and abstracts, with associations decided by 30 general users. This comparison evaluates the performance of VCGS. Detailed analysis of the research results can inform the design of new techniques to enhance the performance of document analysis tools in general and VCGS in particular.

2. RESEARCH METHOD
Six thousand titles and abstracts were downloaded from EBSCOhost Health Source—Consumer Edition databases. VCGS extracted concepts and constructed associations between the extracted concepts based on the analysis of term frequency and weighting. From the 15 clusters generated by the system, we randomly selected 3 clusters. Within each cluster, we randomly selected 5 terms, resulting in 15 terms:

- Cluster 1: diet, carbohydrate, allergy, drug, and food
- Cluster 2: diabetes, hypertensions, blood, coronary, and coronary heart disease
- Cluster 3: symptom, depression, mother, pregnancy, and postpartum

Terms within each cluster were selected to form 30 term pairs (10 pairs for each cluster). Another 15 pairs consisting of terms across clusters were randomly composed as fillers. Each pair of terms was written on a small piece of paper and presented, one pair at a time in a random order, to subjects. Subjects rated each pair of the concepts on a 0-6 scale: 0 corresponds to no association or very weak association and 6 corresponds to very strong association. Subjects were also asked to think aloud during the rating process. The sessions were videotaped.

Thirty subjects, 15 males and 15 females, participated in the study. Of the 30 subjects, 22 were undergraduate students, 3 were graduate students, and 5 were university employees. Subjects' major field of study varied, including psychology, social work, physics, chemistry, journalism, geography, political science, economics, nursing, advertising, international studies, English, and Spanish. Table 1 shows the demographic information of the subjects (Number of subjects, age, computer experience (year),

Conference'04, Month 1–2, 2004, City, State, Country.

and frequency of looking for medical information (times/month)).

Table 1: Demographic information of subjects

	N	Age	Comp. Exp.	Freq.
Male	15	24.60 (8.70)	12.33 (2.85)	3.86 (8.14)
Female	15	25.47 (11.27)	15.40 (4.73)	2.43 (3.83)
Total	30	25.03 (9.90)	13.87 (4.14)	3.12 (6.21)

3. RESULTS

ANOVA tests showed that gender did not have significant impact on people's judgments on associations between terms. Among 45 term pairs, male and female subjects only showed significantly different judgments on 3 pairs of terms: hypertension—blood (p=.029), symptom—depression (p=.015), coronary—postpartum (p=.005). Thus, the subsequent data analysis considered the 30 subjects as one group.

3.1 Comparison of VCGS with human judgment on concept association

The association judgments from human beings produced 30 vectors, each consists of 45 values (After normalization, M= . 758, SD=.087). The similarity between the system and human beings was then determined by the correlations between system's judgment and the mean of human judgments. T-test was performed to examine whether there are significant differences between human and the VCGS system.

Figure 1 shows the results of the comparison between system judgment and human judgments (r=.587). T-test showed that there were significant differences between human and the VCGS system (p=.00).

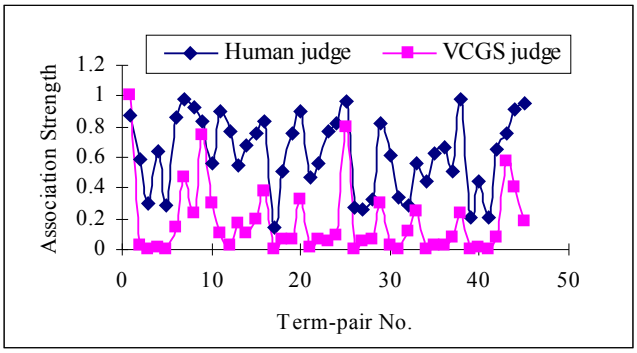

Figure 1. Term association judged by human and VCGS without LSA

3.2 Improving system performance

Since the VCGS system produced term associations that are significantly different from human beings, we worked toward improving the performance of the system. LSA (Latent Semantic Analysis) was incorporated into the original VCGS system, given the fact that LSA has been found favorable to simulate a variety of human cognitive phenomena ranging from word-categorization to discourse comprehension [4]. After incorporating LSA technique, the performance of VCGS was significantly improved. Using the optimized LSA parameter k (k=11), the dimension number of reduced space, the VCGS system produced term associations that were closer to users. The correlation of the VCGS judgment and human judgments was increased to 0.697. From T-test, there were no significantly differences between human and VCGS system (p=.168), as shown in Figure 2. The result showed that VCGS, when incorporates LSA, is able to simulate human judgments in determining medical concepts associations.

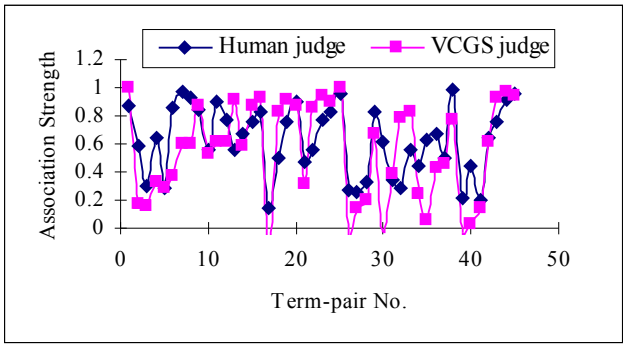

Figure 2. Term association judged by human and VCGS with LSA (k=11)

4. CONCLUSION

We compared VCGS (Vocabulary Cluster Generating System) generated associations between medical terms, based on a consumer health document corpus, to general users' judgments of the associations. Our user study results and the effort to improve the system performance showed that LSA (Latent Semantic Analysis) is needed for the VCGS system to produce estimations of the strength of terms associations that are close to users.

5. REFERENCES

[1] Fu, Y., Bauer, T., Mostafa, J., Palakal, M., and Mukhopadhyay, S. 2002. Concept extraction and association from cancer literature. In Proceedings of the 4th International Workshop on Web Information and Data Management, pages 100-103, McLean, Virginia.

[2] Jacquemin, C., and Bourigault, D. 2003. Term extraction and automatic indexing. In R. Mitkov (Ed.). The Oxford Handbook of Computational Linguistics. NY: Oxford University Press (pp599-615).

[3] Khoo, C.S.G., and Na, J. 2006. Semantic relations in information science. Annual Review of Information Science and Technology, 40, 2006, 157-228.

[4] Landauer, T. K., Foltz, P.W., and Laham, D. 1998. Introduction to latent semantic analysis. Discourse Processes, 25, 259-284.

[5] Lin, X., Marchionini, G., and Soergel, G. 1993. Category-based and association-based map display by human subjects. Proceedings of the 4th ASIS SIG/IR Classification Research Workshop, October (pp. 147-164). Columbus, OH.

[6] Novak, J.D. 1998. Learning, Creating, and Using Knowledge. Mahwah, NJ: Lawrence Erlbaum.

[7] Wang, P. 1999. An empirical study of knowledge structures of research topics. In L. Woods (Ed.), Proceedings of the

62nd ASIS annual meeting (pp. 557-586), Medford, NJ: Information Today.