Extracting Remarkable Temporal Patterns of Technical Terms in Medical Research Documents

Hidena Abe, Shusaku Tsumoto

Department of Medical Informatics, School of Medicine, Shimane University, Japan

Abstract

In temporal text mining, some importance indices such as simple appearance frequency, tf-idf, and differences of some indices play the key role to recognize remarkable trends of terms in sets of documents. However, most of conventional methods have treated their remarkable trends as discrete statuses for each time-point or fixed period. In order to find their trends as continuous statuses, we have considered the values of importance indices of the terms in each time-point as temporal behaviors of the terms. In this paper, we describe the method to extract remarkable temporal patterns by using linear trends of importance indices of technical terms related to migraine drug therapy on the documents from MEDLINE.

Keywords:

Text mining, Migraine, Drug therapy, Temporal patterns, Linear regression

Introduction

In this poster, we used an integrated framework for detecting trends of technical terms by combining automatic term extraction methods, importance indices of the terms, and a trend analysis method [1]. A case study shows the availability for finding emergent, popular, and subsiding temporal patterns of technical terms related to migraine drug therapy on MEDLINE documents.

Experimental Result

In this case study, taking a search scenario, we identify following the temporal patterns by using the degrees and the intercepts of the trend lines for each term: "emergent" for ascending trend lines with negative intercept, "popular" for ascending trend lines with positive intercept, and "subsiding" for descending trend lines. As for a search scenario over time, a user conducts the search for abstracts of articles "migraine/drug therapy [MH:NOEXP] AND YYYY [DP] AND clinical trial [PT] AND english [LA]" through PubMed year by year. As described in [2], the articles related to drug usages about migraine therapy are changed over time. With this search query, we got the articles with abstract mode of PubMed from 1980 to 2009. Then, the automatic term extraction method extracts 61936 terms.

Table 1 shows the numbers of emergent terms, popular terms, and subsiding terms based on the assigned meaning for the linear trends on the abstracts. The numbers of emergent and popular terms are increasing for the terms including the MeSH terms, which increase their appearance in recent years. However, some traditional drugs are still popular for the researches.

Table 1 – The trends of extracted terms including MeSH terms in the scenario about migraine drug therapy.

MeSH Term	# of related terms	tf-idf			Jaccard Coefficient			Odds		
		Emergent	Popular	Subsiding	Emergent	Popular	Subsiding	Emergent	Popular	Subsiding
Propranolol	1	1	0	0	1	0	0	1	0	0
Cinnarizine	1	1	0	0	0	1	0	0	1	0
Cal. Channel blockers	3	1	1	1	1	1	1	1	0	2
Acetaminophen	2	1	1	0	1	1	0	1	1	0
Flunarizine	13	2	1	10	2	1	10	2	1	10
Piperazines	1	0	1	0	0	1	0	0	0	1
Naproxen	11	11	0	0	11	0	0	11	0	0
Clonidine	1	0	0	1	0	0	1	0	0	1
Dihydroergotamine	17	6	8	3	4	9	4	3	10	4
Analgesics	16	9	7	0	6	7	3	6	6	4
Aspirin	11	2	3	6	1	4	6	0	3	8
Sumatriptan	211	93	105	13	80	128	3	79	110	22
Serotonin agonists	2	1	1	0	0	2	0	0	2	0

Conclusion

By using the linear trends of the importance indices of the terms, we identified emergent, popular, and subsiding terms related to migraine drug therapy in both of abstracts and titles.

References

[1] Abe, H., Tsumoto, S.: Detecting temporal trends of technical phrases by using importance indices and linear regression. In Proc. of the 18th International Symposium, ISMIS 2009, Springer, 2009; pp.251-260.

[2] Srinivasan P. MeSHmap: A text mining tool for MEDLINE. Proc. AMIA Symp., 2001; pp.642-646.

Address for correspondence

Hidenao Abe: Department of Medical Informatics, School of Medicine, Shimane University. 89-1 Enya-cho Izumo Shimane 6938501 JAPAN. abe@med.shimane-u.ac.jp