ORIGINAL RESEARCH

A study on accuracy improvement of knowledge extraction from the medical package inserts

Takashi Ikoma[*1], Masakazu Takahashi[2], Kazuhiko Tsuda[1]

[1] *Graduate School of Business Sciences, University of Tsukuba, Japan*
[2] *Graduate School of Innovation and Technology Management, Yamaguchi University, Japan*

Abstract

This paper presents the evaluation method of the effect range from the package insert of medicine with text mining. Most of the people who take the over the counter medicines cannot understand the medicinal effects. This is because they have little knowledge of the medicine. The ingredients of the over the counter medicines are made from prescription products. The prescription product shows various laboratory findings to obtain authorization from the Ministry of Health, Labour and Welfare. The medical information is described in the package insert of the medicine, and anyone is available. It is possible to evaluate the effect of the ingredient in the over the counter medicines, if we analyzed the medical information described in this package insert of the medicine with text mining method. This paper, we focus on both the antipyretic and the antitussive among lots of the over the counter medicines and do intensive research with them. However, variability is observed in the results of the analysis. We analyze the effect of magnification of each experimental configuration. Since the size of this fluctuation range comes from the experimental configuration. Therefore, we summarize the medication type and the experimental data of the medicine in each experimental configuration. As a result, we succeed in the effective range of medicine with text mining, if we extracted the experimental configuration and analyzed above summarization.

Key Words: Package insert, Medicine, Effect range of medicine, Experimental configuration, Quantitative comparison research, Text mining

1 Introduction

In many cases, Over The Counter Medicines to purchase are chosen by consumers without the knowledge of medicine. The criteria for selection are symptoms of the patients primarily. Whereas various ingredients are included in Over The Counter Medicines to cope with many symptoms. For example, in the case of cold medicine, antitussive, stuffy nose, the effect including the alleviation of fever varies. Besides, as for each Over The Counter Medicines, the effect is characteristic, and the ingredients and the quantity contain variables. Also, the information such as labora-tory findings is not described in the package inserts on the Over The Counter Medicines about each ingredient. Therefore, the comparison of effect with other Over The Counter Medicines is difficult. Also, a lot of medicines with the same antipyretic activity are presented. The comparison of effect is enabled easily to obtain the same effect in all Over The Counter Medicines if we contained the same medicine. However, the Over The Counter Medicines is the same purpose, but they do not contain the same medicine. Therefore, the comparison of the effect between the Over The Counter Medicines becomes difficult. At first, it is necessary to de-

*Correspondence: Takashi Ikoma; Email: Ikoma@gssm.otsuka.tsukuba.ac.jp; Address: Graduate School of Business Sciences, University of Tsukuba, 3-29-1, Otsuka, Bunkyo, Tokyo 112-0012, Japan.

termine an effect of one of the medicines contains by Over The Counter Medicines to solve this problem.

The package inserts distributors in Japan are described following points; warning, contraindications, composition and properties, dosage and administration, usage notes, pharmacokinetics, clinical outcomes, efficacy, pharmacology and active ingredient, respectively. Not only in Japan, such as the United States, the package inserts contain information required to dose medicine even though there are some differences in the documents. However, the test data in the documents do not mention all. If we are able to estimate the effect of the medicine from the package insert, it is helpful to people those who buy medicines.

The components which contain by Over The Counter Medicines are an ethical medicine. Among ethical medicines, we examine the likelihood of the comparison of the analgesic effect between ethical medicines in medicines with analgesic effect by performing text mining of the package inserts's information about those medicines. The effect, comparison of the medicine is possible, but unevenness is found in the results. Therefore, we consider whether we lowered the unevenness of the effect, comparison of the medicine.

Experiments to confirm various safety and effects are conducted before the medicine is approved by the Ministry of Health, Labour and Welfare. The experimental data are extracted and described in a package insert. Therefore, performed experimental data are described in the various experiment backgrounds to boil a package insert. Because the data on the various experiment backgrounds are included, as to the effect, comparison that performs text mining for the information on the attached document, and is obtained, unevenness is thought to be wise.

In this paper, we determine the difference in magnification of the effect of the experimental configuration. And we consider whether you can perform a narrowing of an effective range. The rest of the paper is organized as follows: Section 2 discusses related work; Section 3 discusses the backgrounds of the research. Section 4 briefly summarizes the Package Insert of the Medicine; Section 5 describes the analytics of the data and presents analytical results of filtering of the effective range of the package insert of the medicine; and Section 6 gives some concluding remarks and future work.
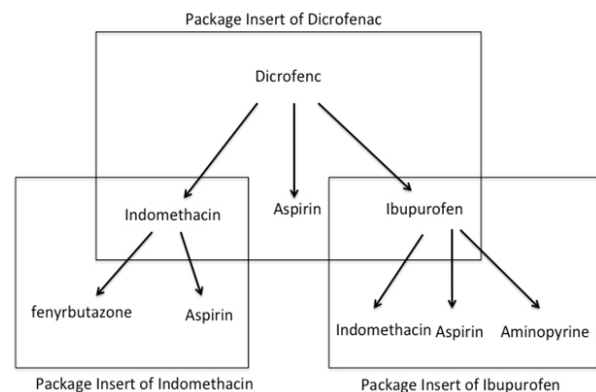
2   Related work

The text mining performs the natural language processing on text documents with data mining techniques. This technique generates new knowledge based on the tendency, co-occurrence or the characteristics among text document.[1–4] Kurohashi et al. also report the studies on the analysis of the document structure.[5, 6]

One of the applications of text mining method, Sugiyama et al. use text mining in order to prevent medical malpractice in a prescription ordering system of hospital.[7, 8] They analyze the contents on which prescription correction is made, the ordering system is changed the contents had much correction to improve the checking of prescription. They also report the study for integrating the ordering system with a package insert of medicine. Since the statement of a package insert is complicated, construction of a database becomes complicated.

Some of the research analyzes text structure is made about notes on the use of a package insert with text mining, so far. For example, Okuya et al. describe cluster analysis of the information on medical supplies, and build the database of a package insert.[9, 10] Kimura et al. analyze the document of the package insert in order to use a package insert.[11, 12] Nabeta et al. analyze the incident report with text mining.[13–15] Nabeta et al. report studies to extract information from large amounts of data to enhance the safety of the medicine.[16, 17] Togashi et al. report a study to create a database.[18] Bisgin et al. report a study utilizing topic model to the extraction of the safety and therapeutic.[19] Although text mining is used in a medical area, the information, which anyone can obtain, is restricted. Among medical information, a package insert is public information and is the information, which can be used for analysis as information, which is easy to come to hand.

There is a report studied about the comparability of the effect of the medicines by which a direct comparison is not carried out using the text information on the package insert of a medicine. In this research, quantitative extraction of the relationship between ingredients is made using the text information indicated to the pharmacology of the ethical pharmaceutical package insert and the possibility of narrowing down of the effective range is examined in quantification of an effect. Ikoma et al. make an effect, comparison of medicines by which direct comparisons are not made by combining the information on the package inserts. Figure 1 indicates the relationship between the effects of medicine.

Figure 1: Relationship between Effect of Medicine

The package inserts are mentioned the relationship of effect with other medicines. First, the text document that contains the medicine name is extracted from the package insert. Then, the words related to the magnification of the effect between the medicines (for example, "the same" or "3 times", etc.) are extracted. Then, it is possible to make effective comparisons between the two medicines if they were compared effects among them. Effect between the two medicines can be compared in one package insert. Even though if we were able to make effective comparisons of two medicines in one package insert of medicine, it is possible to make effective comparisons among multiple medicines. Even if the only comparative effect between two medicines was made of one package insert, comparative effect between the medicines become possible to integrate information on multiple package inserts (see Figure 1). Ikoma et al. examine the effect comparability of the medicine with an analgesic and antihistamine.[20, 21] A report of the comparison of medicine effective using text mining, the width of the effective range is extensive. We suggest a method for increasing the accuracy of the effective range in the present work.

The thesis, which searched for newly published paper and described about the use of text mining[22] and the reviewing paper of text mining methods apply in the medical field.[23]

In this study, we examine even the medically of other effects whether an effective comparison is possible similarly.

3   Ethical pharmaceutical package insert

Many documents exist in the ethical pharmaceutical as information dissemination data about the active ingredient of a medicine. For example, they are a bookmark of a "precautions-for-use" explanatory and medicine, an ethical pharmaceutical package leaflet, etc. Among them, an ethical pharmaceutical package insert is a product description of the ethical pharmaceutical specified by the Pharmaceutical Affairs Law. An ethical pharmaceutical package insert is a document in which the pharmaceutical company made the basic information on the medical supplies to a doctor, dentist, and a pharmacist. It is obliged to be attached for every package unit of medical supplies. The ethical pharmaceutical package insert is established the description items in the insert by Article 52 and 54 of the Pharmaceutical Affairs Law. It is used as a material in the reference at the time of use and prescribes for prescription medicine.

The package insert is described following issues: warning, contraindications, composition and properties, dosage and administration, usage notes, pharmacokinetics, clinical performance, pharmacology, physical and chemical knowledge of the active ingredient, information on packaging. There is a variety of information in the package insert among the world. The similar contents are mentioned by many countries in the package inserts.

It is aimed at the text information indicated to "pharmacology" among the information written in an ethical pharmaceutical package insert in this research. The section of "pharmacology" is briefly indicated the information such as the results of pharmacological tests from application to approval. In many cases, there is basic research with ethical pharmaceutical performed in a test tube or to an animal and clinical study performed to humans. The pharmacological test data are provided as text information. Moreover, as for test data, in addition to information, including the validity of ethical pharmaceutical, etc., relationship with other medicines may be indicated. In addition to information such as the efficacy of medicine, there is a relationship with the other medicines are described in the test data.

The medicine with which brand names differ from making into the main ingredients the medicine of an ingredient with which the patent term has expired. If the text information on the same ingredients is collected, it will become possible to put together as information on the same ingredient. The information acquired from each ethical pharmaceutical package insert for every ingredient is collected. The characteristics of an ethical pharmaceutical ingredient are verified. Furthermore, the relationship between ingredients is combined and it becomes possible to show the relation of the ingredient of the ethical pharmaceutical that cannot be obtained from the ethical pharmaceutical package leaflet information on one ingredient.

4   Further analysis of ethical pharmaceutical package insert of ethical pharmaceutical

By analyzing package inserts with text mining showed that various information can be extracted as a last paragraph showed. Then, we focus on the document that compares the effect between the medicines in which the same effect is shown in this section. By analyzing these documents, text mining is carried out for the purpose of comparison of the effect among medicines. Many package inserts describe which compared the effect with the medicine put on the market previously, in order to verify the effect of a medicine.

This description is generally only comparison between two medicines of precedence medicine and the medicine concerned. However, if the comparative information between two or more two medicines is combined, it is possible to make the effect comparison table between two or more medicines.

Then, the construction of the comparative information between two or more medicines is tried for the medicine with an analgesic effect. An execution procedure is made are following steps:

(1) Two or more package inserts perform text mining and extract the comparative information between the two
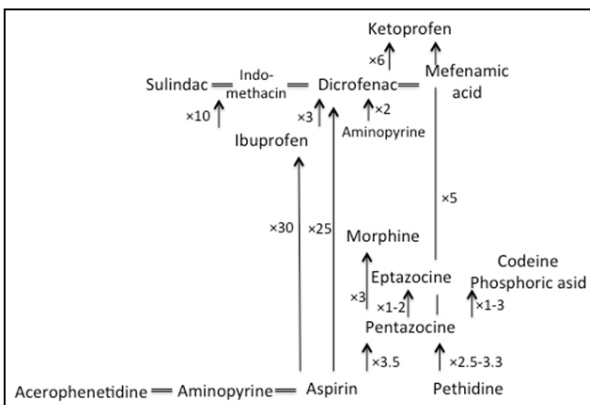
medicines.

(2) Combine the comparative information between two medicines and collect as effective comparison between two or more medicines.

(3) For example, if medicine A is 10 times stronger than medicine B, if then it is described as B → A (ex. × 10). Similarly, the relationship between the medicines describes same, then, it is described as indicated as "=".

The result is shown in Figure 2. It became possible to understand the effect between two or more medicines at a glance with Figure 2.

5 Filtering of the effect range

We examine to extract the effect comparative information focusing to the medicine with an analgesic effect with text mining in section 4. However, mismatching occurs in the effect between each the component of the ingredients.

Medical supply names and words of effect are extracted from one package insert. Then, the effect between the two medicines can be compared. In addition, Medical supply names and words of effect are extracted from the same category of package inserts in the same way. Then, effect comparison among medicine in the same category becomes possible. But, the variation of effect is large in the results obtained. Therefore, it is possible to reduce the variation, if we classified the data for each same experimental configuration.

Figure 2: Analgesic Effect Comparison between Ingredients

If we consider this result from a pharmacological viewpoint, the difference in test method will be considered to be a primary factor. For example, it is from the differences of the animals used in tests such as mice or rats. Therefore, it is assumed that the difference in test methods makes fluctuation of the medicinal effect on this paper. Specifically, the test configuration information that measured the effect between

medicines is extracted from a package insert with text mining.

In this section, integrating the difference in effect of the medicines with the test conditions performs verification.

5.1 Procedure of the effect range squeezing

The extraction procedure of the effective information on the medicine which considered the test method is shown in the next.

Three items of the following information are extracted from the package insert of the medicine A.

- The medicine for comparison of an effect (the medicine B, C)
- Experiment configuration
- The comparative result of the effect for every comparison medicine (multiple)

With these three information, the effective range will be compared such as medicine A and medicine B, and medicine A and medicine C. This procedure is performed to all the package inserts of a medicine with an analgesic effect and then the result is collected to be compared. And it is verified whether the tendency for the magnifications of the effect for every experiment configuration to be different for every experiment configuration is found out with the collected information.

5.2 Results of the effect range squeezing

Among the medicine with analgesic effect, we make a comparison with ibuprofen as test medicine for many package inserts. For a comparison of the effects with ibuprofen, following four ingredients are described in the package insert as follows:

- Aspirin
- Indomethacin
- Aminopyrine
- Oxyphenbutazone

Among them, aspirin is most compared with. Then, following eight experiment confirmations are extracted if we compared with the experimental configuration of the examination which compared ibuprofen with aspirin as follows:

- Lipoic saccharide
- Ultraviolet erythema
- Carrageenin edema
- Adjuvant arthritis
- Yeast suspension
- Randall-Selitto method
- Acetylcholine-induced writhing
- T.T.G. Fever suppression

Table 1 indicates the result of experiment configurations for each magnification of the effect on aspirin of ibuprofen. Even in the same experimental configuration, there is a range in magnification of the effect. Therefore, the Table 1 indicates the range in magnification. From this result, the magnification of the effect on aspirin of ibuprofen is 6 to 9 times with the width of the fluctuation effect is 1.5 in experiment configuration of carrageenin dropsy. There is no effect of fluctuations in experiment configuration of lipoic saccharide for magnification of the effect is 4.0 as well. The fluctuation range of an effect is 2 at the maximum if we classify the experimental configuration in Table 1. From the result, if we compared the effect of two medicines, the difference in the exact effect cannot be grasped to be taken into account experimental background for the Fluctuation width of the effect is large. On the other hand, experiment configurations have an influence on the difference in the effect of two medicines.

Table 1: Range in Magnification

Experimental Configuration	Min.	Max.
Lipoic saccharide	4	4
Ultraviolet erythema	16	32
Carrageenin edema	6	9
Adjuvant arthritis	5	10
Yeast suspension	20	20
Randall-Selitto method	30	30
Acetylcholine-induced writhing	28	28
T.T.G. Fever suppression	25	25

Table 2 indicates the result of the comparison of two same medicines call ibuprofen and aspirin. Nevertheless, among the experimental configuration of Ultraviolet erythema and Lipo-polysaccharides, the difference in the effect arose also 8 times at the maximum.

Table 2: Magnifications of an Effect for Hydrocortisone of Indomethacin

Experimental Configuration	Min.	Max.
Carrageenin edema	2	2.5
Granuloma method	4	4

Then, comparison of other medicines also checks whether the experimental configuration affects the effect of the medicine. The result extracts from the experimental data for the hydrocortisone of indomethacin is as follows. Although there were a few kinds of experiment configuration, it is confirmed that the magnifications of an effect differ for every experiment configuration. Furthermore, the result extract from the experimental data for the diclofenac of indomethacin is as follows. The magnification of the experimental configuration and the effect of this experimental data are indicated in Table 3. The same result is also confirmed in this examination. It is suggested that the experimental configuration is affected, the effect of a medicine as a result.

Table 3: Magnifications of an Effect for Dicrofenac of Indomethacin

Experimental Configuration	Min.	Max.
Adjuvant arthritis	50	50
Carrageenin edema	5	5
Randall-Selitto method	5	5

5.3 Evaluation of other effects

With the package insert of the medicine with an analgesic effect, we confirm that there was a possibility of the effect range squeezing with extracting the magnification of an effect for every experimental configuration. However, it is not clear whether we get the same trend with other medicine. Then, as for the medicine except analgetic, effective range squeezing is considered with same manner.

The antitussive contains in many generic medicines are examined. The package inserts of the medicine classify into an antitussive medicine among ethical pharmaceuticals are accumulated. Paying attention to the text information in connection with, it extracts from the text information written on the package insert. The text information is extracted with a focus on the magnification of an effect from the text information described in the package insert. Some of the experimental results are shown in below. It is confirmed that the experiments are carried out with various configurations even in antitussive medicine same as analgesics.

Next, Comparison among DL-methyl ephedrine and Codeine is indicated. Table 4 indicates the effect magnification in the different test background. Then effect magnification is different even in the same medicine by the experimental configuration. Furthermore, Table 5 indicates a list of comparative medicine and experimental configuration of DL-methyl ephedrine. From this result, if we set the experimental configuration, it is able to forecast the effect of comparison medicine.

Table 4: Magnifications of an Effect for Codeine of DL-Methyl Ephedrine Hydrochloride

Experimental Configuration	Min.	Min.
Guinea pig cough by the sulfurous acid gas inhalation	0.6	0.6
Mechanical irritation	0.4	0.4

Table 5: Effect Comparison Every Experiment Condition of the DL-Methyl Ephedrine Hydrochloride

Experimental Configuration	Experimental Medicine	Experimental animal	Max.
Guinea pig cough by the sulfurous acid gas inhalation	Codeine	Guinea pig	0.6
Guinea pig cough by the sulfurous acid gas inhalation	Ephedrine	Guinea pig	0.7

Among the medicine with antitussive, we make a comparison with dl-methyl ephedrine hydrochloride as test medicine for many package inserts. For a comparison of the effects with dl-methyl ephedrine hydrochloride, following two ingredients are described in the package insert as follows:

- Codeine
- Ephedrine

We can narrow down an effect range from information performed in same experiment system. Similar to an analgesic, it is found in the antitussive that a more correct effect comparison is possible by preparing experiment system. Furthermore, the effect comparison between the medicines that we do not compare directly by preparing experimental configuration is possible. We show the comparison of the medicine of the same experimental configuration of the dl-methyl ephedrine hydrochloride as a case.

Thus, with ephedrine and the codeine, we find that codeine is higher in an effect. Not only the information of the effect comparison of the medicine described in a package insert but also the effect comparison between the medicines that are not compared directly with information in the same experimental system is enabled.

6   Concluding remarks

In this paper, we conduct the effect comparison of medicine with the package insert of medicine. As a result, we succeed in extracting the information on the effect comparison of the medicine.

The effect comparison that is more correct because the experimental method of the medicine is considered and is made possible if we examine the effect comparison in the analgesic.

This method is able to narrow an effect range with the antitussive similarly. Furthermore, the likelihood that the effect comparison between the medicines that we cannot compare directly yet is enabled and is suggested when information with experiment scenery is obtained. A more correct effect comparison is enabled by performing text mining with the experiment scenery of the attached document together, and the effect comparison between the medicines which we do not compare directly more is enabled. It is hard to understand whether any medicine work how the information for a single package insert. But with proposed method, if we extracted and compare the effect of the medicine from the package inserts, the possibility of estimating the effect of medicine that is not compared directly in the package insert.

The prediction of the effect between the medicines that we do not compare directly by using this technique is enabled.

Our future work includes 1) We test whether a similar technique is possible in other ingredients and compare the effect of various Over The Counter Medicines more exactly, and 2) We grasp the magnification of the effect of experimental method and can connect to comparing more various medicines. These works will require practical experiments and further survey studies.

References

[1] Hiroko Y, et al. Evaluation of Methods to Extract Collocational Information from Corpus for Semantic Clustering of Japanese Polysemous Verbs. IEICE Technical Report. 1998; 98(462): 9-16.

[2] Koyama T, et al. Identification of Research Sub-Domain and Term Classification Based on Term Clustering. Special Interest Group of Natural Language Processing. 2008; 4: 87-92.

[3] Oyama T, et al. Term Extraction Using Verb Co-occurrence. Proc. 3rd International Workshop on Comitational Terminology. 2004: 79-82.

[4] Kudo T, et al. Applying Conditional Random Fields to Japanese Morphological Analysis. Science And Technology. 2004: 2-4.

[5] Kurohashi S, et al. Automatic Detection of Discourse Structure by Checking Surface Information in Sentences. Journal of natural language processing. 1994; 1(1): 3-20.

[6] Matsuzawa H, et al. Mining Structured Association Patterns from Large Databases. Transactions of Information Processing Society of Japan. 2001; 42: 21-35.

[7] Sugiyama T, et al. Investigating the Usefulness of a Prescriotion Checking System in Risk Managemet. Iryoyakugaku. 2003; 29(1): 73.

[8] Furukawa H, et al. Investigation of Possibility for Physician Order Entry System on Risk Management Strategy Related To Medication Errors. Japan journal of medical informatics. 2001; 21(1): 69-76.

[9] Okuya R, et al. A proposal for a medicine information database and text templates for generating package inserts. Medicine, Healthcare and Patient Safety. 2013; 5: 161-169.

[10] Boyce R, et al. Using Natural Language Processing to Extract Medicine-Medicine Interaction Information from Package Inserts. Proceedings of the 2012 Workshop on Biomedical Natural Language Processing. 2012: 206-213.

[11] Kimura M, et al. Analysis on Descriptions of Dosage Regimens in Package Inserts of Medicines. Human Interface and the Management of Information. Information and Interaction. Lecture Notes in Computer Science. 2009; 5618: 539-548.

[12] Kimura M, et al. Analysis of Questionnaires Regarding Safety of Medicine Use, Application of Text Mining to Free Description Questionnaires. The Japanese Journal of Ergonomics. 2005; 41(5): 297-305.

[13] Nabeta K, et al. Analysis on descriptions of precautionary statements in package inserts of medicines. Medicine, Healthcare and Patient Safety. 2012; 4: 19-25.

[14] Baba T, et al. Formal Concept Analysis of Medical Incident Reports. KES 2010, Part III, LNAI 6278. 2010: 207–214.

[15] Kawanaka H, et al. Tendency Discovery from Incident Report Map Generated by Self Organizing Map and its Development. Proceedings of IEEE International Conference on Systems, Man and Cybernetics. 2007: 2016-2021.

[16] Nabeta K, et al. A Proposal of a Method to Extract Active Ingredient Names from Package Inserts. Human Interface and the Management of Information. Information and Interaction, Lecture Notes in Computer Science. 2009; 5618: 576-585.

[17] Duke J, et al. A Quantitative Analysis of Adverse Events and "Over-warning" in Medicine Labeling. Arch Intern Med. 2011; 171(10): 944-946.

[18] Togashi H, et al. Analysis for contents data of package leaflet in medicinal supplies. Japan Journal of Medical Informatics. 2006; 26(2): 129-134. Japanese.

[19] Bisgin H, et al. Mining FDA medicine labels using an unsupervised learning technique - topic modeling. BMC Bioinformatics. 2011; 12(Suppl 10): S11.

[20] Ikoma T, et al. Effect verification of the ingredient contained in OTC by text mining. The 27th Annual Conference of the Japanese Society for Artificial Intelligence. 2013.

[21] Ikoma T, et al. A Study on Filtering of the Effect Range with the Package Insert of the Medicine. The 38th Annual International Computers, Software & Applications Conference. 2014.

[22] Wilco WM, Fleuren, et al. Application of text mining in the biomedical domain. Methods. 2015; 74: 97-106.

[23] Juliane Fluck, et al. Text mining for systems biology. Drug Discovery Today. 2014; 19(2): 140-14.