A COMPUTATIONAL SEMANTICS SYSTEM FOR DETECTING DRUG REACTIONS AND PATIENT OUTCOMES IN PERSONAL HEALTH MESSAGES

BY

YUNLIANG JIANG

THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer Science in the Graduate College of the University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Advisor:

Professor Bruce R. Schatz
Case studies are a standard approach to medicine. A physician needs the outcomes of a drug, situationally relevant to a particular patient. We propose a system for patient outcomes utilizing computational semantics, which effectively digests message groups. **Filtering** identifies personal comments, by eliminating useless messages. **Clustering** groups similar topics from different messages, by statistical overlap with specified terms. **Summarizing** labels the clusters so content can be quickly digested. We implemented a prototype system with these functions for mining health messages. Our methods do not require extensive training or dictionaries, while enabling users to specify custom topics for digesting. This system has been tested with sample messages from our unique dataset from Yahoo! Groups, containing 12M personal messages from 27K public groups in Health and Wellness. Evaluated results show high quality of labeled clustering, promising an effective automatic system for discovering useful information across large volumes of health information.
ACKNOWLEDGMENTS

This work was completed with invaluable help from my advisor, Professor Bruce R. Schatz. I am grateful for his advice, pointers, discussions, and teaching.

I thank Cindy Xide Lin for the close collaboration.

I thank Hongfei Tian, Rong Ding, Shanxiang Qi for the testing and content editing.

I thank Yue Lu and Professor Qiaozhu Mei for the algorithm support and discussion.

I thank my family for their love, support, and encouragement. Thanks to my friends for their friendship.
# TABLE OF CONTENTS

List of Tables ................................................................. v

CHAPTER 1  Introduction ...................................................... 1

CHAPTER 2  Related Work ..................................................... 4

CHAPTER 3  Problem Definition ................................................. 6

CHAPTER 4  Methodology ..................................................... 8
  4.1 Step 1: Filtering the Messages ....................................... 8
  4.2 Step 2: Clustering – Outcome Selection and Integration ............... 11
    4.2.1 PLSA Model .................................................... 11
    4.2.2 Integration Progress ......................................... 12
  4.3 Step 3: Summarizing: Automatic Labeling ............................ 13

CHAPTER 5  Experiments and Results ..................................... 15
  5.1 Data and Setup ....................................................... 15
  5.2 Filtering Result ....................................................... 15
  5.3 Clustering Result ..................................................... 16
  5.4 Evaluation ............................................................ 19
    5.4.1 Task 1. Identifying the Extra Outcomes ....................... 19
    5.4.2 Task 2. Classifying the Outcomes with Expert Comments .......... 22
    5.4.3 Task 3. Distinguishing Similar Opinions and Supplementary Opinions 23
    5.4.4 Limitation of PLSA ............................................ 23
  5.5 Labeling Result ....................................................... 24

CHAPTER 6  Conclusion ....................................................... 26

CHAPTER 7  Future Work ..................................................... 27

REFERENCES ................................................................. 29
List of Tables

4.1 Message Examples ................................................. 9
5.1 The result of the SVM classification .......................... 16
5.2 The outcome results for Meridia with expert comments ....... 18
5.3 The extra outcomes for Meridia without expert comments .. 19
5.4 The outcome results for Vioxx with expert comments ....... 20
5.5 The extra outcomes for Vioxx without expert comments .... 21
5.6 The outcome results for Tysabri with expert comments ...... 21
5.7 The extra outcomes for Tysabri without expert comments .. 22
5.8 Task 1 result ....................................................... 22
5.9 Task 2 result ....................................................... 22
5.10 Task 3 result ...................................................... 23
5.11 Label results ...................................................... 25
CHAPTER 1

Introduction

Online discussions now play an essential role in people’s daily life. When a person plans to buy an electronic product, she would like to view other customers’ reviews from shopping websites. When a user tries to trade on some specific stock, she would like to know other traders’ comments from online stock discussion board. Online discussions in specific area are so important, because they collect different users’ recent experiences. Though subjective, they reflect the most direct and comprehensive opinions from actual people using actual products.

When it comes to healthcare, the situation is similar: online bulletin boards and chat groups, such as Yahoo! Groups\(^1\) and WebMD\(^2\), offer patients and doctors a good platform to discuss health problems, e.g., diseases and drugs, diagnosis and treatment. Case studies are a standard approach to medicine. A physician wants to know how well a drug works, what its outcomes are, including side-effects. A patient wants to know the outcomes with similar patients, who did better and who did worse, under what conditions.

Online medical discussions have limitations preventing users from effectively digesting the information. If you use a drug name as the entry to search the database, there are usually thousands of comments or reviews returned. The outcomes need to be integrated and summarized in a targeted fashion. We propose an interactive system for health messages, which addresses the following problems:

\(^1\)http://groups.yahoo.com/
\(^2\)http://www.webmd.com/
1. **Filtering:** Medical discussions contain a huge number of irrelevant messages. For example, some users will quote news articles, while advertisements will also display on the discussion board. The discussions must be filtered, so that only the user comments from actual persons remain.

2. **Clustering:** Unlike product reviews, medical discussion messages are unstructured, *i.e.*, each comment talks about several topics, in just one piece of plain text. The messages must be partitioned into parts, and parts on similar topics from different messages must be grouped together to form a coherent view of all the information available from all the people on each topic.

3. **Summarizing:** Each group contains multiple messages about the topic under discussion. And each dataset contains multiple topics, *e.g.*, among all the messages discussing Meridia, some deal with weight loss, some with side effect, some with expensive price. The groups must accordingly be labeled effectively so that users can quickly find the desired messages within the desired topics.

Our system for patient outcomes solves all of these problems to effectively digest message groups. **Filtering** transforms the database into personal health messages, by deleting news and adverts. **Clustering** groups similar topics from sentences appearing in different messages, where the topic categories can be specified manually by an expert or identified automatically by the system. **Summarizing** labels the groups so that their content can be quickly digested, by giving a sample sentence from the clustered messages or a series of generated phrases. These steps propose to re-organize the information of medical online discussions in a practical way.

We have implemented a prototype system for text mining of health messages. This system has been tested with sample messages from our unique dataset from Yahoo! Groups, which contains 12M personal messages from 27K public groups in Health and Wellness. This outcomes research utilizes deeper processing of natural language, such as SVM and PLSA, than
our previous study on drug reactions with this dataset[1]. Our methods do not require extensive training nor do they require dictionaries. In addition, they enable users to specify their own topics for digesting. Thus these methods are very general and very powerful.

The rest of this paper is organized as follows: Chapter 2 summarizes the related work and techniques. Chapter 3 defines the problem, the key concepts and the symbols in principle. Chapter 4 describes the algorithm and implementation of our approaches. Chapter 5 provides experimental results indicating the high accuracy and usefulness of the system. Finally, we conclude the paper in Chapter 6 and outline the future work in Chapter 7.
CHAPTER 2

Related Work

Our paper focuses on detect the outcome of a specific drug from users’ messages. Richardson et al. [2] identified four aspects: \( P,E,C,O \) as the key aspects of clinical information: \( P \) stands for Patient-problem: what are the patient characteristics (e.g. age range, gender, etc.)? What is the primary condition or disease? \( E \) stands for Exposure-intervention: what is the main intervention (e.g. drug, treatment, duration, etc.)? \( C \) stands for Comparison: what is the exposure compared to (e.g. placebo, another drug, etc.)? \( O \) stands for Outcome: what are the clinical outcomes (e.g. healing, morbidity, side effects, etc.)? These elements are known as the \( PECO \) elements.

However, it is very difficult to extract the \( PECO \) elements from medical information using the traditional entity extraction technique due to the complexities and unsureness of the medical information, even in the formal literature. A lot of evaluation work are based on manually labeled data, such as Nie’s work [3].

A lot of applications utilized the medical formal literature to extract and integrate useful information. Such applications include: generating text summaries [4], topic modeling [5], mining predictive rules [6], etc. Our paper uses informal medical messages which are generated by multiple online users. Compared with the formal one, informal messages are more unstructured, noisy, and difficult to process.

Also some researchers paid attention to the informal medical messages. For instance, Crain et al. [7] worked on consumer medical search by using Yahoo! Answer health messages, while Chee’s work tracked users’ sentiment regarding particular drugs [1] and did information
visualization [8] based on Yahoo! Medical Group Messages. Using the same dataset as [1, 8], we do a comprehensive information extraction task and apply a series of data mining and information retrieval techniques, which are mainly introduced below.

Probabilistic latent semantic analysis (PLSA) [9] has been applied to many text mining problems with good results. Researchers use this model to solve problems such like opinion integration [10], sentiment analysis [11], web search and mining [12], etc. To the best of our knowledge, Lu’s work [10] which focuses on automatically integrating opinions from product reviews using the semi-supervised PLSA model is close to our motivation, though we have different data source and purpose.

After generating multiple topics from a bench of texts, labeling each topic with semantic terms or phrases would help to understand each topic intuitively. From the survey we find that Mei’s work [13] on automatic labeling could be applied on top of the results of our clustered medical outcomes.
CHAPTER 3

Problem Definition

For one particular drug, we collect all the related health messages from Yahoo! Groups, denoted as $M$, in which $N$ is the set of all the news, $C$ is the set of all the user comments (our target input), and $A$ is the set of advertisements. Clearly, we have $M = N \cup C \cup A$.

After successfully extracting $C$, we split it into a set of sentences, denoted as $D$. Each sentence $d \in D$ is a comment unit, which would potentially present one side of outcomes. Our target goal is to group all the comments into $m$ meaningful outcome clusters $O_1, O_2, ...O_m$, given the collection $D$.

Here are several key concepts to be introduced:

1. **Expert comment** $e_i$: To better cluster the outcomes, semi-supervised PLSA model would be applied. Expert comments aim to offer the prior knowledge for PLSA and key points for each $O_i$. For each $O_i$ we have one expert comment $e_i$. Compared with the user comments, the expert comment should be more well-written and semantically vertical to each other. Thus we collect all the set of expert comments $E$ (formed by $e_1, e_2, ...e_n$, $n < m$), for each drug, from a professional drug information website\(^1\).

2. **Similar opinion** $O_{i,sim}$ and **Supplementary opinion** $O_{i,sup}$: Each outcome $O_i$ consists of a group of comment units $D_i$ and one expert comment $e_i$. Some of the comments represent similar opinion with $e_i$, which form $O_{i,sim}$. Others reflect different opinions from $e_i$, though they still talk about the same topic. We call such collection $O_{i,sup}$. Here

\(^1\)http://www.drugs.com/
$O_{i,sup}$ can be split into several sub-clusters, where each represents one of the different opinions.

3. **Outcome Label** $l_i$: For each topic (as a cluster or sub-cluster) we generated, since it consists of multiple sentences, it is difficult to understand the meaning of the topic directly. In addition to selecting one representative sentence for each cluster, we automatically generate a set of phrases $l_i$ for each cluster, which would semantically interpret the topic.
CHAPTER 4

Methodology

From Chapter 1, we know that there are three main steps:

Filtering: The input is the whole collection $M$. We will separate it into three classes: $C, N$ and $A$. $C$ is our target, and we need to filter out $N$ and $A$.

Clustering: Outcome selection and integration: Given $C$ and $n$ expert comments $e_1, e_2, \ldots e_n$, we would generate $m$ ($m > n$) clusters $O_1, O_2, \ldots O_m$. Each refers to one discriminative outcome and can be further split into similar opinion $O_{i,sim}$ and supplementary opinion $O_{i,sup}$. Note the reason why $m > n$ is we want to generate some outcomes beyond any expert comments, called extra comments.

Summarizing: For each semantic topic (as the form of a cluster or sub-cluster), we would automatically generate a set of phrases $l_i$ to label the topic.

We will discuss the methodology of these steps in details.

4.1 Step 1: Filtering the Messages

After roughly observing several messages, we can separate them into three groups:

- **News ($N$):** When users are discussing about their health conditions, they may directly share or quote one article from news. The content of a news article is related to the drug, mainly about the latest FDA approval or scientific discovery. Although containing useful information, it is too long and full of irrelevant information, which is difficult to extract the useful part. So we will eliminate this group of messages.
• **Comments** (*C*): User comments are the most informative part in the whole collection *M* since they reflect users’ actual feelings and experiences on drugs – the outcomes. They have proper lengths and most of the parts can be directly extracted. This group of messages is our target and also the input of Step 2.

• **Ads** (*A*): Since Yahoo! Group is an open discussion board, it contains a huge number of advertisements, sent by individuals or robots. They are useless to the project and should be eliminate also. They are usually very short, and repeatable (entirely repeatable or sentence-base repeatable) which is a good feature to distinguish *A* from others.

Table 4.1 shows a snippet of each kind of messages, which are extracted from the real data.

![Table 4.1: Message Examples](image)

To classify such three groups of messages, we apply SVM classification [14] to split all the messages into three classes (*N, C, A*). Support vector machines (SVM) are a set of related
supervised learning methods that analyze data and recognize patterns, used for classification. It has relatively high accuracy and can handle high-dimensional data. It is a good solution to distinguish the messages and filter the useless ones.

To apply SVM classification in our project, there are five main steps:

1. Manually label some examples as comments, news, and ads, called training data.
2. Select useful features from the training data.
3. Transfer each message into a high-dimensional vector, where each dimension refers to one feature.
4. Train a SVM classifier on the training data, which automatically studies the weight for each dimension and find the most discriminative margin among the different groups.
5. Test the unlabeled data by the SVM classifier.

The most essential step is the feature selection, we find the following features can be considered:

- Word distribution: term frequency of each messages.
- Basic statistics of text: message length, the number of paragraphes, etc.
- Word uniqueness: the ratio of the number of unique words to the message length. (We find ads have a low ratio since the content is always repeatable).
- Some special features: e.g. how many “I” are there in one message. (user comments possibly contain many), etc.

We train a 3-class SVM classifier [15] and test on the real data. The results show a high accuracy. We will introduce the experiment details in Section 5.2.
4.2 Step 2: Clustering – Outcome Selection and Integration

To group the comments into reasonable and discriminative clusters, where each cluster represents one main outcome of the drug. Semi-supervised PLSA [9] model is applied. We would introduce the model first and then describe the integration process.

4.2.1 PLSA Model

In the PLSA model, we consider each document in the collection $D$ is generated from a mixture of $m + 1$ multinomial component models. One component model is the background model $\theta_B$ that absorbs non-discriminative (i.e., meaningless) words and the rest are the $m$ latent theme topic models (saying $\Theta = \{\theta_1, \theta_2, ..., \theta_m\}$) that are estimated by Equation 4.5 via the Expectation Maximization (EM) algorithm [16], a method for finding maximum likelihood estimates of parameters in statistical models, all the parameters and results can be computed and updated using the following formulas:

\[
p(z_d, w, j) = \frac{(1 - \lambda_B) \pi_{d,j}^{(n)} p^{(n)}(w|\theta_j)}{\lambda_B p^{(n)}(w|\theta_B) + (1 - \lambda_B) \sum_{j'=1}^{k} \pi_{d,j'}^{(n)} p^{(n)}(w|\theta_{j'})} \tag{4.1}
\]

\[
p(z_d, w, B) = \frac{\lambda_B p^{(n)}(w|\theta_B)}{\lambda_B p^{(n)}(w|\theta_B) + (1 - \lambda_B) \sum_{j'=1}^{k} \pi_{d,j'}^{(n)} p^{(n)}(w|\theta_{j'})} \tag{4.2}
\]

\[
\pi_{d,j}^{(n+1)} = \frac{\sum_{w \in V} c(w, d)p(z_d, w, j)}{\sum_{j'} \sum_{w \in V} c(w, d)p(z_d, w, j')} \tag{4.3}
\]

\[
p^{(n+1)}(w|\theta_j) = \frac{\sum_{d \in D} c(w, d)p(z_d, w, j)}{\sum_{d' \in D} \sum_{w' \in V} c(w', d)p(z_d, w', j')} \tag{4.4}
\]

\[
p^{(n+1)}(w|\theta_j) = \frac{\sum_{d \in D} c(w, d)p(z_d, w, j)}{\sum_{w' \in V} \sum_{d \in D} c(w, d)p(z_d, w', j')} \tag{4.5}
\]

To better cluster $m$ aspects, we can enroll some prior knowledge by extending the basic PLSA to incorporate a conjugate prior defined base on expert comments $e_1, e_2, ..., e_n$ (i.e.,
semi-supervised). For each outcome cluster $O_i$ ($1 \leq i \leq n$), since we have already acquire the expert comment $e_i$, we can build a unigram language model $\{p(w|e_i)\}$ and interpret it into the above formula, after several computing, Formula 4.5 turns into the following format:

$$p^{(n+1)}(w|\theta_j) = \frac{\sum_{d \in D} c(w, d)p(z_{d,w,j}) + \mu p(w|e_j)}{\sum_{w' \in V} \sum_{d' \in D} c(w', d')p(z_{d',w',j}) + \mu} \quad (4.6)$$

Note, for the cluster $O_i$ ($n < i \leq m$), there is no prior knowledge since we aim to find some extra outcome topics rather than the experts.

### 4.2.2 Integration Progress

To build these $m$ meaningful clusters by applying semi-supervised PLSA model, there are several steps described below:

1. Build the prior knowledge. For each cluster $O_i$ ($1 \leq i \leq n$), we have already acquired an expert comment $e_i$. Based on it, we estimate $p(w|e_i)$ by Maximum Likelihood as the prior estimator. Here only adjectives, adverbs, verbs and nouns are considered in the estimator since they are the terms which express the opinions.

2. Given such priors and the set of the comment units $D$, we could estimate the topic models $\{\theta_1, \theta_2, .. \theta_m\}$ by the formulas in Section 4.2.1.

3. For each comment unit $d \in D$, we assign it to the most suitable cluster by the following formula:

$$\arg \max_j p(d|\theta_j) = \arg \max_j \sum_{w \in V} c(w, d)p(w|\theta_j) \quad (4.7)$$

4. Now in each cluster $O_i$, there are a couple of assigned comment units. For the cluster which has a prior expert comment, we would separate them into Similar opinion $O_{i,sim}$.
and Supplementary opinion $O_{i\text{,sup}}$, by applying the semi-supervised PLSA model to create two clusters with $e_i$ as one cluster’s prior knowledge.

5. In each Supplementary opinion $O_{i\text{,sup}}$, it may contain different opinions and we apply a basic PLSA model without any prior knowledge to create several meaningful sub-opinions.

6. Now for each opinion, we select one sentence from the assigned comment units to represent such topic, which is easier for users to understand compared to topic models, and associated with the supporting value (the size of the clusters).

We apply the above process on the real data and get meaningful results. We will show the result and evaluate it in Section 5.3 and 5.4.

### 4.3 Step 3: Summarizing: Automatic Labeling

From Mei’s analysis [13], a good label for text should be understandable to the users, and could capture the semantic meaning of the topic. Compared with sentences or individual words, a phrase is coherent and concise for a user to read and understand. Avoiding manually labeling each topic $i$, we can use an automatic labeling technique to generate a set of phrases $l_i$ for each semantic cluster. The main steps are as below:

1. Candidate label generation: The first step is to generate a collection of labels as the format of phrases. There are several methods, such as: chunking/parsing, Ngram testing, etc. We use the Ngram testing to generate bigram phrases from the original sentence clusters. We generate the most significant bigrams using the N-gram Statistics Package [17], and use T-test to choose the ones which have the highest T-Score [18].
2. Relevance scoring: Intuitively, a phrase which contains more important words in the topic, would be seen as a good label. Here, the importance can be defined as the high frequency of the word in the topic model. We use $p(w|\theta)$ to present the word distribution of the topic $\theta$, and the relevance score of a candidate phrase $l = w_0w_1...w_k$ could be defined as:

$$Score = \log \frac{p(l|\theta)}{p(l)} = \sum_{0 \leq i \leq k} \log \frac{p(w_i|\theta)}{p(w_i)}$$

(4.8)

3. Choosing $l_i$: After computing the score for each candidate label, we choose the ones which have the top score, forming the set of labels of one cluster $i$.

In Section 5.5, for each drug, we automatically generate semantic labels for each similar opinion $O_{i,\text{sim}}$, as well as each extra outcome cluster without expert comments.
CHAPTER 5

Experiments and Results

5.1 Data and Setup

We utilize health messages from Yahoo! Groups to build the SVM classifier, topic model and test. The forum consists of 27,290 groups and over 12,519,807 messages totally, spanning seven years and multiple topics. The data is also applied in Chee’s work [1].

All the experiment is running on a 4TB-disk, 4GB-RAM, and 10-core server.

5.2 Filtering Result

Since we build a drug-based system, for each drug, we need to select the related messages which form the whole collection $M$, and then spit it into three categories: $N$, $C$ (target) and $A$ by applying the SVM classification method.

We choose three drugs as our investigation objectives: Meridia(diet drug), Vioxx(arthritis drug) and Tysabri(multiple sclerosis drug), all of which have been withdrawn by FDA since serious side efforts. However, it would be an advantage for our task since many clear-stated outcomes would be described by users.

We randomly choose 141 messages from the collection of Meridia as the training data. After manually labeling them we acquire 33 pieces of news, 37 comments and 71 pieces of ads. From this we can see the personal health messages are really dirty – a huge number of data is useless and should be eliminated.
We choose the features according to Section 4.1 and get an 800-dimensional vector space (most of the dimensions represent the word distribution). Following the steps of the SVM classification, we transfer the data into vector form, test the unlabeled data, and chunk the messages in $C$ into sentences. Table 5.1 shows the statistics of the classification.

Table 5.1: The result of the SVM classification

<table>
<thead>
<tr>
<th>Drug Name</th>
<th># of messages</th>
<th># of user comments</th>
<th># of commend units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meridia</td>
<td>672</td>
<td>160</td>
<td>3752</td>
</tr>
<tr>
<td>Vioxx</td>
<td>3799</td>
<td>748</td>
<td>19236</td>
</tr>
<tr>
<td>Tysabri</td>
<td>919</td>
<td>56</td>
<td>1317</td>
</tr>
</tbody>
</table>

To test the robust and accuracy of our SVM classifier, first we test the training data back and get 139 correct labels and 2 incorrect ones. Second we randomly choose 60 messages from the test pool (20 for each drug) and manually label them as the standard, we get 56 correct labels and 4 incorrect ones. The accuracy rate is 93.3%. Such results show our SVM classifier can successfully filter out useless messages and acquire the target group $C$.

5.3 Clustering Result

Now we have the set of comment units $D$ for each drug, according to the semi-supervised PLSA model introduced before, we would generate the outcome results for each drug.

Table 5.2 and Table 5.3 show the clustering results with expert comments, and the extra outcomes without prior knowledge, respectively, for the drug of Meridia. From Table 5.2, eventually we build 9 clusters for Meridia, where each of them focuses on one meaningful semantic outcome. For example, Outcome ID 1 mainly talks about the topic of Meridia’s use. From the expert comment $e_1$ we know taking Meridia may help to lose weight. And
298 comments support the similar opinion. While others express different opinions although the same topic: weight and function. For example, 77 comments say the wish to lose weight. 70 comments talk about another function of Meridia. From such one row, we would easily get the general knowledge and statistics of user comments on this outcome of Meridia. This information is scattered in the huge amount of messages and impossible to integrate by hand without such system.

Besides the 9 outcomes for Meridia which cover the main aspects that user may be interested in: effectiveness, why taking it, how much and how long to take it, various of side efforts, and the price, we also find 3 extra outcomes which are not mentioned in the expert comments. For example, Outcome ID 12 shows the appreciation of the usefulness of Yahoo! Health Group. These outcomes would make the whole condition of one particular drug more completely, rather than just following the professional’s.

Read the information from both of the tables, one may find the result of our system would effectively collect and integrate the drug-based information and display it in a well-readable way.

Table 5.4 and Table 5.5 show the corresponding clustering results for Vioxx. And Table 5.6 and Table 5.7 refer to the condition of Tysabri. From the results we can see our system would not only work well for one drug, but also fit other drugs.

One of the advantages of the PLSA model against other simple statistic methods relies on its capacity of capturing the coherence of terms (e.g., appositive, synonym) the PLSA model is able to detect such connection between the two comments which contain since they share the similar contexts. Take one of the Meridia’s outcome (ID 2) as an example, the expert comment is “Take Meridia exactly as prescribed by your doctor and follow the directions on your prescription label.”. The 133 similar sentences contain not only “i am seriously considering going to a new doctor” which capture the word “doctor” exactly, but also “tracy you might ask your
Table 5.2: The outcome results for Meridia with expert comments

<table>
<thead>
<tr>
<th>ID</th>
<th>Keywords</th>
<th>Expert comment</th>
<th>Similar Opinions</th>
<th>Supplementary Opinions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>lose</td>
<td>Meridia is used together with diet and exercise to treat obesity. It helps to lose pounds in weeks.</td>
<td>[s=298] I have been on meridia for 14 months and have lost 93 pounds</td>
<td>[s=77] I just know that i want the extra weight off of me. [s=60] and I worked out with weights 3 times per week [s=70] evidently maybe the drugs really bombed out her immune system</td>
</tr>
<tr>
<td>2</td>
<td>doctor</td>
<td>Take Meridia exactly as prescribed by your doctor and follow the directions on your prescription label.</td>
<td>[s=133] the doctor gave me a prescription for 2 month and we will see if we can not get at least a good start on weightloss for me</td>
<td>[s=105] I called the medical hotline at meridia a couple of days ago to ask a few questions [s=61] I simply asked a short question</td>
</tr>
<tr>
<td>3</td>
<td>prescription</td>
<td>Meridia is usually taken once daily. Meridia can be taken with or without food.</td>
<td>[s=140] I am on 10 mg once per day at lunch time just because it is easier for me to remember</td>
<td>[s=65] It is the week before and the week during that I feel the thyroid symptoms the worse [s=62] yesterday marked the beginning of month 11 for me [s=53] how long have you been on meridia</td>
</tr>
<tr>
<td>4</td>
<td>2 years</td>
<td>Meridia should not be taken for longer than 2 years.</td>
<td>[s=228] I was just going to get another prescription after taking it for 2 years</td>
<td>[s=66] It is the week before and the week during that I feel the thyroid symptoms the worse [s=62] yesterday marked the beginning of month 11 for me [s=53] how long have you been on meridia</td>
</tr>
<tr>
<td>5</td>
<td>avoid</td>
<td>Avoid taking cough and cold or allergy medications while taking Meridia. Avoid drinking alcohol while taking Meridia.</td>
<td>[s=108] I may have shared this earlier was the fact that you have no desire for alcohol</td>
<td>[s=91] I do not understand your reaction to the water issue [s=69] whether I was drinking a sports drink drinking a health food bar or Campbell's soup I was consuming some type of detrimental sugar</td>
</tr>
<tr>
<td>6</td>
<td>allergy,</td>
<td>dangerously high blood pressure</td>
<td>[s=143] Watch your blood pressure as some patients have had sharp rises in their BP</td>
<td>[s=61] and he believes if the blood pressure elevates at all the patient should be taken off [s=42] he wanted to wait 6 months and do another blood test</td>
</tr>
<tr>
<td>7</td>
<td>alcohol</td>
<td>fast, pounding, or uneven heartbeats</td>
<td>[s=137] I guess meridia can be hard on the heart so my doctor wants to be sure my heart is good                                                                                                                                                                                                                                                                 72] they are tracking any unusual symptoms [s=76] although I would have never considered myself depressed in the past</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>pain,</td>
<td>chest pain or heavy feeling, pain spreading to the arm or shoulder, general ill feeling, headache, back pain, joint pain</td>
<td>[s=132] In the beginning I had severe headaches and insomnia but that went away within 3 4 days</td>
<td>[s=49] bend the knee of your other leg and bring that knee up toward your stomach [s=79] It is really getting me down and I am an emotional eater so that is not good</td>
</tr>
<tr>
<td>9</td>
<td>expensive,</td>
<td>It is expensive, most cannot afford it.</td>
<td>[s=101] It is expensive but worth it</td>
<td>[s=75] I have had no side effects at all [s=92] the cost is 196</td>
</tr>
</tbody>
</table>
Table 5.3: The extra outcomes for Meridia without expert comments

<table>
<thead>
<tr>
<th>ID</th>
<th>Extra Opinions</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>[s=372] i wish you luck on meridia</td>
</tr>
<tr>
<td>11</td>
<td>[s=276] i started taking it in november 2004 when i weighed in at 325lbs</td>
</tr>
<tr>
<td>12</td>
<td>[s=286] it's nice seeing more postings its encourage to know that there are more out there with lots in common</td>
</tr>
</tbody>
</table>

or “i have learned to ask my physician questions concerning the particular medication” which capture the word “dr”, “physician”, the same meaning of “doctor”.

5.4 Evaluation

In this part we would evaluate the effectiveness of our integration system. We design three tasks of evaluation, aiming to check how well our system-generated results would match human choices. In the experiment, we choose three volunteers to be the judges, who are from the major of computer science, nutrition and bioinformatics.

5.4.1 Task 1. Identifying the Extra Outcomes

In this task, we plan to identify the extra outcomes from the outcomes with prior knowledge. For each drug, we randomly choose 2 assigned sentences from each of the system-generated outcome clusters (including clusters with prior knowledge and clusters without it). So we totally have 24 sentences for Meridia and Vioxx, and 18 sentences for Tysabri. All the judges need to select 6 sentences which do not fit into the clusters with prior knowledge.

Table 5.8 shows how the system recover the human choice. In the table, the coverage rate refers to the ratio of the number of sentences agreed on by the system and the judge to the total number (here is 6). It could be seen as a measure of accuracy. For all three drugs, our system would acquire a relatively high accuracy on average (66.7%, 55.6%, 66.7%)
Table 5.4: The outcome results for Vioxx with expert comments

<table>
<thead>
<tr>
<th>ID</th>
<th>Keywords</th>
<th>Expert comment</th>
<th>Similar Opinions</th>
<th>Supplementary Opinions</th>
</tr>
</thead>
</table>
| 1  | reduce inflammation, pain             | Vioxx works by reducing substances that cause inflammation, pain, and fever in the body. Vioxx is used to reduce pain, inflammation | [s=632] hope everyone is not in too much pain this week i am still hoping this vioxx will work | [s=470] positive results that reduces inflammation  
[s=363] he said it is a new anti inflammatory similar to vioxx and it works with narcotic pain  
[s=417] i went back to the doctor today and he gave me vioxx to reduce the muscle spasms and inflammation |
| 2  | doctor directed                       | Do not take Vioxx without first talking to your doctor. Take Vioxx exactly as directed by your doctor. | [s=816] i called the doctor and he said the vioxx would not make you feel this way | [s=470] it was a lot easier to talk to my doctor than i thougt it would be  
[s=477] i was using it on a regular basis till i had my follow up labs taken my liver function |
| 3  | taken dose, water, shake              | Take each dose with a full glass of water. Shake the Vioxx suspension well before measuring a dose. | [s=701] you will have to consume approximately the same amount of vioxx each day and your doctor may need to adjust | [s=430] during the worse period i had my doctor even had me on a daily dose cloriquin  
[s=478] my gi doc told me it was ok to take 2000mg daily |
| 4  | no restrictions                       | There are no restrictions on food, beverages, or activity during treatment with Vioxx. | [s=590] as for those medicines that shouldn’t be taken with food, this one is different. | [s=548] i can eat very little and some of the foods i can eat  
[s=425] i kept a diligent journal of all my side effects and made sure i was seen by one of them |
| 5  | allergic reaction                     | an allergic reaction (difficulty breathing; closing of your throat; swelling of your lips, tongue, or face; or hives) | [s=520] i get bad reaction to but pain is worsening and have to find what can help me | [s=434] no horrible reaction  
[s=479] adults who get adverse reactions and demented from hep |
| 6  | heart, nausea                         | have congestive heart failure and heart disease, nausea, heartburn                | [s=526] i was one of the unfortunate people given double doses of vioxx for my as and ended up with heart damage | [s=410] in my experience i have found that pharmaceutical drugs are too pure and concentrated causing many unbearable side effects such as nausea  
[s=448] the nausea and diarrhea from the colchicine was much worse for me than from the combo |
| 7  | pain, discomfort                      | abdominal pain, tenderness, or discomfort                                       | [s=702] i have the redness and inflammation but not constant pain and tenderness | [s=359] the ulnar pain that you mention that you have does it feel like when you have hit your funny bone  
[s=424] i also get alot of numb tingly achy pain up the sides of my  
[s=356] when i explained the probs about my spine he palpated all my vertebrae and agreed there was definite tenderness at the two spots |
| 8  | liver, kidney disease                 | have liver disease, have kidney disease                                          | [s=789] even if i didn’t have a liver disease i shouldn’t take any other | [s=383] when a person suffers nerve damage a doctor is supposed to look for a cause  
[s=365] the risk that you should most be concerned with is the possibility of liver problems which should be monitored by the doctor  
[s=411] in april i had to discontinue my methotrexate due to elevated liver enzymes |
| 9  | aspirin, affect                       | aspirin or or an aspirin-like medication such as salsalate will affect Vioxx     | [s=420] that has touched my pain besides aspirin which thinned my blood way too much | [s=482] vitamins i take contain 400mcg of folate or folic acid  
[s=494] c and a good multivitamin will help you reach your goal sooner rather than later |
Table 5.5: The extra outcomes for Vioxx without expert comments

<table>
<thead>
<tr>
<th>ID</th>
<th>Extra Opinions</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>[s=1337] I moved on to some prescriptions but those have been pulled off the market vioxx and bextra</td>
</tr>
<tr>
<td>11</td>
<td>[s=1377] The vioxx scandal has already taught us how ruthless the pharmaceuticals can be in pursuit of their profits</td>
</tr>
<tr>
<td>12</td>
<td>[s=1703] After all when the commotion about vioxx was going on I didn’t give it too much thought either because I wasn’t feeling it myself</td>
</tr>
</tbody>
</table>

Table 5.6: The outcome results for Tysabri with expert comments

<table>
<thead>
<tr>
<th>ID</th>
<th>Keywords</th>
<th>Expert comment</th>
<th>Similar Opinions</th>
<th>Supplementary Opinions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>immune system</td>
<td>Tysabri is a monoclonal antibody that affects the actions of the body’s immune system. Monoclonal antibodies are made to target and destroy only certain cells in the body.</td>
<td>[s=71] They would like to kill the whole thing but the immune system is pretty resilient</td>
<td>[s=36] It sounds the one unfortunate soul was allergic to one of the antibodies used and died after an alternative proved too potent</td>
</tr>
<tr>
<td>2</td>
<td>doctor talk</td>
<td>Before taking Tysabri, talk with your doctor.</td>
<td>[s=53] As for comments from doctors of Tysabri I suggest you contact Dr</td>
<td>[s=32] They offered to put me up for the night but I couldn’t stay</td>
</tr>
<tr>
<td>3</td>
<td>infection, death, risk, brain</td>
<td>Tysabri may cause a serious viral infection of the brain that can lead to disability or death. This risk is higher if you have a weak immune system or are receiving certain medicines.</td>
<td>[s=49] New cases of brain infection linked to Tysabri</td>
<td>[s=26] My neuro called tonight and we had an extensive conservation about PML Tysabri</td>
</tr>
<tr>
<td>4</td>
<td>injection, clinic</td>
<td>Tysabri is injected into a vein through an IV. You will receive this injection in a clinic or hospital setting every 4 weeks.</td>
<td>[s=63] I know that some weeks go by where I never miss an injection despite potential obstacles</td>
<td>[s=31] And the study confirm the safety profile from previous clinical studies of Tysabri</td>
</tr>
<tr>
<td>5</td>
<td>headache, depression</td>
<td>Headache, tired feeling, depression</td>
<td>[s=61] Side effects headache around 1 to 2 hours after but gone later that night</td>
<td>[s=43] These fractures may go undiagnosed for years</td>
</tr>
<tr>
<td>6</td>
<td>joint pain</td>
<td>Joint or muscle pain; stomach pain, diarrhea</td>
<td>[s=42] Many people have fractures and it is a leading cause of low back and leg pain</td>
<td>[s=30] Has Tysabri eliminated her MS and all its damage</td>
</tr>
</tbody>
</table>
Table 5.7: The extra outcomes for Tysabri without expert comments

<table>
<thead>
<tr>
<th>ID</th>
<th>Extra Opinions</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>[s=146] the agency’s removal of clinical hold could pave the way for re entry of</td>
</tr>
<tr>
<td></td>
<td>Tysabri to the marketplace</td>
</tr>
<tr>
<td>8</td>
<td>[s=126] it seems to me that Tysabri needs to achieve to not lose US money is</td>
</tr>
<tr>
<td></td>
<td>11-14 market share</td>
</tr>
<tr>
<td>9</td>
<td>[s=165] access to Tysabri is currently limited to ongoing clinical trials due</td>
</tr>
<tr>
<td></td>
<td>to product availability</td>
</tr>
</tbody>
</table>

Table 5.8: Task 1 result

<table>
<thead>
<tr>
<th>Coverage rate</th>
<th>Meridia</th>
<th>Vioxx</th>
<th>Tysabri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge1</td>
<td>4/6=66.7%</td>
<td>4/6=66.7%</td>
<td>4/6=66.7%</td>
</tr>
<tr>
<td>Judge2</td>
<td>4/6=66.7%</td>
<td>3/6=50%</td>
<td>4/6=66.7%</td>
</tr>
<tr>
<td>Judge3</td>
<td>4/6=66.7%</td>
<td>3/6=50%</td>
<td>4/6=66.7%</td>
</tr>
<tr>
<td>Average</td>
<td>66.7%</td>
<td>55.6%</td>
<td>66.7%</td>
</tr>
</tbody>
</table>

5.4.2 Task 2. Classifying the Outcomes with Expert Comments

In the second task, we aim to evaluate the effectiveness of our system in grouping sentences into \( m \) outcome clusters with prior knowledge. For each drug, we still choose 2 sentences from each of the system-generated outcome clusters (only the clusters with prior knowledge). All the judges are asked to assign each sentence to one of the \( m \) groups (for Meridia and Vioxx, \( m = 9 \), for Tysabri, \( m = 6 \)).

Table 5.9: Task 2 result

<table>
<thead>
<tr>
<th>Coverage rate</th>
<th>Meridia</th>
<th>Vioxx</th>
<th>Tysabri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge1</td>
<td>12/18=66.7%</td>
<td>13/18=72.2%</td>
<td>9/12=75%</td>
</tr>
<tr>
<td>Judge2</td>
<td>15/18=83.3%</td>
<td>10/18=55.6%</td>
<td>10/12=83.3%</td>
</tr>
<tr>
<td>Judge3</td>
<td>16/18=88.9%</td>
<td>13/18=72.2%</td>
<td>12/12=100%</td>
</tr>
<tr>
<td>Average</td>
<td>79.6%</td>
<td>66.7%</td>
<td>86.1%</td>
</tr>
<tr>
<td>Agreement rate</td>
<td>9/18=50%</td>
<td>10/18=55.6%</td>
<td>8/12=66.7%</td>
</tr>
</tbody>
</table>
The results are shown in Table 5.9. From the last row, we know that three judges agree on
50%, 55.6%, 66.7% of the sentences about the class labels, respectively, which indicates that
the classification is a subjective work – the results are controversial even among human judges.

Even such a subjective task, our system shows a high match with the human opinions. (e.g.,
the coverage rate of Tysabri is 86.1%).

5.4.3 Task 3. Distinguishing Similar Opinions and Supplementary
Opinions

The third task tries to explore how well we can separate the similar opinions from the supple-
mentary ones. For each drug, we choose 4 outcome clusters and for each cluster $i$, we choose
1 sentence from $O_{i, sim}$, mixed with 2 sentences from $O_{i, sup}$. The judges are asked to pick up
the most similar one with the expert comment from three.

Table 5.10: Task 3 result

<table>
<thead>
<tr>
<th>Coverage rate</th>
<th>Meridia</th>
<th>Vioxx</th>
<th>Tysabri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge1</td>
<td>4/4=100%</td>
<td>3/4=75%</td>
<td>3/4=75%</td>
</tr>
<tr>
<td>Judge2</td>
<td>4/4=100%</td>
<td>3/4=75%</td>
<td>3/4=75%</td>
</tr>
<tr>
<td>Judge3</td>
<td>4/4=100%</td>
<td>2/4=50%</td>
<td>3/4=100%</td>
</tr>
<tr>
<td>Average</td>
<td>100%</td>
<td>66.7%</td>
<td>75%</td>
</tr>
</tbody>
</table>

The results are shown in Table 5.10. The high accuracy shows that our method can suc-
cessfully recognize the semantic difference among the sentences in one outcome cluster.

5.4.4 Limitation of PLSA

From the results we observe that Vioxx always has the lowest accuracy among three. A poten-
tial reason would be its oversize clusters may affect the quantity. PLSA model requires to set
the number of clusters manually and lacks a way to dynamically determine the proper number
of clusters. We would try to solve the problem in the future work.
5.5 Labeling Result

Following the methods described in Section 4.3, for three drugs, we automatically generate top 3 semantic labels for each similar opinion $O_{i, \text{sim}}$, as well as the extra outcome without expert comments, Table 5.11 shows the overall results. (Extra outcomes: OID 10 − 12 for Meridia/Vioxx, 7 − 9 for Tysabri). We can see most of the labels are not only understandable but also reflect the meaning of the corresponding cluster. And we also find that the labels for the extra outcomes either discuss the extra topic, or does not make sense since too scattered.
Table 5.11: Label results

<table>
<thead>
<tr>
<th>OID</th>
<th>Top3 labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>lost pounds, 4 lbs, 8 pounds</td>
</tr>
<tr>
<td>2</td>
<td>private doctor, toxicity ones, the prescription</td>
</tr>
<tr>
<td>3</td>
<td>x1 daily, units daily, picolinate 500mg</td>
</tr>
<tr>
<td>4</td>
<td>seven years, point not, 2 years</td>
</tr>
<tr>
<td>5</td>
<td>suit against, minimally impact, alcohol allergic</td>
</tr>
<tr>
<td>6</td>
<td>bloody pressure, consistent normal, tiredness dizziness</td>
</tr>
<tr>
<td>7</td>
<td>caused heart, beat fast, developing ischemic</td>
</tr>
<tr>
<td>8</td>
<td>arms back, relief pain, joint pain</td>
</tr>
<tr>
<td>9</td>
<td>price high, worth of, size 1</td>
</tr>
<tr>
<td>10</td>
<td>willing to, support good, experience them</td>
</tr>
<tr>
<td>11</td>
<td>started new, go after, worried about</td>
</tr>
<tr>
<td>12</td>
<td>loving it, stimulant treatment, few questions</td>
</tr>
</tbody>
</table>

(a). Meridia

<table>
<thead>
<tr>
<th>OID</th>
<th>Top3 labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>reducing inflammation, probably caused, pain reduced</td>
</tr>
<tr>
<td>2</td>
<td>refer him, visit doctor, ask ahead</td>
</tr>
<tr>
<td>3</td>
<td>400mg tablet, msm 400mg, take aggressively</td>
</tr>
<tr>
<td>4</td>
<td>cooked food, meat fresh, contributing factor</td>
</tr>
<tr>
<td>5</td>
<td>arthus reaction, supportive help, bad reaction</td>
</tr>
<tr>
<td>6</td>
<td>heart problem, anixety attack, physiological treatment</td>
</tr>
<tr>
<td>7</td>
<td>abdominal pain, neuropathic pain, classic symptom</td>
</tr>
<tr>
<td>8</td>
<td>liver problem, not taken, meds pain</td>
</tr>
<tr>
<td>9</td>
<td>besides aspirin, cure yourself, proton pump</td>
</tr>
<tr>
<td>10</td>
<td>out market, against vioxx, market bextra</td>
</tr>
<tr>
<td>11</td>
<td>patients from, training me, the tremors</td>
</tr>
<tr>
<td>12</td>
<td>not interested, surgery you, treatment hope</td>
</tr>
</tbody>
</table>

(b). Vioxx

<table>
<thead>
<tr>
<th>OID</th>
<th>Top3 labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>flu symptoms, weakened immune, body enervation</td>
</tr>
<tr>
<td>2</td>
<td>doctor call, doctor prescription, matter process</td>
</tr>
<tr>
<td>3</td>
<td>calculated risk, greater risk, sudden death</td>
</tr>
<tr>
<td>4</td>
<td>intramuscular injection, dominant medicine, receive training</td>
</tr>
<tr>
<td>5</td>
<td>heavy feeling, extremely tired, muscle ache</td>
</tr>
<tr>
<td>6</td>
<td>instant pain, arm pain, spasticity problems</td>
</tr>
<tr>
<td>7</td>
<td>prior review, get approved, rule breaker</td>
</tr>
<tr>
<td>8</td>
<td>los angeles, senior citizens, transgression although</td>
</tr>
<tr>
<td>9</td>
<td>undisclosed november, appointment september, august 9</td>
</tr>
</tbody>
</table>

(c). Tysabri

25
CHAPTER 6

Conclusion

In this paper, we build a useful system to filter, integrate and label drug-based medical information. SVM-classifier, semi-supervised PLSA model and automatic labeling techniques are applied in the system. The experiment results with high accuracy show that our system can successfully reorganize the online medical information in a readable and understandable way.
CHAPTER 7

Future Work

In the future, we will focusing on the following potential directions:

1. In this paper, we take three drugs: Meridia, Vioxx, Tysabri as examples since they are the FDA-withdrawn drugs with serious side efforts. In this case, the task is relatively easy – the user comments would talk a lot about the side efforts which can be distinguished as outcomes from all the sentences. When it comes to the routine drugs, or even treatments, the clusters would be less discriminative. We will try build a web-based comprehensive system to show the outcomes and detailed opinions of each drug (withdrawn, routine) and treatment.

2. Other online sources also contain much information about personal health messages, e.g., Twitter \(^1\). Twitter contains not only many users’ comments but also the network relationships. We can apply network analysis on the current system and develop new topics: cluster content evolving analysis; content-based popularity analysis, etc.

3. \textit{PECO} pattern mining from the extracted results. This would be helpful to label \textit{PECO} sentences from the medical literatures and informal messages automatically.

4. As Section 5.4 discussed, PLSA model requires to set the number of clusters manually and lacks a way to dynamically determine the proper number of clusters. We would focus on how to automatically decide the number of the semantic clusters instead of a

\(^1\)http://twitter.com/
fixed number. A combination of multiple clustering methods can be applied to achieve such goal.

5. In this paper, we utilize a straightforward labeling method which lacks the semantic analysis. In the future, we would enhance the labeling method by enroll first-order relevance [13], which would capture the semantic meaning better. Also good labels would offer more accurate prior knowledge to create new clusters for new outcomes, which are not specified by an expert, but discovered automatically by the system.

6. Speech-based reorganization and detection. In Theorell’s work [19], mobile device can be used to record health diaries by patients’ voice. How to transfer speech into text, recognize and segment the useful information are all challenging and promising topics.

Based on the current progress and plan, we could build solid work to extend this topic and publish some qualified conference/journal papers. It can be a promising PhD thesis in the future.
REFERENCES


