Recommending alternative drugs by using generic drug names to minimise side effects

Sohee Hwang, Jungrim Kim,
Jeongwoo Kim and Sanghyun Park*

Department of Computer Science,
Yonsei University,
Seoul, Korea
Email: Soheeee2@yonsei.ac.kr
Email: kimgogo02@yonsei.ac.kr
Email: jwkim2013@yonsei.ac.kr
Email: sanghyun@yonsei.ac.kr
*Corresponding author

Abstract: Healthcare and the treatment of illnesses are one of the most fundamental aspects of modern human life, and drugs are the easiest approach to healthcare. However, consuming drugs lead to diverse effects. We propose the use generic medicine names and it is important to note that while drugs with the same generic name serve similar purposes, they may also cause different side effects. This paper presents a strategy to address the issue of side effects by recommending alternative drugs that have the same therapeutic effect but with less detrimental effects. By integrating the generic names of drugs and data from social networks, more data can be obtained to arrive at meaningful conclusions. This paper proposes a new approach for analysing drug-induced side effects, with collecting, processing, and using data from social networks.

Keywords: data mining; drug recommendation; adverse drug reaction; social data; side effect; generic name; drug-induced; user comment; alleviated side effect; alternative drug.

Reference to this paper should be made as follows: Hwang, S., Kim, J., Kim, J. and Park, S. (2017) 'Recommending alternative drugs by using generic drug names to minimise side effects', *Int. J. Data Mining and Bioinformatics*, Vol. 18, No. 4, pp.301–314.

Biographical notes: Sohee Hwang is MS candidate at Yonsei University supervised by Professor Sanghyun Park. Her research interests in bioinformatics and data mining algorithms with biomedical data to provide clinical resolutions.

Jungrim Kim received the BS degree in 2013, from Ajou University, Kyunggi, Republic of Korea. He is currently a PhD candidate in the Yonsei University. He has interests in development of data mining algorithms with large-scale biomedical data.

Jeongwoo Kim is currently a PhD candidate in the Yonsei University supervised by Prof. Sanghyun Park. He has interests in data mining and text-mining algorithms. His research goal is extracting meaningful knowledge from the bunch of data.

Sanghyun Park received his BS and MS degrees (Computer Engineering) from Seoul National University in 1989 and 1991, respectively, under the supervision of Prof. Sukho Lee. Then he received a PhD degree (Computer Science) from University of California at Los Angeles (UCLA) in 2001 under the supervision of Prof. Wesley W. Chu. His thesis title is Indexing Techniques for Similarity Searches in Sequence Databases. His research interest includes database, data mining and bioinformatics, and he is currently supervising the Data Engineering Lab in Yonsei University.

This article is a revised and expanded version of a paper entitled 'Recommend Alternative Drugs to Minimize Side-effect Using Generic Name of Drug' presented at the '6th International Conference on Emerging Databases: Technologies, Applications, and Theory', Jeju Island, Korea, 17–19 October 2016.

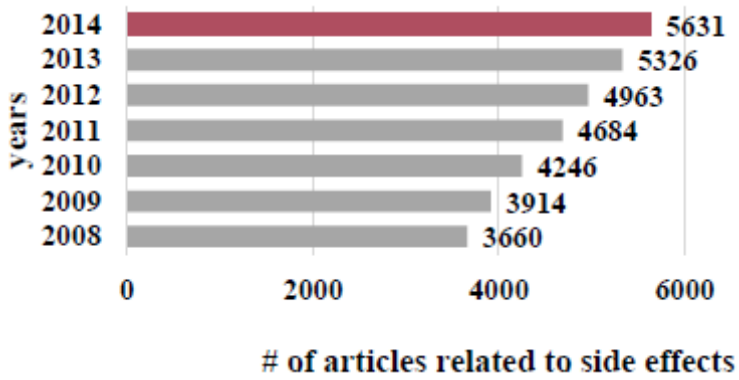
1 Introduction

The rapid growth of social networks has brought fundamental changes in modern society. We realise that the problem now is not of the lack of information, but rather the excess of it. A natural consequence is the issue of dealing with such abundant data, one example being biological data. Various studies have produced valuable results from analysing biological data which contributed to human life. The analysing biological data are named as Bioinformatics. Bioinformatics is the one major part of biotechnology revolution which extracted valuable information from analysing biological data.

Healthcare and the treatment of illnesses are one of the most fundamental aspects of modern human life, and drugs are the easiest approach to providing healthcare. Drugs reduce pain, cure diseases, and maintain health; however, consuming drugs leads to diverse effects. Furthermore, unfortunately, because of the high cost of drugs, not everyone can have access to the same benefits. We propose using the generic names of medicine, which identify drugs that contain similar active chemical ingredients, and, therefore, perform similar purposes, to address such hindrances. Thereby, people would have the opportunity to select preferred drugs that are suited to their needs. It is possible to identify the chemical, generic, and brand names of any given drug. Chemical names provide information about the chemical composition of a drug, generic names group drugs with similar purposes, and brand names are chosen by the drug manufacturers. For instance, Tylenol (n.d.) is a brand name, and the drug has the generic name acetaminophen. However, drugs do not necessarily have a single unique brand name. Different companies might have different brand names for the same basic drug with the same generic name. If two or more drugs have the same generic name, the first one developed or launched is usually called the original medicine, and the subsequently released forms are called generic medicines. Most physicians consider generic medicines to have the same effectiveness as the original medicines (Kersnik and Peklar, 2006; Shrank et al., 2011; Howland, 2010). However, although a generic medicine may contain similar active ingredients, its chemical formulation, or effect on the consumer may not be the same as the original formulation. In fact, some studies have demonstrated differences between various medicines with the same generic name from different manufacturers.

Therefore, generic medicines often have different effectiveness and side effects even when they contain the same active ingredients (Faasse et al., 2013).

Figure 1 Number of articles about side effects over 7 years (see online version for colours)



By integrating the drug characteristics and data from social networks, we can obtain more data to arrive at meaningful conclusions on specific drug effects. The rapid growth of social networks has fundamentally changed modern society. We realise that the current problem is not a lack of information but, rather, its excess. A natural consequence is the challenge of dealing with the overwhelmingly abundant data: one example being social network data. Numerous studies have produced valuable results by analysing social networks data, which have contributed to human life. One significant subject of studies is the analysis of drug data from social networks. The consumption of drugs by humans induces various effects such as curing diseases, reducing pain, and maintaining health. However, drugs can also cause side effects, which can also significantly harm the human body. Therefore, the side effects of drugs should be a topic of significant importance. As shown in Figure 1, the number of publications on drug side effect has increased over the past 7 years, illustrating the increased effort to analyse and address the problem.

Users of social networks voluntarily share their healthcare-related experiences. Drugs are a part of healthcare, and user reviews often contain the opinions or experiences of people after they have taken certain drugs. Some examples of healthcare group websites include Medications.com (Medications.com, n.d.), DailyStrength.org (DailyStrength, n.d.), drugs.com (Drugs.com, n.d.), webmd.com (webmd, n.d.), drugratingz.com (DrugRatingz, n.d.), and askpatients.com (Ask a Patient, n.d.). Furthermore, numerous literature reports deal with the healthcare information from social networks (Grajalesm, 2014). The paper by Sampathkumar et al. (2014) included comments from healthcare forums that mentioned adverse effects using hidden Markov models (HMM). An earlier study by Yang et al. (2013) combined text mining and supervised learning methods to extract messages, and then divided comments into positive and negative examples. Other studies have extracted the adverse effects of drugs and divided them into two groups (Benton et al., 2011; Sarker and Ginn, 2015; Nikfarjam and Gonzalez, 2011): expected and unexpected. Similarly, the paper by Sarker and Gonzalez (2015) performed feature selection with three binary classification approaches: Naïve Bayes (NB), Support Vector Machines (SVM) and Maximum Entropy (ME). Most publications aimed to propose a

highly accurate method of detecting and classifying side effects from user reviews utilising machine learning (Yates and Goharian, 2013; Laszlom et al., 1998; Chee et al., 2001; Jiang and Zheng, 2013; Ginn et al., 2014).

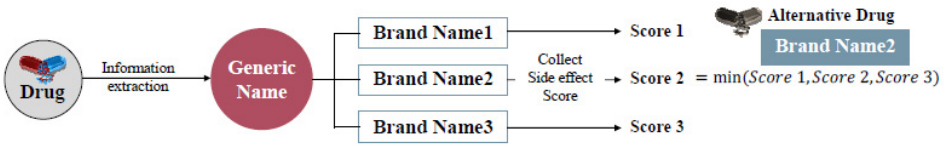
In this paper, we report our suggested strategy to minimise the side effects of drugs by using their characteristics. The severity of the side effects was searched in social networks associated with healthcare, and we used generic medicine names to identify drugs with similar actions but better performances and have fewer side effects. Based on our study results, we will recommend alternative drugs that have a better side-effect grade than others similar drugs did.

The paper is organised as follows. In Section 2, we presented the data sets collected from two healthcare groups and showed the methodology for identifying alternative drugs and examining the results. In Section 3, we evaluated the performance of the suggested method. The last section concludes with brief examples of contributions of our method and a proposal for future study directions.

2 Methods

In this section, we present our methodology for identifying alternative drugs and examining the results as follow: (1) collect data, (2) identify alternative drugs, (3) recommend alternative drugs, and (4) investigate the extracted data. An overview of our method is illustrated in Figure 1. In step 1, we used data from drugratingz.com to determine the drugs with a low rating for side effects, and, therefore, require alternative drugs. In step 2, we described the process used to identify alternative drugs. In step 3, we described the process used to select an appropriate alternative drug based on the data obtained in step 2. In our last step, we proceeded to examine our methods.

Figure 2 Example of recommending alternative drug



2.1 Data sets

Although there are numerous types of drugs developed for different purposes, this study focused on drugs for depression and anxiety disorders. Data from drugratingz.com suggest that depression and anxiety disorder drugs are one of the most widely used. In addition, the depression and anxiety disorder drugs had the most customer reviews, from which we drew data for our analysis.

We used five different data sets: (1) the list of drugs with ratings on side effects, (2) the list of generic names that identified each drug, (3) data of drugs with identical generic names, (4) rating score for the drugs, and (5) the number of citation of side-effect symptoms. Our first data set was extracted from 709 user reviews on drugratingz.com, three were from drugs.com, and the last data set was from treato.com.

Table 1 The details of data sets

	# of posts	*# of drugs*	*Sites*
Total drugs	5,959	1,815	Drugrationz.com
Depression	709	140	Drugratingz.com
Drug – Gname	–	60	Drugs.com
Gname – Bname	–	123	Drugs.com
Side effects	1,016,418	4	Treato.com

Details of first data set from drugratingz.com provide information on 1815 drugs classified into 38 categories including but not limited to cancer, addiction treatment, birth control, and pain relief. There were 140 drugs in the depression and anxiety disorder category, accounting for 7.71% of the drugs listed on drugratingz.com, which has the third highest number of all the categories. The core data analysed in this research were based on user ratings of drugs. Therefore, we ruled out drugs that had no reviews. In addition, we determined that drugs with a side-effect rating of more than 2.95 did not require alternative drugs and, thus, ruled them out as well.

The other three data sets were from drugs.com, which is an online pharmaceutical dictionary that offers drug information to users and healthcare professionals. From drugs.com, we obtained the generic and brands names of the drugs.

To demonstrate the results, we used the data set from Treato.com, which is a data analysis company known for providing health-associated insights to patients, healthcare professionals, pharmaceutical companies, and hospitals. Treato.com extracts the details of side-effect symptoms from user comments and shows the citation counts. For instance, two user reviews of Paxil shown in Figure 2 report the detailed side-effect symptom of 'withdrawal', which is cited thrice in user comments. From the search of all user comments, 'withdrawal' was cited 51,337 times. Thus, we used the information from treato.com as the measure of side effects.

2.2 Step 1: extraction of user-review data

In step 1, we extracted the list of drugs and their ratings with python programming from reviews written by users concerned with depression and anxiety disorders. The drug-related data were collected from drugratingz.com. There were four categories for rating drugs: effectiveness, no side effects, convenience, and values. The scores ranged from 1 to 5, where 1 indicated the worst and 5 the best ratings. Since the goal of this research was to reduce the side effects of drugs, we focused on the side-effects category. Moreover, we filtered out drugs with rating over 4, they are worthless to find alternative one. Thus, we stripped our data of unnecessary information, retaining only the names and side-effect ratings of drugs. Afterward, we calculated the average side-effect score for each drug and sorted them in ascending order.

Figure 3 The overview of proposed method

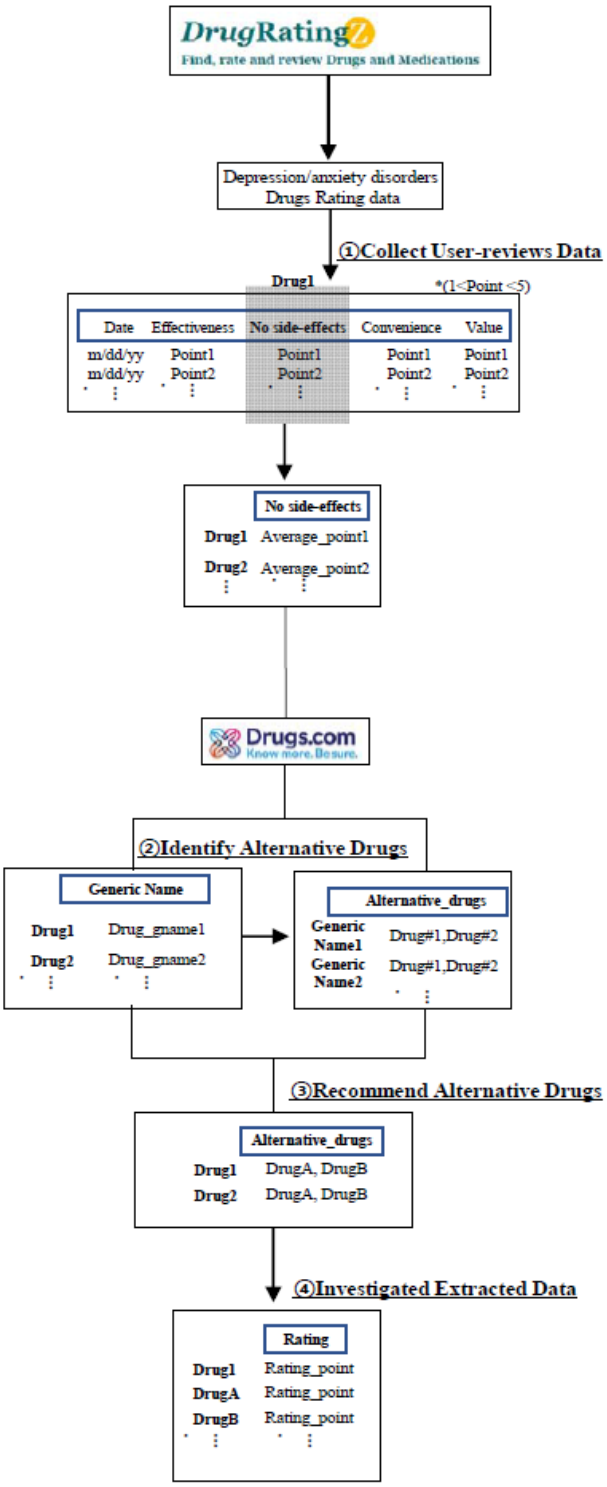
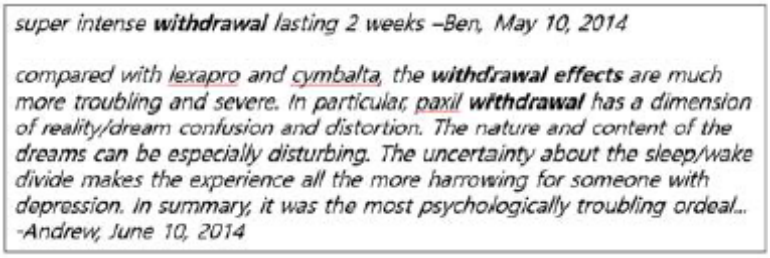


Figure 4 The example of user-comments (treato.com)



2.3 Step 2: identification of alternative drugs

In step 2, we discuss our methodology for identifying alternative drugs. It is important to note that all drugs have a generic name, which is the name of pharmaceutical drugs that are equivalent to a brand name in effects, dosage, intended use, quality, strength, and route of administration. Drugs do not necessarily have a unique brand name, however. Drugs sold by different companies might have different brand names, but it is possible they contain the same active drug with the same generic name. This makes it possible to find alternatives to a drug – if the generic names of brand drugs A and drug B are the same, drug A could replace drug B, and vice versa. Importantly, we already obtained data on the names and the average side-effect ratings of drugs in step 1. Based on the average ratings, we easily filtered out the drugs with high ratings (determined by a specific threshold) as safe drugs. To identify the generic name of each drug, we used the information from drugs.com, which presents not only the generic name of each drug but also a list of drugs with the same generic name. Our two-step process involved (1) identifying the generic name of a drug, and (2) obtaining a list of drugs with the same generic name.

2.4 Step 3: recommend alternative drugs

In step 2, we obtained data that allow us to identify drugs using their generic names. In this section, we report our selection of alternative drugs using the data on generic names. Similar to step 2, we used information from the data sets to obtain the brand names of drugs with a given generic name. For instance, the generic name of Effexor is venlafaxine, and we found two brand names with that generic name: Effexor and Effexor XR. Therefore, we determine that Effexor XR is a valid alternative candidate drug for Effexor. Some drugs have numerous brand names with the same generic name. For instance, Wellbutrin, whose generic name is bupropion, has ten other brand names. Using step 3, we were able to acquire a list of candidate alternative drugs based on the generic name matching process.

2.5 Step 4: investigation of extracted data

In step 4, we determined the validity of step 3. The data used in the analysis were from drugs.com, which provide an overall rating of each drug. It should be noted that the ratings from drugs.com, as opposed to those from drugratingz.com, do not provide

evaluations on multiple dimensions. That is, they do not have ratings specifically for the side effects of a drug. Nevertheless, we used the data from drugs.com because the ratings for some drugs were not available on drugratingz.com. The investigation was executed using the following process: (1) we obtained the rating scores of the original and alternative drugs and then (2) we compared each rating. We have confirmed that the alternative drugs are safer than the original drugs.

3 Results

In this section, we present the result of our proposed method. Since the goal of this research was to reduce the side effects of drugs, we tested whether our method successfully reduced the incidence of side effects. Ultimately, the output of our method would be a list of alternatives for each drug. Our analysis shows that 43.33% of drugs had an alternative, while the rest did not. That is, 26.66% of the drugs investigated had a one-on-one correspondence between the brand and generic names. Furthermore, 20% of the drugs that had alternatives were from drugratingz.com. The paper will proceed as follows: we will describe the data sets in 4.1, analyse alternative drugs from drugratingz.com in 4.2, and analyse the rest of the alternative drugs in 4.3.

3.1 Extracted data analysis

Using the process above, we acquired a list of recommendations for alternative drugs. The recommendation list was divided into two types, with one consisting of drugs mentioned in drugratingz.com and the other were drugs that were not cited there. In 3.2.1 and 3.2.2, we examined the first and second types of recommendations, respectively.

3.1.1 Drugratingz.com

To identify alternative drugs, we examined the generic name of each drug. The data on generic names from drugratingz.com were very limited, which inevitably led to some restrictions in the search for alternative drugs. This problem was addressed by using data from drugs.com, which provides comprehensive information on drugs and their generic names. Using the data from drugs.com, we were able to retrieve data that were more meaningful. As mentioned earlier, we ruled out drugs with a side-effect rating greater than 2.95, which was the average value of the side-effect rating and a high rating implies no alternative drug is required. Furthermore, 13% of the drugs were replaceable by alternatives from the drugs available on drugratingz.com. Comparing the side-effect ratings of the original and recommended alternative drugs, we observed that the rating of the alternative drugs was higher. For instance, mirtazapine was the drug with the lowest side-effects score of 1. We obtained information about the generic name of mirtazapine from drugs.com and based this information we were able to identify Remeron as an alternative drug that had the same generic name as mirtazapine. The side-effects score of Remeron was 2.23, which was much higher than that of mirtazapine. Details of our analysis results are presented in Table 2, which shows the side-effects scores of the original and alternative drugs. As shown in Table 2, alternative drugs effectively reduced the side effects.

Table 2 Result of detected alternative drugs (drugratingz.com)

Drugs	*Score of side effects*	*Alternative drugs*	*Score of side effects*
Mirtazapine	1	Remeron	2.23
Paxil	1.97	Pexeva	4.5
Alprazolam	2	Xanax	3.4
Effexor	2.17	Effexor XR	2.77
Wellbutrin	2.225	Bupropion HCI Tablet	3
Amitriptyline	2.33	Elavil	3.25

3.1.2 Drugs.com

From Section 3.1.1, we noticed that if the alternative drug was cited in the drugratingz.com database, we could use the side-effect score to compare the resulting side-effect scores. However, not all drugs were listed on drugratingz.com. For such exceptional cases, we decided to analyse the ratings of alternative drugs from drugs.com. We collected the ratings of the original and alternative drugs from drugs.com. For instance, the side-effect score of Buspar was 1, which is lower than our set threshold of 2.95. Based on the generic name identification process, we found that Vanspar is a valid alternative drug to Buspar. To examine the effectiveness of this alternative, we collected and compared the ratings of Buspar and Vanspar from drugs.com. Since the rating for Vanspar was higher than that for Buspar, we concluded that we successfully found a better alternative drug. Details of our results are presented in Table 3. The table presents reactions of drugs whose information was not available from drugratingz.com. We observed that the results were significant because we succeeded in identifying an alternative with a higher rating.

Table 3 Detected alternative drugs (drugs.com)

Drugs	*Score of side effects*	*Rate in drugs.com*	*Alternative drugs*	*Rate in drugs.com*
Buspar	1	6.5	Vanspar	6.7
Topamax	1.83	6.7	Topiragen	9
Bupropion HCI Tablet	2	7.6	Zyban	9.1
Dexedrine	2	8.7	Dextrostat	9.2
Abilify	2.11	6.6	Abilify Maintena	7.3
Dexadrine	2.66	8.7	Dextrostat	9.2
Prozac	2.95	7.7	Sarafem	8.8

3.2 Evaluation

From the additional information from treato.com, we verified the accuracy of the extracted data. Treato automatically collects, indexes, and analyses the significant amount of content patients and caregivers generate online to extract relevant information for a broader understanding of individual treatment regimens and condition-related experiences. It is the world's largest source of patient insights gathered from online conversations across the social web. Treato covers all healthcare forums and provides

details of side-effect scores, obtained from user comments, and counts the number of symptoms. For acetaminophen, 869 different concerns and 387,368 posts are present; the five highest counts for symptoms were addiction, liver damage, drowsiness, withdrawal symptoms, and dry mouth. To determine the degree of influence of side effects, we measured the effective evaluation score:

$$evaluation\_score = \sum_{i=1}^{20} \frac{\#\ of\ i^{th}\ citations}{\#\ of\ comments}$$

$*is\ the\ ranking\ of\ citation\ count$

The numerator denotes the number of relevant symptoms and the denominator denotes the total number of comments on one drug. Therefore, the evaluation score presents the degree of side effects reported in user reviews.

Table 4 Side-effects ranking (Paxil vs. Pexeva)

Paxil				*Pexeva*			
Rank	*Symptoms*	*No. of comments*	*Score*	*Rank*	*Symptoms*	*No. of comments*	*Score*
1	Withdrawal symptoms	51,351	0.1599	1	Withdrawal symptoms	97	0.1026455
2	Weight gain	21,226	0.0661	2	Weight gain	55	0.0582010
3	Dizziness	8576	0.0267	3	Nausea	30	0.0317460
4	Nausea	7994	0.0248	4	Addiction	15	0.0158730
5	Wight loss	6719	0.0209	5	Sexual dysfunction	15	0.0158730
6	Sexual dysfunction	5300	0.0165	6	Head zaps	12	0.0126984
7	Numbness	4395	0.0136	7	Dizziness	11	0.0116402
8	Suicide	4114	0.0128	8	Drowsiness	11	0.0116402
9	Head zaps	2888	0.0089	9	Numbness	11	0.0116402
10	Suicidal thoughts	2870	0.0081	10	Attempted suicide	9	0.0095238
1–20	Sum score		0.4240	1–20	Sum score		0.348

Pexeva was recommended as an alternative drug to paxil. According to drugratingz.com, the side-effect score of pexeva is much higher than that of paxil. To verify the accuracy of the alternative drug choice, we used the evaluation scores of paxil and pexeva (0.3450 and 0.2906, respectively). We applied 319,614 posts for paxil and 945 posts for pexeva. The scores implied that the number of people who suffered side effects was higher in the paxil group than in the pexeva group. Therefore, the results demonstrated that pexeva was an effective alternative drug for paxil. Moreover, as shown in Table 4, suicide accounted for 2% of deaths among individuals who consumed paxil (6984 posts) but 0.9% among those who consumed pexeva (nine posts). Thus, collectively, pexeva had fewer adverse effects than paxil did. Xanax was recommended instead of alprazolam. According to drugrationgz.com, the side-effect score of xanax was considerably higher than that of alprazolam. The evaluation scores of xanax and alprazolam were 0.1133 and 0.0620, respectively. We applied 52,293 posts for alprazolam and 659,239 posts for

xanax. From these scores, it proved that our proposed method could extract the correct value. Furthermore, this draws a meaningful conclusion, because of extracted rating data from drugs.com. The total ratings of alprazolam and xanax were the same in public website (9.0) even though the individual rating for the side effects differed. Proposed method derived evaluation about particular part as side effect. Remeron was proposed as an alternative drug to mirtazapine. As observed in a previous evaluation, remeron was found to be an effective alternative for Mirtazapine. There are 85,543 discussions on the side effects of remeron and 62,985 discussions on the side effects of mirtazapine.

Table 5 Side-effect ranking (Alprazolam vs. Xanax)

Alprazolam				*Xanax*			
Rank	*Symptoms*	*No. of comments*	*Score*	*Rank*	*Symptoms*	*No. of comments*	*Score*
1	Euphoria	1615	0.03088367	1	Weight gain	8704	0.0134
2	Irritability	829	0.01585298	2	Weakness	7621	0.0117
3	Weakness	647	0.01237259	3	Euphoria	6587	0.0101
4	Blisters	396	0.00757271	4	Weight loss	3549	0.0054
5	Psychosis	257	0.00491461	5	Hallucinations	2713	0.0042
6	Hallucinations	249	0.00476163	6	Attempted suicide	2272	0.0035
7	Weight gain	228	0.00436004	7	Dry mouth	1131	0.0017
8	Migraines	174	0.00332740	8	Impotence	1025	0.0015
9	Weight loss	162	0.00309792	9	Blisters	961	0.0014
10	Nightmares	121	0.00231388	10	Chronic fatigue	823	0.0012
1–20	Sum score		0.1133	1–20	Sum score		0.062

Table 6 Side-effect ranking (Remeron vs. Mirtazapine)

Remeron				*Mirtazapine*			
Rank	*Symptom*	*No. of comments*	*Score*	*Rank*	*Symptom*	*No. of comments*	*Score*
1	Weight gain	10,320	0.120642491	1	Weight gain	6478	0.102849885
2	Withdrawal symptoms	5929	0.069310982	2	Withdrawal symptoms	4685	0.07438279
3	Drowsiness	2212	0.025858643	3	Drowsiness	1939	0.030785108
4	Nausea	1987	0.023228356	4	Nausea	1685	0.026752401
5	Sleepy	1879	0.021965818	5	Weight loss	1277	0.020274669
6	Weight loss	1756	0.020527928	6	Sleepy	1200	0.019052155
7	Increased appetite	1459	0.017055949	7	Increased appetite	1143	0.018147178
8	Dizziness	1220	0.014262	8	Dizziness	1086	0.017242201
9	Grogginess	1165	0.013619041	9	Sexual dysfunction	990	0.015718028
10	Tinnitus	878	0.010263964	10	Grogginess	722	0.011463047
1–20	Sum score		0.405321363	1–20	Sum score		0.423418274

The evaluation scores of remeron and mirtazapine were approximately 0.40 and 0.42, respectively, which proved that mirtazapine can be replaced with Remeron.

In general, the ranking of a drug measures its overall effect. Therefore, it is difficult to obtain isolated information about a drug's side effect. The proposed method successfully provided data on a specific dimension of a drug represented by the variable evaluation score and proved that the effectiveness of the alternative drugs was above that demonstrated clinically.

4 Conclusion

We have designed a method for recommending alternative drugs to minimise side effects. We obtained the data by extracting useful information from user reviews. Our method applies to drugs in other categories as well. Although the scope of this study was limited to drugs concerned with depression and anxiety disorders, a similar approach could be taken for identifying alternative drugs in any category, as long as there are a sufficient number of user reviews.

We have demonstrated a strategy for minimising the side effects of drugs used for depression and anxiety disorders. We combined the information from two sources: drugratingz.com, which has information on drugs and their side-effect scores and drugs.com with information on the drug generic and brand names. The identification of alternative drugs would not have been possible if any of the information above were missing. In general, the ratings of drugs only measured the overall effect, and information on specific dimensions was usually not available. Therefore, it was difficult to obtain ratings that specifically measured the side effects. The proposed method successfully provides data on a specific dimension of a drug, which is represented by the variable evaluation score. For instance, Pexeva was presented as a recommended alternative for Paxil, and we verified that the evaluation score of Pexeva was higher than that of Paxil. Furthermore, the results revealed Xanax as an alternative drug for Alprazolam, although the overall rating scores for both drugs were equal. Even with an equal overall score, our method made it possible to distinguish between the two drugs by comparing the side-effect values, which were different.

In future studies, we will expand our method to all drug classes and determine how to set the thresholds of the side-effects score. Furthermore, we will ensure that or study design reflects all categories of rating scores from drugs.com.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2015R1A2A1A05001845).

References

Ask a Patient (n.d.) Available online at: http://www.askapatient.com

Benton, A., Ungar, L., Hill, S., Hennessy, S., Mao, J., Chung, A., Leonard, C.E. and Holmes, J.H. (2011) 'Identifying potential adverse effects using the web: a new approach to medical hypothesis generation', *Journal of Biomedical Informatics*, Vol. 44, pp.989–996.

Chee, B.W., Berlin, R. and Schartz, B. (2011) 'Predicting adverse drug events from personal health messages', *Proceedings of the American Medical Informatics Association (AMIA)*, Washington, DC, pp.217–226.

DailyStrength (n.d.) *Online Support Groups and Forums at DailyStrength*. Available online at: http://www.dailystrength.org

DrugRatingz (n.d.) *DrugRatingz: Find Rate and Review Drugs and Medications*. Available online at: http://www.drugratingz.com

Drugs.com (n.d.) *Drugs.com: Know More. Be Sure*. Available online at: http://www.drugs.com

Faasse, K., Cundy, T., Gamble, G. and Petrie, K. (2013) 'The effect of an apparent change to a branded or generic medication on drug effectiveness and side effects', *Psychosomatic Medicine*, Vol. 75, No. 1, pp.90–96.

Ginn, R., Pimpalkhute, P., Nikfarjam, A., Patki, A., O'Connor, K. and Sarker, A. (2014) 'Mining Twitter for adverse drug reaction mentions: a corpus and classification benchmark', *Proceeding of the Fourth Workshop on Building and Evaluating Resources for Health and Biomedical Text Processing*. Available online at: http://www.nactem.ac.uk/biotxtm2014/papers/Ginnetal.pdf

Grajalesm, F. (2014) 'Social media: a review and tutorial of applications in medicine and health care', *Journal of Medical Internet Research*, Vol. 16, No. 2, pp.1–23.

Howland, R.H. (2010) 'Are generic medications safe and effective?', *Journal of Psychological Nursing and Mental Health Services*, Vol. 48, pp.13–16.

Jiang, K. and Zheng, Y. (2013) 'Mining twitter data for potential drug effects', *International Conference on Advanced Data Mining and Applications*, Springer Berlin Heidelberg, Berlin, pp.434–443.

Kersnik, J. and Peklar, J. (2006) 'Attitudes of Slovene general practitioners towards generic drug prescribing and comparison with international studies', *Journal of Clinical Pharmacy and Therapeutics*, Vol. 31, No. 6, pp.577–583.

Laszlom, A., Smit, S. and Jeremy, D. (1998) 'Low-dose venlafaxine treatment in panic disorder', *Psychopharmacology*, Vol. 34, No. 2, pp.207–209.

Medications.com (n.d.) *The Premier Community to Talk about Health*. Available online at: http://www.medications.com/

Nikfarjam, A. and Gonzalez, G.H. (2011) 'Pattern mining for extraction of mentions of adverse drug reactions from user comments', *AMIA Annual Symposium Proceedings*, Vol. 2011, p.1019.

Sampathkumar, H., Chen, X. and Luo, B. (2014) 'Mining adverse drug reactions from online healthcare forums using Hidden Markov Model', *BMC Medical Informatics and Decision Making*, doi:10.1186/1472-6947-14-91.

Sarker, A. and Ginn, R. (2015) 'Utilizing social media data for pharmacovigilance: a review', *Journal of Biomedical Informatics*, Vol. 54, pp.202–212.

Sarker, A. and Gonzalez, G. (2015) 'Portable automatic text classification for adverse drug reaction detection via multi-corpus training', *Journal of Biomedical Information*, Vol. 53, pp.196–207.

Shrank, W.H., Girdish, C., Brennan, T.A. and Choudhry, N.K. (2011) 'Physician perceptions about generic drugs', *Annals of Pharmacotherapy*, Vol. 45, No. 1, pp.31–38.

Tylenol (n.d.) Available online at: http://www.tylenol.co.kr/index.jsp

webmd (n.d.) Available online at: http://webmd.com

Yang, M., Wang, X. and Kiang, M. (2013) 'Identification of consumer adverse drug reaction messages on social media', *PACIS 2013 Proceedings*. Available online at: http://aisel.aisnet.org/pacis2013/193

Yates, A. and Goharian, N. (2013) 'ADRTrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites', *Advances in Information Retrieval*, Vol. 7814, pp.816–819.