HHS Public Access

Author manuscript

Proceedings (IEEE Int Conf Bioinformatics Biomed). Author manuscript; available in PMC
2018 January 03.

Published in final edited form as:

Proceedings (IEEE Int Conf Bioinformatics Biomed). 2017 November ; 2017: 1258–1261. doi:10.1109/

BIBM.2017.8217839.

Evaluating Automatic Methods to Extract Patients' Supplement Use from Clinical Reports

Yadan Fan,

Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA

Lu He, and

Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA

Rui Zhang

Institute for Health Informatics, and College of Pharmacy, University of Minnesota, Minneapolis, MN, USA

Abstract

The widespread prevalence of dietary supplements has drawn extensive attention due to the safety and efficacy issue. Clinical notes document a great amount of detailed information on dietary supplement usage, thus providing a rich source for clinical research on supplement safety surveillance. Identification the use status of dietary supplements is one of the initial steps for the ultimate goal of the supplement safety surveillance. In this study, we built rule-based and machine learning-based classifiers to automatically classify the use status of supplements into four categories: Continuing (C), Discontinued (D), Started (S), and Unclassified (U). In comparison to the machine learning classifier trained on the same datasets, the rule-based classifier showed a better performance with F-measure in the C, D, S, U status of 0.93, 0.98, 0.95, and 0.83, respectively. We further analyzed the errors generated by the rule-based classifier. The classifier can be potentially applied to extract supplement information from clinical notes for supporting research and clinical practice related to patient safety on supplement usage.

Keywords

Clinical Notes; Electronic Health records; Machine Learning; Natural Language Processing

I. Introduction

Electronic Health Records (EHR) has stored a great amount of patient-specific information, especially in clinical notes, which documents and aggregates the clinical data related to all aspects of healthcare, including medication, diagnosis, laboratory test results, and radiological results [1]. The widespread adoption of EHR has facilitated the secondary use of clinical notes in clinical research, clinical decision support, drug safety surveillance, etc [2]. However, the full benefits of EHR remain unrealized because of the detailed information buried in the notes by physicians during the patient encounter [3].

With the widespread prevalence of dietary supplements in the United States, the safety and efficacy of supplements have drawn extensive attention. It was estimated that around 23,000 emergency department visits in the United States were attributed to adverse events associated with dietary supplements [4]. Since dietary supplements are promoted and regulated as food by the Dietary Supplement Health and Education Act (DSHEA) in 1994 [5], they are subject to regulations which are different from prescription drugs, thus exempting from premarket approval by the U.S. Food and Drug Administration (FDA). The information regarding the adverse events associated with supplement use mainly comes from voluntary reporting by the physicians through post-market surveillance. Thus, the number of adverse events associated with supplements is underestimated. In addition, current supplement research is only limited to pharmaceutical studies which usually focus on few supplements. Therefore, a knowledge gap related to the supplement safety is ubiquitous among not only patients but also healthcare providers.

Supplement safety information, same as drugs, can be indirectly mined through different data sources, including biomedical literature, adverse event reports, or electronic health records (EHR). We have demonstrated the feasibility of using clinical notes to investigate the supplement research. Our prior study has investigated the term coverage of supplements in EHR and found out that some of the supplements mentioned in the unstructured clinical notes were not documented in the structured medication lists [6]. Additionally, we also found that some detailed information regarding the supplement use status, such as “starting” or “discontinuation”, is embedded in the notes. Clinical notes might store supplement usage information that does not exist in structured data in EHR, which can be further extracted to complement medical lists for patient safety surveillance on supplements [7]. Through an extensive literature search, based on our knowledge, there were no studies investigating the extraction of supplement use information in clinical notes so far.

Identifying the historical and current status of dietary supplements use is a critical step for assessing the efficacy and adverse effects of dietary supplements. However, one of the challenges is how to accurately extract the detailed supplement use information in clinical notes for further use. In the clinical domain, rule-based and machine learning-based methods are two major approaches for extracting information from clinical narratives. In this study, we compare the performance of rule-based and machine learning-based classifiers in classifying use status of dietary supplements from the clinical narratives.

II. Methods and Materials

A. Study design

We collected sentences related to the 25 selected supplements from clinical notes. Experts manually reviewed all 1300 sentences to generate a gold standard, which further split into training and test sets. We trained both rule-based and machine learning-based classifiers on the training set and then compared the performance of classifiers in the test set. We analyzed the errors generated by the rule-based classifier. Details in each step are described below.

B. Dataset

Compared with the previous rule-based method [7], we expanded our dataset in this study by incorporating 300 sentences on 15 supplements (20 sentences for each of the 15 supplements), including bilberry, biotin, black cohosh, coenzyme Q10, cranberry, dandelion, flax seed, folic acid, glucosamine, glutamine, kava, lecithin, milk thistle, saw palmetto, and turmeric. We retrieved the notes containing the above-mentioned 15 supplements from the Clinical Data Repository (CDR) at the University of Minnesota using supplements names and their lexical variations. The list of the lexical variations was generated through manual chart review and suggested by a pharmacist. A total of 1000 sentences on 10 supplements (100 sentences for each of the 10 supplements), including alfalfa, Echinacea, fish oil, garlic, ginger, ginkgo, ginseng, melatonin, St John's Wort, and Vitamin E served as the training set (~77%). The dataset with another 15 supplements was used as a held-out test set (~23%) for comparing the performance of the rule-based and machine learning-based classifiers.

C. Annotations and gold standard

The annotation guideline was developed based on the consensus of two independent annotators [7]. In summary, "Continuing" status indicates that there is evidence showing the patients continue on the current supplements (e.g., "She continued on herbal supplements including echinacea"). "Discontinued" refers that the patient has stopped or is to stop the supplement for certain reasons such as allergic reactions or adverse interactions with prescribed medications (e.g., "Stopped taking her garlic tablets a week ago"). "Started" status indicates that the patient is being started on or will be started on the supplements (e.g., "Begin melatonin 10mg 1 hour before bedtime"). "Unclassified" status is related to supplement mentions in the clinical notes, which do not indicate the use status, such as patients' education, recommendation, or negation (e.g., "Denies using St. John's Wort"; "Advised over-the-counter melatonin"; "I recommend ginger for nausea"). Initially, two annotators reviewed the guideline and made a consensus on disagreement for a small sample. Then 100 sentences were randomly selected and independently annotated by the two annotators. The inter-annotator agreement on 100 sentences was measured as Cohen's Kappa of 0.83 and percentage agreement of 95%, indicating excellent agreement between annotators. In total, the gold standard extended to the 1300 annotated sentences.

D. Development, evaluation, and comparison of the classifiers

All the sentences were normalized using Lexical Variant Generation (LVG) [8]. The rule-based classifier was built on the training data using a series of semantic features indicating the use status of the supplement, such as "stop", "start", "begin", "advise", "discuss", "deny" and also the abbreviations for the indicator words like "d/ced" meaning "discontinued" for short. According to the previous rule-based study [7], it's significant to set the distance boundary between the indicator words and the supplement mentions. Several experiments were done to choose the optimal window size. The window size was selected as 8 tokens on the left side of the supplement mentions based on the results of our previous study. Compared with the previous study, in this study, we further considered the passive syntactic structure, where the use status indicator words are on the right side of the supplements when building the rules. For example, "melatonin has been stopped", "Vitamin

E is recommended”, “St John’s Wort restarted”. The rules were manually crafted on the training data until every record was covered by the rule. We assumed that if the record does not belong to any of the status, it fell into the “Unclassified” category. The rules were tested on the held-out test set on 300 sentences of 15 supplements for generalizability. Evaluation metric including precision, recall, and F-measure were used in the measurement. For each use status, commonly used rules were listed in Table 1.

In the previously reported machine learning study [9], we trained 7 different feature sets along with 5 classification algorithms: support vector machine (SVM), maximum entropy, Naïve Bayes, decision tree, and random forest. The result showed that the SVM with type 5 feature set (unigrams + bigrams + indicator words with distance) had the best performance. The indicator words include “increase”, “decrease”, “continue”, “resume”, “start”, “allergy”, etc. In the current study, we built the rules in the same training set and compared the performance of the rule-based classifier with above-mentioned machine learning classifier on the same test data.

III. Results and Discussion

A. Dataset

Among the training data, there were 380 sentences in the Continuing (C) status, 156 sentences in the Discontinued (D) status, 139 sentences in the Started (S) status, and 325 sentences in the Unclassified (U) status. In the test data, there were 112, 47, 74, and 67 sentences in the C, D, S and U status.

B. Performance of rule-based and machine learning based classifiers

Compared with the former rule-based study [7], additional 38 rules were generated on the extended dataset. These additional rules are mainly generated for sentences with passive syntax, namely, the indicator words are on the right side of the supplement mentions. Based on the training data, a total of 153 rules were generated, among which 45 were for C status, 21 for D status, 12 for S status, and 75 for U status. The performance of the rule-based classifier on the test dataset was shown in Table II. Overall, F-measures for use status of C, D, and S are over 0.90. However, the performance in the U category (F-measure: 0.84) is slightly worse. We assumed that if the record doesn’t belong to any of the categories, it should fall within the “unclassified” status. There might be some patterns we failed to find in the other three categories, leading to false positives under the “unclassified” category.

The optimal model in the machine learning study was SVM with features including unigrams, bigrams, and indication words with distance [9]. The performance of the machine learning model in the test dataset was shown in Table III. From the results, we can see that the performance of the machine learning method is slightly worse than the rule-based method, especially in D and U status. Comparing the performances of both classifiers on the test dataset, the rule-based classifier outperformed the learning classifier on the range of 0.02 to 0.11.

C. Error analysis

We performed an error analysis on the rule-based classifier in the test data. The main source of erroneous classification came from that the distance between the supplement mention and the indicator words was beyond the 8 tokens. For example, “She uses alternative medications for her lung cancer, including juniper, parsley, dandelion, and chamomile.” However, we conducted several experiments to expand the distance over 8 tokens, it would increase additional false positives. Thus, it is a tradeoff to choose the appropriate threshold window size to balance precision and recall.

The second erroneous source is due to the indicator issue. There might be more than one indicator word in the sentence. For example, “The patient also adds she has been taking milk thistle daily as an alternative supplementation for liver toxicity.” In this sentence, there are two indicators words “adds” and “taking” which imply two different use status “Started” and “Continuing”, respectively. The two indicators are both within a distance of 8 tokens with respect to the supplement mention. In this case, the ordering of the rules in terms of the four categories is significant for correct classification.

The third type of error is from some new patterns we failed to find in the training data while they occurred in the test data. For example, “she applied turmeric over her joints.”, “wonders about two supplements: Astaxanthin and biotin and whether it is safe.”, such patterns are not common in the training data. This is a common issue that the training set is not perfectly representative of the test set.

D. Comparing two classifiers

From the results we can see that the rule-based classifier outperformed the machine learning-based one in the test set. For the rule-based classifier, when the sample size of the training data was enlarged (by ~ 30%) compared with the previous study [7], more patterns (increase 33%) can be generated. The rule-based method is widely applied in the biomedical natural language processing systems. For example, NegEx applying two simple regular expression rules with a set of negated phrases negation was originally designed to detect negated findings or disease in discharge summaries, which is now extended to multiple languages and leveraged to extract negated information from clinical narratives [10]. The rule-based method can be easily implemented and modified. However, it’s a time-consuming iterative process to generate the rules. Also, it requires a sufficient and representative data set for generating the rules. The advantage is the rules were analyzed and generated by human which is more generalizable. With regard to the machine learning classifier, we only tested limited types of feature sets [9]. Also, the corpus for machine learning-based study is relatively small. The performance of the machine learning classifier is likely to be improved when more types of feature sets are tested and more training data is incorporated.

One of the advantages of our study is that we achieved a promising result using only 1300 sentences for both classifiers. We didn’t use much data and still got good results. However, in order to improve the generalizability of the rules, we are going to expand the dataset by incorporating more supplements in the future work. As we can see from both classifiers, the indicator words play a significant role in this classification task. In the future, we will

conduct additional linguistic analysis work to include more useful indicator words and refine the current list of the indicator words.

IV. Conclusion

In this study, we evaluated and compared a rule-based classifier and a machine learning based classifier on the identification of patients' supplement use status. We collected 1300 sentences for 25 dietary supplements. The gold standard was generated through manual review by experts. The rule-based classifier showed promising results on the test dataset with F-measure in C, D, S and U status of 0.93, 0.98, 0.96, and 0.84, respectively. Comparing with the machine learning-based classifier, the rule-based classifier performed slightly better. We further analyzed the errors on the rule-based classifier. Future work includes further expansion of gold standard and development of the hybrid system.

Acknowledgments

Research reported in this publication was supported by the National Institutes of Health, National Center for Complementary & Integrative Health Award (R01AT009457) (PI: Zhang), the University of Minnesota Clinical and Translational Science Award (8UL1TR000114) (PI: Blazer), and the University of Minnesota Grant-In-Aid award (PI: Zhang). The authors thank Fairview Health Services for data access support of this research.

References

1. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*. 2012 Jun 1; 13(6):395–405.
2. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *Journal of the American Medical Informatics Association*. 2011 Mar 1; 18(2):181–6. [PubMed: 21233086]
3. Patrick, Jon, Min, Li. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. *Journal of the American Medical Informatics Association*. 2010; 17.5:524–527. [PubMed: 20819856]
4. Geller AI, Shehab N, Weidle NJ, Lovegrove MC, Wolpert BJ, Timbo BB, Mozersky RP, Budnitz DS. Emergency department visits for adverse events related to dietary supplements. *N Engl J Med*. 2015 Oct 15; 2015(373):1531–40.
5. Radimer K, Bindewald B, Hughes J, Ervin B, Swanson C, Picciano MF. Dietary supplement use by US adults: data from the National Health and Nutrition Examination Survey, 1999–2000. *American journal of epidemiology*. 2004 Aug 15; 160(4):339–49. [PubMed: 15286019]
6. Zhang, Rui, et al. AMIA Annual Symposium Proceedings. Vol. 2015. American Medical Informatics Association; 2015. Evaluating term coverage of herbal and dietary supplements in electronic health records.
7. Fan, Y., He, L., Zhang, R. Bioinformatics and Biomedicine (BIBM), 2016 IEEE International Conference on 2016 Dec 15. IEEE; Classification of use status for dietary supplements in clinical notes; p. 1054-1061.
8. LVG. https://www.nlm.nih.gov/research/umls/new_users/online_learning/LEX_004.html
9. Fan Y, He L, Pakhomov SV, Melton GB, Zhang R. Classifying Supplement Use Status in Clinical Notes. *AMIA Summits on Translational Science Proceedings*. 2017:493.
10. Chapman, Wendy W., et al. Extending the NegEx lexicon for multiple languages. *Studies in health technology and informatics*. 2013; 192:677. [PubMed: 23920642]

Table I

Selected regular expressions and examples

Use Status Class	Frequently Used Regular Expressions	Selected Examples
Continuing (C)	$(\text{increase} \text{continue} \text{lower} \text{take} \text{use}) (\backslash s+\backslash S+)^{\{0,7\}}\backslash s+\text{sup}$	<ul style="list-style-type: none"> • She has increased her alfalfa tabs and this has eliminated her symptoms and chest tightness. • Continue fish oil to reduce inflammation. • She lowered her fish oil to 2 capsules daily without any noticeable affect on her joint pain. • She is also taking a Vitamin E supplement and Tylenol as needed for pain. • Reports she uses garlic for her high cholesterol.
Discontinued (D)	$(\text{stop} \text{discontinue} \text{hold} \text{allergy})(\backslash s+\backslash S+)^{\{0,7\}}\backslash s+\text{sup}$ $(\text{be} \text{stay})\backslash s+\text{off}(\backslash s+\backslash S+)^{\{0,7\}}\backslash s+\text{sup}$	<ul style="list-style-type: none"> • Stop Vitamin E supplement. • She is to discontinue her St. John's Wort. • You are already holding the fish oil and aspirin. • The patient has a drug allergy to sulfa and Echinacea. • She was off her fish oil and garlic for about 3 weeks.
Started (S)	$(\text{start} \text{resume} \text{begin} \text{restart} \text{add})(\backslash s+\backslash S+)^{\{0,7\}}\backslash s+\text{sup}$	<ul style="list-style-type: none"> • Started echinacea 1 week ago for cold. • We have resumed fish oil from 1g bid. • I have asked him to begin using fish oil 3 capsules a per day, and he is agreeable to this. • Patient is going to restart calcium, zinc, magnesium, and Vitamin E. • Add supplements with ginger.
Unclassified (U)	$(\text{suggest} \text{avoid} \text{deny} \text{recommend} \text{decline})(\backslash s+\backslash S+)^{\{0,7\}}\text{sup}$	<ul style="list-style-type: none"> • Also suggested that she could consider trying otc ginkgo biloba. • Avoid use of st. john's wort on methadone as it can affect systemic level. • He denied any supplements such as ginkgo or other antiplatelet factors. • I did recommend taking over-the-counter fish oil, either 500 or 1000 mg per day. • Pt declined use of ginger ale and crackers.

TABLE II

THE PERFORMANCE OF THE RULE-BASED CLASSIFIER ON THE TEST DATA

Status	Number	Precision	recall	F-measure
Continuing (C)	112	0.94	0.91	0.93
Discontinued (D)	47	1.00	0.96	0.98
Started (S)	74	0.96	0.96	0.96
Unclassified (U)	67	0.81	0.88	0.84

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE III

THE PERFORMANCE OF THE MACHINE LEARNING-BASED CLASSIFIER ON THE TEST DATA

Status	Number	Precision	recall	F-measure
Continuing (C)	112	0.87	0.95	0.91
Discontinued (D)	47	0.93	0.89	0.91
Started (S)	74	0.90	0.93	0.91
Unclassified (U)	67	0.79	0.66	0.72

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript