Modeling Drug Exposure Data in Electronic Medical Records: an Application to Warfarin

Mei Liu, PhD[1], Min Jiang, MS[1], Vivian K. Kawai, MD[2], Charles M. Stein, MD[2], Dan M. Roden, MD[2], Joshua C. Denny[1], Hua Xu, PhD[1]

[1]Department of Biomedical Informatics, [2]Department of Medicine-Clinical Pharmacology Division, Vanderbilt University, School of Medicine, Nashville, TN

Abstract

Identification of patients' drug exposure information is critical to drug-related research that is based on electronic medical records (EMRs). Drug information is often embedded in clinical narratives and drug regimens change frequently because of various reasons like intolerance or insurance issues, making accurate modeling challenging. Here, we developed an informatics framework to determine patient drug exposure histories from EMRs by combining natural language processing (NLP) and machine learning (ML) technologies. Our framework consists of three phases: 1) drug entity recognition - identifying drug mentions; 2) drug event detection - labeling drug mentions with a status (e.g., "on" or "stop"); and 3) drug exposure modeling - predicting if a patient is taking a drug at a given time using the status and temporal information associated with the mentions. We applied the framework to determine patient warfarin exposure at hospital admissions and achieved 87% precision, 79% recall, and an area under the receiver-operator characteristic curve of 0.93.

INTRODUCTION

The rapid growth of Electronic Medical Record (EMR) systems has created an unprecedented practice-based longitudinal data resource for conducting observational research. Structured medication data have long been utilized for such research efforts, including pharmacoepidemiology, pharmacogenomic, and service-related health care investigations[1]. More recently, a number of efforts have linked new and existing EMR databases with archived biological material to accelerate research in personalized medicine[2]. These programs seek to identify common and rare genetic variants that contribute to variability in drug response within the context of relevant clinical covariates.

Despite the promise of EMRs as research tools for drug-related clinical research, challenges exist for large-scale studies using practice-based medication data. Perhaps one of the most significant challenges is the accurate quantification of drug exposure histories (e.g. start date, stop date, duration of exposure) from EMR data, especially in the outpatient setting. The challenge is fueled by frequent changes in medication regimens over time due to factors like intolerances, insurance issues, generics, and treatment plan adjustment to reach clinical targets. As such, medications may continue to be mentioned in the EMR even when they are not being prescribed. For outpatients, medication mentions usually exist as heterogeneous data types including both structured (e.g., e-prescribing systems) and unstructured (e.g. clinical notes) formats. Much of the detailed drug information such as status changes is embedded in narrative text, which is not immediately available for use and analysis. Thus, to determine drug exposure periods, a domain expert typically reconciles multiple sources of drug information to compile a longitudinal history of drug usage – a difficult and costly task. The manual extraction approach to identify drug exposure is one rate-limiting step for EMR-based drug research, and it often results in study cohorts with limited sample size[3]. For inpatients, structured drug information from electronic applications such as physician order entry systems, pharmacy fills records, or medication administration records, ameliorates some of these problems, but such resource is relatively new (containing recent years of data only) and it lacks ubiquity and ignores over-the-counter medications.

In this study, we introduce a natural language processing (NLP) and machine learning (ML) based framework to determine patient drug exposure from outpatient EMRs for clinical research. More specifically as an application, we investigated the framework in determining whether a patient was on warfarin (a widely prescribed anticoagulant) or not at the time of hospital admission, which is crucial to an ongoing EMR-based warfarin pharmacogenetic study at Vanderbilt.

Background

Over the years, various techniques, including machine learning (ML), natural language processing (NLP), heuristic rules, and regular expressions, have been investigated for automated extraction of medication information. Rule-

based approach was examined by Kraus et al.[4] to automatically extract drug name, dosage, and route information from clinical notes. Levin et al.[5] utilized open source tools to implement a system that extracted drug information from narrative anesthesia health records and normalized the drug information by mapping it to the controlled nomenclature of RxNorm. The project team developed an NLP system, MedEx, combining method of lookup, regular expression, and chart parser to accurately extract comprehensive drug information (e.g. drug name, strength, route, and frequency) from EMRs[6]. MedEx has been integrated with other existing NLP tools for medication extraction, ranking as the 2nd best system out of 20 entries in the 2009 i2b2 medication extraction challenge[7].

A number of studies have investigated recognition of drug status changes in clinical narratives. Pakhomov et al.[8] built a Maximum Entropy classifier on rheumatology clinical notes with a variety of feature sets to classify medication status changes into four categories: past, continuing, discontinued, and started. Breydo et al.[9] carried out an assessment in identification of inactive medications by analyzing the context space for presence of semantic keys that indicate discontinuation of the drugs. More recently, Sohn et al.[10] investigated a rule-based method and support vector machine (SVM) classifier to automatically identify five categories of medication status (i.e. no change, start, stop, increase, and decrease) in clinical narratives using only indication words as features.

Medication reconciliation (MR) is also related to the current study. MR is a process of creating an accurate list of a patient's medications for the purposes of resolving discrepancies and supporting accurate medication orders. Many processes developed for MR include manual audits and survey. Recently, with the increasing availability of EMRs, the MR processes are being integrated into the clinician's workflow, but they still largely remain manual tasks[11,12]. Poon et al.[13] developed a system to support creation of a preadmission medication list (PAML) by first extracting the medication information from four electronic sources (i.e. current medication list from two outpatient EMRs and most recent discharge medication orders within two CPOE systems) and then presenting it to the clinician, who will finally consolidate and reconcile the information into a single coherent list. Recently, Peterson et al.[14] studied PAMLs, admission history and physical notes, and outpatient records as sources of accurate MR data when compared to an adjudicated gold standard including patient and family interview and pharmacy fill records, finding that the PAML had the best recall while the admission note was the most precise source of admission medication data. Cimino and colleagues[15] presented an automatic approach to MR from a mixture of coded and narrative sources. All in all, these approaches seek to assure a patient's medications from sources only at time points related to admissions. In contrary, the proposed informatics framework in the current study is to make patient drug exposure status predictions at any given time utilizing all surrounding clinical notes and corresponding temporal information.

METHODS
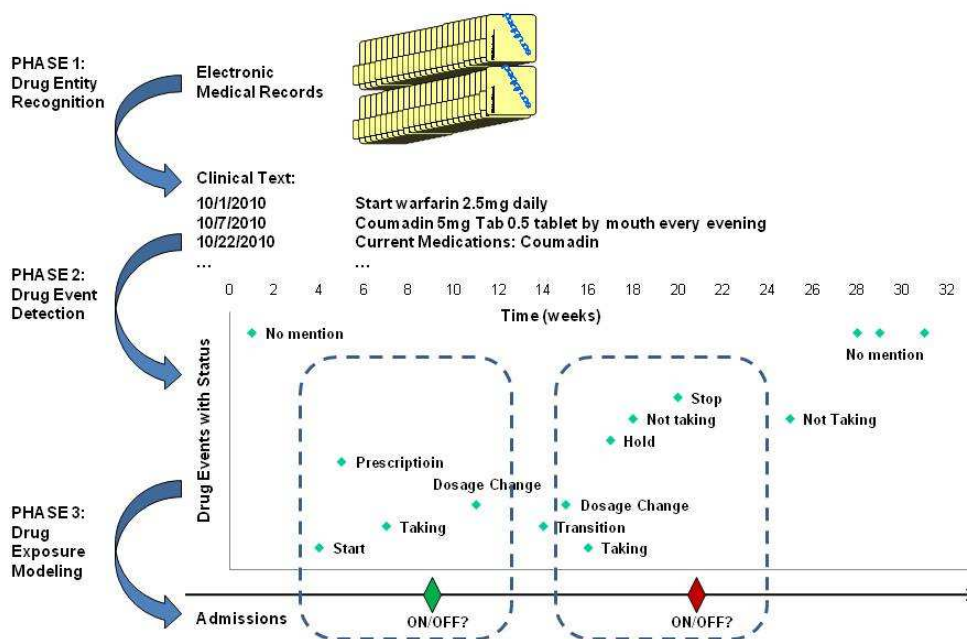
Figure 1. Overview of the proposed informatics framework for patient drug exposure modeling

The proposed framework consists of three phases (Figure 1): (1) drug entity recognition - to identify drug mentions in clinical notes; (2) drug event detection - to automatically label the free text drug mentions with predefined categories such as '*on*' or '*stop*'; (3) drug exposure modeling - to predict whether a patient is currently taking (e.g., ON or OFF) a drug at a given time using the status and temporal information associated with the drug mentions around the time point.

Drug Entity Recognition

To identify free text drug mentions in EMRs, we employed the medication extraction system, MedEx, developed by our group at Vanderbilt[6,7]. It consists of a semantic tagger and a context-free grammar parser that parses medication sentences using a semantic grammar. It has been shown to accurately identify not only drug names but also signature information such as strength, route, and frequency from discharge summaries and clinical visit notes. In this study, MedEx was used to extract sentences containing specific drug mentions from all types of clinical notes at Vanderbilt University Hospital.

Drug Event Detection

The goal of this phase is to determine why a drug is mentioned in EMR. Based on a manual analysis of drugs in clinical records, we defined a list with 7 status classes: '*on*', '*stop*', '*hold*', '*dosage change*', '*complication*', '*transition*', and '*other*'. Mentions belonging to the '*on*' class describe a patient being put on, staying on, or resuming a medication, for example, "…started on warfarin anticoagulation…" and "continued the warfarin to present." On the other hand, '*stop*' mentions describe the intent of either physician or patient to discontinue the drug (e.g., "…please call patient to stop her Coumadin…", "he had stopped taking his Coumadin"). Sentences in the '*hold*' class convey the idea that the medication should be held temporarily for observation or other purposes (e.g., "while the patient was on antibiotics, his Coumadin was held several days"). The '*complication*' class consists of sentences describing possible adverse reactions of the medication, for example, "XXX was on chronic anticoagulation with Coumadin who presented to an OSH with vaginal bleeding…" Sentences in the '*dosage change*' class discusses medication dosage change and the '*transition*' status characterizes change of medications (either from the drug of interest to others or vice versa). Lastly, any comments or general information regarding a medication that is insufficient to be categorized into the above classes are labeled with the '*other*' status.

The drug event detection task is converted into a text classification task in which we assign one of the seven labels to each mention of a drug based on context of the mention. We developed a SVM classifier that uses three types of evidence as features to assign drug mentions with appropriate status: (1) local textual features around drug mentions such as a window of contextual words; (2) semantics of the medication, such as the signature information; (3) discourse information such as note type, source, and section in which the drug mentions appeared. The medication signature information (derived unchanged from MedEx) consists of seven categories: drug form (e.g., 'tablet'), strength (e.g., '10mg'), dose amount (e.g., '2 tablets'), route (e.g., 'by mouth'), frequency (e.g., 'b.i.d.'), duration (e.g., 'for 10 days'), and necessity (e.g., 'prn'). For each sentence identified in the previous phase, different sets of features (i.e., contextual, semantic, and discourse) were generated. Finally for the sentence classification, we employed a SVM learner called LIBLINEAR[16], which is the winner of the International Conference on Machine Learning (ICML) 2008 large-scale learning challenge. It can quickly and accurately train on a dataset with much larger feature space and instances.

Drug Exposure Modeling

For patient drug exposure modeling, we formulated the task as a typical two-class classification problem. Essentially, is a patient ON/OFF a particular drug at a given time? An immediate challenge we face is how to represent the intrinsic time-series data for classification. Time-series data have been extensively explored in other application domains such as signal processing, econometrics and finance. However, the time series data in those areas are typically measured at successive times spaced at uniform time intervals. For example, the daily closing values of the Dow Jones index or the annual flow volume of the Nile River. Past health informatics research in time-series analysis also relied on well-behaved data such as signals from intensive care unit (ICU)[17], electrocardiogram (ECG)[18] and hemodialysis (HD) sessions[19]. In contrast, the time-series data in EMR medication records is noisy and incomplete as notes are created only when patients interact with providers or in follow-up communications. Moreover, for drug exposure modeling from EMRs, there are scenarios where notes are available but no mentioning of the drug of interest, which we refer here as '*no mention*' events. These '*no mention*' events may still provide useful information

for our classification because one may speculate a patient is off a drug when the '*no mention*' event repeats many times over a period of time.

As an initial study to explore the temporal medication data for exposure modeling, we implemented the classification task using the above mentioned SVM classifier with investigations on three temporality-related issues: 1) length of the time window around a given point - how much medication history is required for accurate prediction; 2) representation of the temporal information - how time intervals should be represented; and 3) contribution of the drug status information and '*no mention*' events. The drug status derived from the previous phase can be integrated with the temporal and discourse information to form features. In this study, the note creation date was taken as the time-stamp for each drug mention, and we explored time intervals '*weekly*' and '*monthly*' to represent temporality for classification. For example, if a patient had a drug mention status of "on" in discharge summary 10 days before a time point, its derived feature would be '*2_week_before+DS+on*' with the '*weekly*' interval and '*1_month_before+DS+on*' with the '*monthly*' interval.

Three types of models were analyzed to predict if a patient is ON/OFF a particular medication at a given time. The first model makes prediction purely based on the occurrences of drug mentions without interpreting their meanings or status, which is referred here as the baseline model. More specifically, it does not differentiate an '*on*' mention from a '*stop*' or any other types of mentions; thus the derived feature from the previous example becomes '*2_week_before+DS*' with the '*weekly*' interval and '*1_month_before+DS*' with the '*monthly*' interval. The remaining two models were designed to investigate contributions of the drug status and '*no mention*' events. The second model utilizes the drug status information but not the '*no mention*' cases and the third model takes both drug status and '*no mention*' cases into consideration.

Data Source and Evaluation

As a pilot study, we applied the proposed framework to determine patient warfarin exposure at hospital admissions. We used clinical notes from the Synthetic Derivative (SD) database, which is a de-identified copy of the EMR at Vanderbilt University Hospital. A physician manually reviewed 107 admissions from 61 randomly selected patients in the SD and determined whether a patient was on/off warfarin at each admission. All available clinical notes of the 61 selected patients were collected and 6,046 sentences containing the keyword warfarin/Coumadin were extracted and manually reviewed to assign 1 of the 7 status classes. The manually curated datasets were then used as gold standards to build and evaluate our models.

To assess the performance of each model, the study calculated accuracy (ACC), precision (P), recall (R), F-measure (F), specificity (S), and area under a receiver-operating characteristic (ROC) curve (AUC). Accuracy is commonly known as the proportion of true results obtained (i.e. $ACC = (TP + TN)/(TP + FP + FN + TN)$). Precision is defined as the proportion of true positives against all predicted positive results (i.e. $P = TP/(TP+FP)$). Recall, also known as the true positive rate or sensitivity, is the fraction of true positives among all positives (i.e. $R = TP/(TP+FN)$). F-measure is the harmonic mean of precision and recall (i.e. $F = 2PR/(P+R)$). Specificity, also known as the true negative rate, is defined as $S = TN/(TN+FP)$. Finally, ROC is a graphical plot of the sensitivity vs. 1 - specificity and the AUC is the area under the ROC curve. AUC is a commonly used statistic for classification model comparison in machine learning in which AUC of 1 represents a perfect test.

RESULTS

Drug Event Detection Results

To evaluate the drug status classification module in Phase II, we reserved 2/3 of the total 6,046 sentences as a training set and remaining 1/3 as a testing set. Since sample distribution varies among status classes with the '*on'* class having the most samples (i.e. 4,192 sentences) and the '*transition*' class having the least number of samples (i.e. 32 sentences), the samples in each class were separated individually into training and testing sets according to the above proportions.

Before evaluation, the optimal window size for the contextual feature was empirically determined and the SVM parameters were optimized through a 5-fold cross validation (CV) on the training data for different feature sets (i.e. semantics only, discourse only, contextual only, contextual + semantics, contextual + semantics + discourse). Generally, the more features the higher the performance. The best mean accuracy of 92% was achieved on the 5-fold

CV of the training dataset using all features (Table 1). Finally, we evaluated the trained SVM classifier on the testing dataset and the best performance observed was 93% in accuracy.

Feature Sets	ACC on Training Set	ACC on Test Set
Semantics	0.69	0.68
Discourse	0.78	0.77
Contextual	0.91	0.91
Contextual + Semantics	0.91	0.92
Contextual + Semantics + Discourse	**0.92**	**0.93**

Table 1. Drug event detection for all status classes using different feature sets

Precision and recall for each drug status class is displayed in Table 2. As shown, high F-measure (from 82% to 95%) was achieved for most status classes except for the '*dosage change*' and '*transition*' classes. The poor performance may be due to the fact that the two classes had the least number of instances compared to others. For classification using ML algorithms, the most common class often tends to dominate the decision process, thus producing a biased classification performance on the majority class.

	Precision	Recall	F-Measure
stop	0.91	0.75	0.82
start	0.93	0.98	0.95
hold	0.95	0.81	0.87
other	0.93	0.90	0.91
dosage change	0.83	0.64	0.72
complication	0.94	0.84	0.89
transition	1.0	0.18	0.31

Table 2. Drug event detection for each status class using the contextual+Semantics+Discourse feature set.

Drug Exposure Modeling Results

For patient drug exposure determination in Phase III, we investigated '*weekly*' and '*monthly*' time intervals for temporal feature representation and combined with the drug status information within various time frames (i.e.,180, 150, 120, 90, 60, 30, 14, 7 days before and 0, 7 days after). Parameters for the SVM classifier were again empirically determined with 5-fold CV. All 32 combinations of feature sets were examined. To further ensure correctness in test results, we ran each experiment 20 times with 5-fold CV in each run and reported average over all runs as the final result (Table 3).

The evaluation results (Table 3) demonstrated that the '*Drug-mention + Drug-status + No-mention*' model in which both the drug-status and '*no mention*' information were considered performed the best with the highest AUC of 0.93 and F-measure of 83%. Drug status information indeed contributed to the accurate prediction of patient drug exposure status with increases of 2% - 7% in AUC and 1% - 4% in F-measure. The '*no mention*' information contributed significantly with increases in AUC of 4% - 7% and F-measure of 1% - 11%. Although for the first case (i.e. '*monthly*' time interval, 60 days before, and 0 days after), recall of the '*Drug-mention + Drug-status*' model appeared to be better compared to the '*Drug-mention + Drug-satus + No-mention*' model, the corresponding precision and specificity is dramatically lower, which yielded 6% decrease in AUC. For two-class classification problems, it is more important to focus on the AUC because it measures how well the classifier is in predicting negative cases as negatives and positive cases as positives. AUC of the '*Drug-mention + Drug-status*' model is better (2% - 7%) compared to the baseline '*Drug-mention*' model as a result of the utilization of the drug status information, but it is still much lower (4% - 7%) opposed to the '*Drug-mention + Drug-status + No-mention*' model.

Moreover, for both the '*Drug-mention + Drug-status*' model and the '*Drug-mention + Drug-status + No-mention*' models, we noticed that utilization of information after the admission time yielded greater AUCs. With the '*Drug-mention + Drug-status + No-mention*' model, when no information after the admission time (i.e. 0 days after) was utilized, the '*monthly*' time interval performed better. On the other hand, if we incorporated information after the

hospital admission (i.e. 7 days after), the '*weekly*' time interval performed better compared to the '*monthly*' counterpart. Interestingly, the '*weekly*' interval consistently performed better in the '*Drug-mention + Drug-Status*' model but worse in the baseline '*Drug-mention*' model.

Model	Feature Set	Precision	Recall	Specificity	AUC	ACC	F-Score
Drug-mention	(Monthly, 60, 0)	0.70±0.02	0.91±0.04	0.58±0.05	0.82±0.02	0.75±0.03	0.79±0.02
	(Monthly, 14, 7)	0.63±0.02	0.77±0.05	0.60±0.04	0.77±0.03	0.68±0.03	0.69±0.03
	(Weekly, 30, 0)	0.65±0.02	0.82±0.03	0.52±0.04	0.81±0.03	0.67±0.03	0.72±0.03
	(Weekly, 30, 7)	0.62±0.02	0.74±0.04	0.60±0.03	0.79±0.02	0.67±0.03	0.68±0.03
Drug-mention + Drug-status	(Monthly, 60, 0)	0.68±0.03	0.93±0.04	0.53±0.06	0.82±0.03	0.74±0.04	0.78±0.03
	(Monthly, 14, 7)	0.64±0.04	0.76±0.05	0.62±0.06	0.81±0.03	0.69±0.04	0.70±0.04
	(Weekly, 30, 0)	0.64±0.03	0.85±0.05	0.46±0.05	0.83±0.03	0.67±0.04	0.73±0.04
	(Weekly, 30, 7)	0.65±0.03	0.81±0.05	0.61±0.05	0.86±0.03	0.71±0.03	0.72±0.03
Drug-mention + Drug-status + No-mention	(Monthly, 60, 0)	0.73±0.04	0.80±0.06	0.84±0.03	0.88±0.03	0.82±0.03	0.76±0.04
	(Monthly, 14, 7)	0.78±0.04	0.76±0.05	0.88±0.02	0.91±0.03	0.84±0.03	0.77±0.04
	(Weekly, 30, 0)	0.80±0.04	0.69±0.04	0.90±0.03	0.87±0.03	0.82±0.02	0.74±0.03
	(Weekly, 30, 7)	**0.87±0.02**	**0.79±0.04**	**0.93±0.01**	**0.93±0.02**	**0.88±0.01**	**0.83±0.02**

Table 3. Evaluation of patient warfarin exposure status prediction. Three models were evaluated: (1) 'Drug-mention' (baseline) model: only discourse and drug signature information used (neither drug status nor '*no mention*' information was used); (2) 'Drug-mention + Drug-status' model: only discourse, drug signature, and drug status information was used; and (3) 'Drug-mention + Drug-status + No-mention' model: included all information. The feature set column contains the three features we investigated for evaluation: (1) temporal representation (i.e. 'monthly' or 'weekly'); (2) retrospective time window - number of days to look back from the given time for data collection; (3) number of days to look after the given time. Results are average of 20 runs of 5-fold CV and standard errors are shown following ±. Various retrospective time windows (i.e. 180, 90, 60, 30, 14, 7) were experimented but only the ones with the best performance are shown here.

DISCUSSION

In this study, we developed a combined NLP and ML-based framework to automatically and accurately determine patient drug exposure status from EMRs, and tested the model on a set of patients exposed to the anticoagulant warfarin. Our methods consisted of three phases: 1) drug entity recognition; 2) drug status detection; and 3) drug exposure modeling. Our results with warfarin were promising, demonstrating that the best system achieved an overall AUC of 0.93 with F-measure of 0.83, suggesting that such methods may assist in identification of automatic modeling of patient drug exposure histories. Our methods were also able to accurately predict a number of individual medication status representations, giving a richer semantic understanding to the mentions of the medicines. This paper presents a novel attempt at temporal-based mining of drug exposure histories in EMRs combining NLP and ML technologies.

For the drug event detection in the Phase II, the trained SVM classifier was able to correctly label warfarin mentions in the test set to their 'true' statuses with an accuracy of 93%. High F-measures (82% to 95%) were achieved for most status classes except for the '*dosage change*' and '*transition*' classes as a result of small sample size (less than 100 training samples). In an imbalanced classification problem such as this one, the large preponderance class often dominates the decision process, which produces classification bias toward the majority class. As shown in Table 2, the precision for these two classes are relatively high, and the main problem is actually the low recall which is attributed to the numerous false negatives.

As a further analysis, we manually analyzed the false negatives in the '*dosage change*' and '*transition*' classes and observed that many errors may be corrected easily with heuristic rules. Some sentences in the '*dosage change*' class

contain explicit indicative verbs but were misclassified due to the lack of medication signature information, for example, "…adjusting Coumadin dose…" and "…lowered her dose of Coumadin…" In addition, some sentences were misclassified because they contained more ambiguous verbs (e.g., 'put on', 'back on', 'given') that usually denote other statuses. For instance, "…put on increased dose of Coumadin…" was misclassified as '*on*'. We observed 14 out of 17 (83%) false negatives in the '*dosage change*' class to be caused by these two types of errors, which can be easily fixed by designing linguistic rules that searches for verbs such as 'lower', 'change', 'increase', 'decrease', 'adjust', 'correct', and 'cut back' in close proximity of the drug name. Most of the false negatives in the '*transition*' class such as "…d/c heparin and transition to Coumadin…" and "…stop his Coumadin and get Lovenox…" are also retrievable by heuristic rules. For instance, we may look for patterns like "D1 transition to D2", "stop D1 and start D2", "start D1 … with transition to D2", and "transition off D1 and use D2". In future work, to improve performance, we will integrate the heuristic rules with the current proposed ML-based module to label drug status for status classes with less number of samples. Moreover, we noticed a small number of the errors were due to incorrect annotation, which can be ameliorated through double annotation in the future for quality assurance.

For the warfarin exposure prediction at admissions in Phase III, the largest AUC achieved was 0.93 with 87% precision and 80% recall. Different temporal representation and time window for information inclusion were examined and we observed that including information after the admission time definitely helps in prediction and the '*weekly*' time interval tends to favor a smaller time window. Analysis also demonstrated significant contributions of the drug status and '*no mention*' events in modeling patient drug exposure from EMRs. Nevertheless, status of the drug mentions could be complex to define as granularity of the definition is of concern. Actually from the 7 predefined status classes, we noticed hierarchical structures in those definitions. For example, the '*on*' class includes sentences describing a patient being put on, staying on, or resuming a medication, and in this sense, prescription orders of a medication would also belong to this category. To evaluate the difference in which status granularity would make in predicting patient drug exposure, we refined the predefined 7 status classes by expanding them into 12 classes and performed experiments using the manually annotated status information of 12 classes. Evaluation results showed an increase of 0.01 in AUC.

Most ML models are "black boxes" to users from which one cannot easily interpret how the results were derived and why certain errors were made; however, these hidden steps or rules are often the most interesting. We manually reviewed some of the false negatives and false positives made by the classifier in predicting patient warfarin exposure status and discovered most errors were due to the loss of sequential information by the classifier. For instance, we noticed drug mentions in some false negatives have the following pattern: a sequence of on-off and off-on events and then a '*on*' followed by a number of '*no mention*' events immediately before the given admission. Likewise for the false positives, we observed similar pattern but with a '*stop*' followed by a number of '*no mention*' events immediate before admission. It seems that the last event before admission is important for prediction. However, such important event order information is missed by the SVM classifier because it cannot fully utilize the sequential nature of the data. Sequential or time-series data mining has been extensively studied by researchers in computer science and their methods include Hidden Markov Models[20] and clustering algorithms[21]. As a future work, we will exploit other computational models to consider sequential order of the medication events.

The current study has several limitations. First of all, the gold-standard data sets were annotated by one clinician and the sample size was small which may lead to over-fitting. A double annotation experiment would insure quality and allow investigation of the complexity of the task for humans. And future effort to increase the sample size will alleviate the over-fitting problem. Second, our framework was applied to one drug with complex dosing regimens and associated with significant morbidity; thus, physicians are more likely to carefully note the status of warfarin more frequently than perhaps other medications. In later work, we plan to extend these methods to other drugs as administration pattern varies greatly among different drug classes. Third, the note creation date was taken as the time-stamp for each drug mention, which may not be the true time. Truly understanding the time for events in narrative documents is another challenging NLP task. We[22] and others[23,24,25] have begun investigations on temporal information extraction from clinical narrative, but such work has not yet been applied robustly to medication information. Compliance is another issue related to drug exposure in the outpatient setting. The model we developed here can combine all evidence in EMR (including both structured and narrative sources) to make predictions for drug status of patients. With more and more recorded drug information in EMR (e.g., medical reconciliation tools at transitional points and text messages between patients and physicians), we hope such methods would improve the capability to determine drug compliance status of patients.

CONCLUSION

Patient drug exposure history is critical for drug-related clinical research using EMRs, and may assist medication reconciliation efforts. Although EMRs are an exceptional resource for longitudinal information, it is challenging to quantify the patient exposure histories. Evaluation of our framework in determining warfarin exposure at hospital admissions showed promising results of 87% in precision, 79% in recall, and 0.93 in AUC. Future work should include other drug event detection methods such as using linguistic clues for drug status labeling and other predictive models for drug exposure prediction.

ACKNOWLEDGEMENT

REFERENCES

1. Strom BL. Pharmacoepidemiology. 4th ed. 2005, Chichester; Hoboken, NJ: J. Wiley. Xvii, p889.
2. McCarty CA, Wilke RA. Biobanking and pharmacogenomics. Pharmacogenomics. 2010;11(5):637-41.
3. Wilke RA, Moore JH, Burmester JK. Relative impact of CYP3A genotype and concomitant medication on the severity of atorvastatin-induced muscle damage. Pharmacogenet Genomics. 2005;15(6):415-21.
4. Kraus S, Blake C, West SL. Information extraction from medical notes. Proceedings of the 12th World Congress on Health Informatics Building Sustainable Health Systems (MedInfo). 2007;:1662-4.
5. Levin MA, Krol M, Doshi AM, Reich DL. Extraction and mapping of drug names from free text to a standardized nomenclature. AMIA Ann Symp Proc. 2007;:438-42.
6. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR. MedEx: a medication information extraction system for clinical narratives. J Am Med Inform Assoc. 2010;17:19-24.
7. Doan S, Bastarache L, Klimkowski S, Denny JC, Xu H. Integrating existing natural language processing tools for medication extraction from discharge summaries. J Am Med Inform Assoc. 2010;17(5):528-31.
8. Pakhomov SV, Ruggieri A, Chute CG. Maximum entropy modeling for mining patient medication status from free text. AMIA Ann Symp Proc. 2002;:587-91.
9. Breydo EM, Chu JT, Turchin A. Identification of inactive medications in narrative medical text. AMIA Ann Symp Proc. 2008;:66-70.
10. Sohn S, Murphy SP, Masanz, JJ. Classification of medication status change in clinical narratives. AMIA Ann Symp Proc. 2010;:762-6.
11. Pronovost P, Weast B, Schwarz M et al. Medication reconciliation: a practical tool to reduce the risk of medication errors. J Crit Care. 2003 Dec;18(4):201-5.
12. Poole DL, Chainakul JN, Pearson M et al. Medication reconciliation: a necessity in promoting a safe hospital discharge. NAHQ. 2006 May-June 2006.
13. Poon EG, Blumenfeld B, Hamann C et al. Design and implementation of an application and associated services to support interdisciplinary medication reconciliation efforts at an integrated healthcare delivery network. J Am Med Inform Assoc. 2006;13(6):581-92.
14. Peterson JF, Shi Y, Denny JC et al. Prevalence and clinical significance of discrepancies within three computerized pre-admission medication lists. AMIA Ann Symp Proc. 2010;:642-6.
15. Cimino JJ, Bright TJ, Li J. Medication reconciliation using natural language processing and controlled terminologies. . Proceedings of the 12th World Congress on Health Informatics Building Sustainable Health Systems (MedInfo). 2007;:679-83.
16. Fan R-E, Chang K-W, Hsieh C-J, Wang X-R, Lin C-J. LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research. 2008;9:1871-4.
17. Sarkar M, Leong T. Characterization of medical time series using fuzzy similarity-based fractal dimentions. Artif Intell Med. 2003;27(2):201-22.
18. Kalpakis K, Gada D, Puttagunta V. Distance measures for effective clustering of ARIMA time-series. Proc of the 2001 IEEE International Conference on Data Mining (ICDM). 2001;:273-80.

19. Bellazzi R, Larizza C, Magni P, Bellazzi R. Temporal data mining for the quality assessment of hemodialysis services. Artif Intell Med.2005;34:25-39.
20. Chatzis SP, Kosmopoulos DI, Varvarigou TA. Robust sequential data modeling using an outline tolerant hidden markov model. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009;31(9):1657-69.
21. Liao TW. Clustering of time series data - a survey. Pattern Recognition. 2005;38(11):1857-74.
22. Denny JC, Peterson JF, Choma NN et al. Extracting timing and status descriptors for colonoscopy testing from electronic medical records. J Am Med Inform Assoc. 2010;17(4):383-8.
23. Zhou L, Melton GB, Parsons S, Hripcsak G. A temporal constraint structure for extracting temporal information from clinical narrative. J Biomed Inform. 2006;39(4):424-39.
24. Zhou L, Hripcsak G. Temporal reasoning with medical data - a review with emphasis on medical natural language processing. J Biomed Inform. 2007;40(2):183-202.
25. Zhou L, Parsons S, Hripcsak G. The evaluation of a temporal reasoning system in processing clinical discharge summaries. J Am Med Inform Assoc. 2008;15(1):99-106.