Toward Understanding Clinical Context of Medication Change Events in Clinical Narratives

Diwakar Mahajan[1], MS, Jennifer J. Liang[1], MD, Ching-Huei Tsou, PhD
IBM T.J. Watson Research Center, Yorktown Heights, NY

Abstract

Understanding medication events in clinical narratives is essential to achieving a complete picture of a patient's medication history. While prior research has explored identification of medication changes in clinical notes, due to the longitudinal and narrative nature of clinical documentation, extraction of medication change alone without the necessary clinical context is insufficient for use in real-world applications, such as medication timeline generation and medication reconciliation. Here, we present a framework to capture multi-dimensional context of medication changes documented in clinical notes. We define specific contextual aspects pertinent to medication change events (i.e. Action, Negation, Temporality, Certainty, and Actor), describe the annotation process and challenges encountered while creating the dataset, and explore models based on state-of-the-art transformers to automate the task. The resulting dataset, Contextualized Medication Event Dataset (CMED), consisting of 9,013 medications annotated over 500 clinical notes, will be released to the community as a shared task in 2021-2022.

Introduction

An accurate medication history is foundational for providing quality medical care, allowing healthcare providers to better assess the appropriateness of current treatments, detect potential medication-related pathologies or symptoms, and direct future treatment options. Although medication information is captured in various sources of clinical data, in practice, providers often rely on structured medication orders due to easier access in electronic health record systems. However, many medication events are documented only in unstructured clinical notes[1,2], which contain richer information but are difficult to search through especially at the point-of-care. In clinical practice, several scenarios may occur where a medication change is documented in clinical narratives but not in structured medication data. For example, if the patient already has a medication, the provider may instruct the patient to adjust the dosage or temporarily hold the medication without indicating the change in the medication order. Patient-initiated medication changes are also rarely captured in structured medication data, but would be documented in unstructured clinical narratives. Therefore, capturing medication event information from unstructured data within the patient medical record is required to achieve a full and complete picture of the patient's medication history.

When extracting medication changes from clinical text, it is important to consider the surrounding contextual information because of the longitudinal and narrative nature of clinical documentation. Specifically, the longitudinal quality of clinical text results in documentation of events over the course of the patient's medical history, from past and present events to future possible events. In addition, when documenting a clinical interaction with a patient, providers not only record what has occurred in the patient and the plan forward, but may also describe their clinical reasoning behind any medical decisions, including why certain treatment options were deferred. Such qualities of clinical documentation result in complex clinical events that would be insufficiently captured by extraction of medication change alone without consideration of the surrounding clinical context. This is especially true when developing a medication change extraction system to support real-world applications, such as medication timeline generation[3,4] or medication reconciliation[5,6]. For example, to generate a medication timeline, a system must extract not only the dosage adjustment action, but also place it in the correct point in time, i.e. knowledge of *when* the action is occurring. Similarly, medication reconciliation requires awareness of not just what medications were prescribed by healthcare providers, but also whether the patient is taking (or not taking) a medication, i.e. knowledge of *patient*-initiated actions.

To address the need for consideration of contextual information during medication change extraction, here we propose a conceptual framework to organize multi-dimensional context for medication events in clinical narratives, and present the resulting dataset – Contextualized Medication Event Dataset (CMED). CMED captures pertinent context for medication change events along five orthogonal dimensions (i.e. Action, Negation, Temporality, Certainty, and Actor), and

[1]Equal contribution

will be released to the research community as a shared task in 2021-2022. To the best of our knowledge, CMED will be the first dataset on medication change events made available to the research community. The major contributions of this paper are:

- a dataset of 9,013 medication mentions annotated with contextualized medication change events over 500 clinical notes,
- insights into the language used by clinicians when documenting medication changes and the challenges in interpreting this language during the annotation process, and
- experimental results showing that support vector machines (SVM) and transformer-based models can partially automate this task, with error analysis providing insights into the most challenging aspects of this task.

Related Work

Previous attempts in classifying medication change events in clinical notes were driven by a variety of use cases. For example, some prior works focused only on specific medications, while others focused on specific types of medication changes but for all medications. Past works targeting specific medications include work on warfarin (labels *on* or *stop*)[7], heart failure medications (labels *active*, *discontinued*, or *negative*)[8], beta blockers (labels *active*, *medication list*, *negated*, *discontinued*, or *other*)[9], and dietary supplements (labels *continuing*, *discontinued*, *started*, *unclassified*)[10,11]. In other works targeting medication changes but for all medications, Liu et al. focused only on medication discontinuation events[12], Sohn et al. considered a greater range of medication changes (labels *start, stop, increase, decrease, no-change*)[13], Lerner et al. also included labels that indicate sequential changes (labels *start, start+stop, stop, continue, switch, decrease, increase*)[14], and Pakhomov et al. introduced temporal information in one of their labels (labels *past, continuing, stop, start, not classified*)[15]. Note that these works, driven by a variety of use cases, have resulted in a mixed set of labels, some of which may not cover all types of changes (e.g. labels used by Pakhomov et al. - *past, continuing, stop, start*[15], do not differentiate between increase and decrease), or do not cover all aspects of a medication event (e.g. labels used by Sohn et al. - *start, stop, increase, decrease, no-change*[13], cover all types of changes but do not provide temporal information).

In another body of work, attempts have been made to recognize context or assertions for medical concepts such as problems and tests[16–19]. There have also been attempts at identifying negated medical concepts in clinical text[17,20–23]. Although some of these works identify aspects such as certainty and negation in medical events, none of them have been applied to medication change events. Further, no previous work has attempted to identify the actor behind an event, which is especially important for medication change events due to implications for patient adherence. Thus, there is need for a more organized schema of label definitions that better contextualizes medication events.

Our work differs from previous research in several ways. First, we capture relevant contextual information on top of just identifying medication change in clinical narratives. Second, the contextual information captured is organized along multiple orthogonal dimensions. Finally, we will be releasing our annotated dataset on contextualized medication change events as part of a shared task.

Methods

Data and Annotation

For this study, we used the 2014 i2b2/ UTHealth Natural Language Processing shared task corpus[24–26], which contain a total of 1,304 clinical notes over 296 patients. This corpus was selected due to its longitudinal nature, with 2-5 notes per patient, to allow potential future work in reconciling medication events extracted across different notes. To benefit from the longitudinal nature of this corpus while ensuring sufficient variation in our dataset, we first selected 44 patients from which all notes were chosen (total 199 notes), and then randomly selected 1-2 notes from each of the remaining 252 patients (total 301 notes). The resulting corpus consisted of 500 notes over 296 patients, of which 120 notes were doubly-annotated and adjudicated to measure inter-annotator agreement (IAA). Notes were annotated by a team of three annotators led by a physician. To assist the annotators, medications in the notes were pre-annotated using a medication extraction model[27], and corrected if necessary by the annotators during the annotation process (i.e. incorrect pre-annotations were removed and missed medication mentions were added).

We define a medication change event as any discussion about a medication change for a given patient. The annotation process is as follows. For each medication mention, the annotator first determines whether a medication change is being discussed, and assigns it one of the following medication change event labels:

- **NoDisposition**: no medication change is being discussed, e.g. "*pt is on Coumadin*", "*continue lisinopril*"
- **Disposition**: presence of a medication change being discussed, e.g. "*Start Plavix*"
- **Undetermined**: unclear if a medication change is being discussed and additional information is required to make the determination, e.g. "*Plan: Lasix*" – unclear if just stating a medication patient is on (NoDisposition) or starting a new medication (Disposition)

Next, for identified Disposition events, the annotator labels the clinical context for the event along five dimensions:

- **Action**: What is the change discussed? (Start, Stop, Increase, Decrease, OtherChange, UniqueDose, Unknown)
- **Negation**: Is the change being discussed negated? (Negated, NotNegated)
- **Temporality**: When is this change intended to occur? (Past, Present, Future, Unknown)
- **Certainty**: How likely is this change to have occurred / will occur? (Certain, Hypothetical, Conditional, Unknown)
- **Actor**: Who initiated the change? (Physician, Patient, Unknown)

Figure 1 summarizes the annotation process.

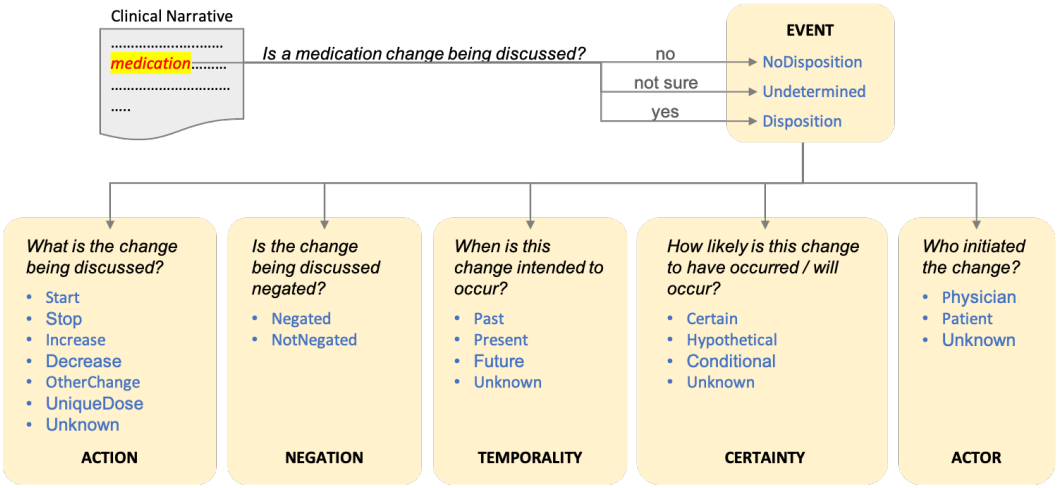

Figure 1: Annotation process.

Some sample annotations are shown in Table 1. Note that a single medication mention can have 0, 1, or more associated Disposition events. For example, "*May try amlodipine if develops cough*" has 1 associated Disposition event for *amlodipine* (Action|Negation|Temporality|Certainty|Actor = Start|NotNegated|Future|Conditional|Physician), while "*Was started on metformin at last visit but pt stopped after one week because of GI upset*" has 2 associated Disposition events for *metformin* (Start|NotNegated|Past|Certain|Physician and Stop|NotNegated|Past|Certain|Patient). These dimensions were designed to extract the relevant clinical context for medication change events, allowing us to capture a wide variety of attributes, such as references to past events (Temporality: Past), treatments being considered but not yet decided upon (Certainty: Hypothetical), or episodes of patient nonadherence (Actor: Patient).

While defining this task, we identified certain nuances of the annotation process that led to specific guidelines in our annotation protocol, specifically, the effect of (1) surrounding contextual information and (2) external medication-specific knowledge on label assignment. For contextual information, because providers can copy-and-paste text from previous notes[28], a sentence that appears to be Present based on tense (e.g. "*Start prednisone taper*"), when found in the Past Medical History (PMH) note section, actually indicates a Past action. To balance the need for consideration of

Table 1: Sample annotations demonstrating how labels change depending on the surrounding text.

Text	Event	Action	Negation	Temporality	Certainty	Actor
Pt currently on *lisinopril*	NoDisp	–	–	–	–	–
Plan: incr *losartan* from 1 tab qd to bid.	Disp	Increase	NotNegated	Present	Certain	Physician
Patient also given a prescription for *Hctz* 12.5mg QD but told not to start it yet until his next BP check and start only if BP $>$ 140/85	Disp	Start	NotNegated	Future	Conditional	Physician
In ED, given *ativan* 1 mg IV x1	Disp	UnqDose	NotNegated	Past	Certain	Physician
She was experiencing a bad episode of dry cough so stopped taking *lisinopril*	Disp	Stop	NotNegated	Past	Certain	Patient
Will hold off on empirically starting *abx* based on urinalysis	Disp	Start	Negated	Present	Certain	Physician
Pt's *lasix* dose was increased from 80 mg to 120 mg for four days, then reduced back to 80	Disp	Increase	NotNegated	Past	Certain	Physician
	Disp	Decrease	NotNegated	Past	Certain	Physician

contextual information while avoiding annotator error due to cognitive overload from having to remember the entire note, we asked annotators to consider only local context when determining the appropriate label, roughly defined as the immediate sentence containing the medication mention +/- 1 sentence and the note section. For knowledge of a specific medication's attributes (e.g. drug class, route of administration) affecting label assignment, because some medications are only available by prescription, while others are available over-the-counter, knowledge of a specific medication's availability can affect whether the Actor is more likely to be Physician or Patient. Similarly, some medications, such as Z-Pak, are dispensed as a set dosage pack with time-limited administration. Therefore, with that knowledge, the medication in a phrase such as "*Plan: Z-Pak*" should be annotated as Start|NotNegated|Present|Certain|Physician and Stop|NotNegated|Future|Certain|Physician. To best reflect the reality of the event, annotators were asked to annotate based on what they believe is going on, even if the assigned label depends on medical knowledge not in the text.

System Description

To automate the task of medication change classification, we split the task into five classification subtasks organized in a two-step process:

1. One subtask to classify each given medication mention into Disposition, NoDisposition, or Undetermined
2. Four subtasks to classify each given Disposition medication along four context dimensions: Action, Temporality, Certainty, and Actor. Negation was excluded from our experiments due to limited number of Negated instances.

For each subtask, we train a classification model based on two approaches: (1) SVM, and (2) transformer-based language models. As CMED will be used in a shared task in 2021-2022, we set aside 100 notes from our annotated data as blind data for the task. The remaining annotations in 400 notes are split into 75% for training, 5% for development and 20% for test. In addition to Negation Dimension, OtherChange for Action dimension and Unknown for all dimensions were excluded in our experiments due to the limited number of instances ($<$40).

Support Vector Machines. To explore the dataset and various dimensions, we elected to use a feature-based approach with a discriminative classification algorithm in our experiments. Our features were divided into the following classes: *lexico-syntactic*, *window*, *dep-parse*, *note-section*, and *RxNorm* features.

- **Lexico-syntactic**. Standard lexical features were utilized such as n-grams of whole, stemmed and lemmatized words, and part-of-speech tags.
- **Window**. We incorporated windowed lexico-syntactic features where we combined the positional information of word tokens with their lexico-syntactic features for the tokens that appear within of a window of 5 from the target medication mention in either direction. The window size of 5 was empirically determined.

- **Dep-parse**. We utilized features based on the dependency parse structure of the sentence containing the target medication, e.g. neighbors, parent and children of the medication concept in the dependency parse tree.
- **Note-section**. We experimented with a rule-based note section classification model that classifies each sentence based on the section headers and location of the sentence within the note.
- **RxNorm**. We leveraged features derived from RxNorm[29] for the target medication such as Anatomical Therapeutic Chemical class, ingredient etc.

Experiments were performed using SVM with a linear kernel and one-vs-all classification strategy as implemented in scikit-learn[30]. As initial experiments with different kernels (linear, polynomial and radial basis function) showed comparable performance, we chose linear kernel for all of our experiments. We used scispaCy[31] for sentence segmentation and extraction of lexico-syntactic and dep-parse features. RxNorm features were extracted using the RxNorm public API[32]. To deal with skew in class distribution, we adjusted the class weights to be inversely proportional to class frequencies.

Transformer-based Language Models. We experimented with state-of-the-art Bidirectional Encoder Representations from Transformer (BERT[34])-based language models, with the models pretrained on general-domain (BERT-base[34]) and in-domain (ClinicalBERT[35]) datasets. We formulated each of the five tasks as a sentence classification task by providing the surrounding sentence of the annotated medication as the context. In this process, we employed the pre-trained transformer to obtain a distributed representation, 0.2 dropout, and a fully connected layer of size 5 with softmax activation to make the classification for each of the five subtasks. We used the transformers package[36] to tune our models with the train and development splits, and present our results on the test split.

We evaluated our two system approaches with the following experiments on the 20% test data, (1) SVM, (2) BERT-base and (3) ClinicalBERT. We present the micro and macro precison, recall, and F_1 scores in the Results section. We also present a majority baseline on the training dataset with majority label for each dimension being chosen as the prediction for comparison against our model.

Results

Inter-annotator agreement and dataset statistics

Overall, our annotators achieved high IAA based on Cohen's kappa. On determining whether a medication change was being discussed (Disposition vs NoDisposition vs Undetermined), IAA was 0.88 on 2,495 annotated medication mentions. For the 367 instances of agreed Disposition events, Action, Negation and Temporality had the highest IAA at 0.87, 0.83, and 0.94, respectively, while Certainty and Actor had lower IAA at 0.75 and 0.72, respectively.

The resulting dataset, CMED, consists of 9,013 annotated medication mentions over 500 clinical notes. As CMED will be used in a shared task in 2021-2022, we only report detailed statistics on the training dataset in this paper. The training dataset consists of 7,230 annotated medication mentions over 400 notes, with distribution for specific labels presented in Table 2.

Table 2: Label distribution for the training dataset

Task	Label	Count	Task	Label	Count
Event	NoDisposition	5260	Temporality	Past	745
	Disposition	1413		Present	494
	Undetermined	557		Future	145
Action	Start	568		Unknown	29
	Stop	341	Certainty	Certain	1177
	Increase	129		Hypothetical	134
	Decrease	54		Conditional	100
	UniqueDose	285		Unknown	2
	OtherChange	1	Actor	Physician	1278
	Unknown	35		Patient	107
Negation	Negated	32		Unknown	28
	NotNegated	1381			

As observed in Table 2, less than 20% of medication mentions have an associated Disposition event. Further, a substantial percentage (7.7%) of medication mentions could not be resolved into Disposition or NoDisposition events due to the lack of sufficient information. Within the Action dimension, Start and Stop account for over 64.3% of Disposition events, UniqueDose for 20.2%, and titration events (Increase and Decrease) for 13%. Labels in the Temporality dimension reveal that over half of Disposition events occur in the Past (52.7%), which is reflective of the longitudinal and narrative nature of clinical notes.

Labels in the Temporality dimension reveal that over half of Disposition events occur in the Past (52.7%), which is reflective of the longitudinal and narrative nature of clinical notes. For Certainty, 16.6% of Disposition events are discussed in a Hypothetical or Conditional context. The prevalence of Past events as well as Hypothetical and Conditional events further confirms the need for contextualized medication event information. For Actor, as expected, the majority of medication changes in clinical text are initiated by healthcare providers (90.4%). Finally, Unknown in all dimensions, Negated in Negation dimension, and OtherChange in Action dimension are rare labels, each with less than 40 instances in the dataset.

System Results

Table 3 shows the results of (1) majority baseline, (2) SVM, (3) BERT-base, and (4) ClinicalBERT against the test dataset. While SVM provides a reasonable baseline, ClinicalBERT yields the best results across all five subtasks.

Table 3: Evaluation of our system under four settings: (1) majority baseline, (2) SVM, (3) BERT-base, and (4) ClinicalBERT. Bold indicates the highest micro and macro precision, recall, and F_1 scores for each subtask.

Experiment		**Task**														
		Event			**Action**			**Temporality**			**Certainty**			**Actor**		
		P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
majority baseline	micro	0.73	0.73	0.73	0.41	0.41	0.41	0.54	0.54	0.54	0.83	0.83	0.83	0.92	0.92	0.92
	macro	0.24	0.33	0.28	0.08	0.20	0.12	0.18	0.33	0.23	0.28	0.33	0.30	0.46	0.50	0.48
SVM	micro	0.79	0.79	0.79	0.59	0.59	0.59	0.71	0.71	0.71	0.83	0.83	0.83	0.88	0.88	0.88
	macro	0.62	0.63	0.63	0.50	0.51	0.50	0.60	0.59	0.59	0.59	0.53	0.56	0.63	0.68	0.65
BERT-base	micro	**0.88**	**0.88**	**0.88**	**0.75**	**0.75**	**0.75**	0.81	0.81	0.81	**0.90**	**0.90**	**0.90**	0.92	0.92	0.92
	macro	**0.79**	0.77	0.78	**0.75**	0.62	0.64	0.77	0.71	0.73	**0.83**	0.74	0.77	0.79	**0.72**	0.75
ClinicalBERT	micro	**0.88**	**0.88**	**0.88**	**0.75**	**0.75**	**0.75**	**0.83**	**0.83**	**0.83**	**0.90**	**0.90**	**0.90**	**0.93**	**0.93**	**0.93**
	macro	**0.79**	**0.79**	**0.79**	**0.75**	**0.63**	**0.65**	**0.80**	**0.74**	**0.75**	**0.83**	**0.76**	**0.79**	**0.83**	**0.72**	**0.76**

Discussion

Annotator disagreements

The lower IAA in Certainty can be attributed to the unclear language in physician documentation, for example:

- "*This was discussed by her neurologist who suggested starting with cholesterol medication and aspirin. She is here to discuss this further.*"
- "*We might suggest that she be started on Cisapride 10 mg qd*"
- "*It might be worthwhile to reduce his atenolol from 75 to 50mg once daily to see if this helps his Sx.*"

Hedging language is common in clinical documentation to express the uncertainty inherent in medical decision-making[33]. Reflecting this, different terms expressing varying degrees of certainty are used to indicate a prescribed intervention. For example, terms such as *try*, *recommend*, or *advise* are more definitive, while *consider*, *suggest*, or *may benefit from* require some interpretation to decide whether the action has actually taken place (Certain) or is only being discussed (Hypothetical). Although we attempted to address hedging language in the annotation guidelines, these disagreements reflect the inherent subjectivity in interpreting clinical language given the complexity and ambiguity in certain aspects of clinical documentation.

Disagreements along the Actor dimension were mostly found in references to past events, e.g. "*Improved breathing with spiriva, mistakenly stopped Advair*", where the subject who initiated the action is unspecified and was interpreted

differently by different annotators. Another interesting kind of disagreement was due to shared decision-making in clinical care, where both the patient and physician contribute to medical decisions and treatments, e.g. "*Discussed changing from Avapro to losartan for cost issues but he is uninterested and does not want to rock the boat.*"

Error Analysis

We conducted error analysis on the best performing model (i.e. ClinicalBERT) and identified three major categories of errors: (1) medication mentions with multiple annotations, (2) multiple medications within the same sentence, and (3) medication mentions that require context beyond the immediate sentence to determine the label. Examples for each of these error categories are shown in Table 4.

Table 4: Common error categories with examples, across five classification subtasks for the ClinicalBERT model.

Error Category	Example	Medication	Ground Truth	Prediction
Medication mentions with multiple annotations	In addition, 8 days prior to admission, pt's regular *lasix* dose was increased from 80 to 120 mg for four days, then reduced back to 80 mg.	*lasix*	Increase Decrease	Decrease Decrease
Multiple medications within the same sentence	We could change his statin from *Mevacor* to *Lipitor* to increase the HDL.	*Mevacor* *Lipitor*	Stop Start	Start Start
Limited context	P: He will restart *glyburide* 5 mg q.d. ~~when his blood sugar is greater than 200.~~	*glyburide*	Conditional	Certain

A significant percentage of errors occur due to medications having multiple event annotations. For example, in "*In addition, 8 days prior to admission, pt's regular lasix dose was increased from 80 to 120 mg for four days, then reduced back to 80 mg.*" the medication *lasix* has two labels for the Action dimension (Increase & Decrease). Since our current classification setup only allows for one prediction per medication mention, the model predicts only a single label Decrease for the Action dimension, leading to an error. Although medication mentions with multiple annotations form a small fraction of our overall dataset (1.2%), this error category accounts for a large percentage of errors across all dimensions (33% of Action errors, 23% for Temporality, 22% for Certainty, and 32% for Actor). One way to address this is to reformulate our task from a sentence classification task to a multi-label classification task.

Next, we observed that multiple medications present within the same sentence lead to errors as the model is unable to differentiate between the target medications. For example, "*We could change his statin from Mevacor to Lipitor to increase the HDL.*" contains two medications, *Mevacor* and *Lipitor*, that each have an Action label of Stop and Start, respectively. However, since they share the same context (i.e. sentence), the model predicts both the labels as Start. To resolve this, the system needs to more precisely identify the context for each medication mention. This can be achieved by reformulating this task as a named entity recognition task.

Finally, we observed a number of errors due to the limited context of a single sentence being available to the model for prediction. For example, in Table 4, the strike-through text "*when his blood sugar is greater than 200*" was not fed into ClinicalBERT. Hence the model made the correct prediction (Certainty: Certain) under the limited context given ("*P: He will restart glyburide 5 mg q.d.*"). While improved sentence segmentation will help this instance, a more general solution such as sequential sentence classification is likely to improve this error category.

Limitations

We acknowledge certain limitations to our work, specifically, those due to the nature of the underlying corpus and those that can be attributed to our annotation guidelines.

CMED is built on top of the corpus used in the 2014 i2b2/UTHealth Natural Language Processing shared task[24–26]. Since this corpus was selected for the purposes of the 2014 i2b2 shared task and therefore focused heavily on diabetes and heart disease patients, it is not representative of a typical patient population. Further, the corpus is limited to a single data warehouse i.e. Partners HealthCare Electronic Medical Records. Reproduction of our work on more diverse corpora is needed to better understand the effectiveness and applicability of our schema.

The current task focuses primarily on the identification and classification of contextual information for medication

change events. Medication mentions that do not discuss change are all grouped under a single label, i.e. NoDisposition, including descriptions of medication status (e.g. "*currently taking lisinopril*"), explicit directions to continue an existing medication (e.g. "*continue metformin*"), documented allergies to medications (e.g. "*sulfa (rash)*"), and other incidental mentions of medications. Depending on the specific use case and application, there may be value in further teasing out these different types of NoDisposition events. Further, although the current schema captures coarse temporality information of medication change events, extraction of more specific temporal references (e.g. "*at last visit*", "*x 10 days*") is needed to place these events in a more precise point in time. Finally, there may be additional contextual information that could contribute to improved understanding of medication changes but was not included in our annotation schema, such as the magnitude of change (i.e. what is the degree of change?) and the reason behind the change (i.e. why was this change introduced?). Such information was excluded from our current effort because of difficulties in defining a set of discrete labels to capture all possible values. Future work can be undertaken to provide such information through an extraction task built on top of CMED.

Applications and Future Work

Medication change events classified under the proposed schema can be directly leveraged in several real-world applications. Various visualizations and dashboard displays have been proposed to improve the usability of EHR systems, many of which include a medication timeline based on structured medication data[3,4]. Future research can be undertaken to apply analytics developed on CMED to such applications. For example, Present and Certain actions identified from clinical narratives under our schema can be incorporated into such medication timelines to further enrich them for a more comprehensive representation of a patient's medication history. These same events can also be presented alongside structured medication data to surface potentially missed or incorrect medication information in the structured data for purposes of medication reconciliation. Separately, patient-initiated actions captured under our schema can be used to supplement pharmacy prescription filling data towards improved understanding of medication nonadherence. Figure 2 shows an example of how such extracted multi-dimensional medication events may be used at the point-of-care, allowing users to control the information flow depending on their needs and specific use case.

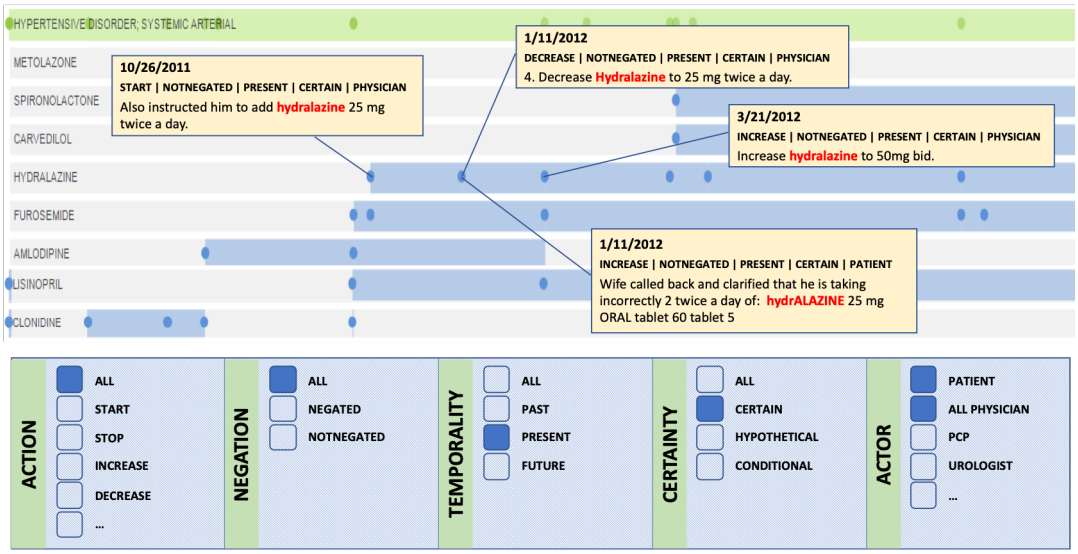

Figure 2: Prototype visualization incorporating extracted medication change events into structured EHR data showing all Present Certain events associated with the medication hydralazine.

CMED was purposefully generated on a longitudinal corpus (2014 i2b2 corpus with 2-5 notes per patient) to allow future work in tracking events for a specific medication over time, to support applications such as medication timeline visualization and summarization systems. To better support such applications and also more fully utilize the clinical context captured in CMED, additional research is necessary in several directions. First, 7.7% of medication mentions in CMED were classified as Undetermined, which can potentially be classified as either Disposition or NoDisposi-

tion given additional information. Further research in incorporating other sources of medication information (e.g. structured medication orders) is needed to better understand and characterize Undetermined events. Second, although temporal information is captured in the current schema, additional research is needed for applications that need temporal alignment or to place events in a more precise point in time. Third, the current schema classifies medications at an instance level, meaning that a single medication change occurrence that is documented multiple times in a patient's medical record will result in multiple instance-level events. Therefore, further work in concept normalization and coreference resolution is needed to resolve instance-level events in and across clinical narratives. Finally, to make CMED more useful in real-world settings, more work needs to done to identify the contradictions between different events recorded for the same medication and resolve them.

Conclusion

We introduce CMED, a dataset capturing contextual information – Action, Negation, Temporality, Certainty, and Actor – for medication change events documented in clinical notes, consisting of 9,013 annotated medication mentions over 500 notes. We describe our annotation guidelines, discuss specific nuances observed during the annotation process, and explore state-of-the-art transformer-based models to automate the task. As the first dataset on medication change events to be made available to the research community, CMED provides the necessary first step towards improved understanding of medication events in clinical narratives. We hope this effort will encourage future research and exploration into leveraging medication information from clinical narratives, and also contribute to other use cases that require consideration of contextual information for clinical events.

References

1. Turchin A, Shubina M, Breydo E, Pendergrass ML, Einbinder JS. Comparison of information content of structured and narrative text data sources on the example of medication intensification. JAMIA 2009;16(3):362-370.
2. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet. 2012;13(6):395-405.
3. Plaisant C, Mushlin R, Snyder A, Li J, Heller D, Shneiderman B. LifeLines: using visualization to enhance navigation and analysis of patient records. Proc AMIA Symp. 1998;76-80.
4. Belden JL, Wegier P, Patel J, Hutson A, Plaisant C, Moore JL, et al. Designing a medication timeline for patients and physicians. J Am Med Inform Assoc. 2019;26(2):95-105.
5. Cadwallader J, Spry K, Morea J, Russ AL, Duke J, Weiner M. Design of a medication reconciliation application: facilitating clinician-focused decision making with data from multiple sources. Appl Clin Inform. 2013;4(1).
6. 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):581-592.
7. Liu M, Jiang M, Kawai VK, et al. Modeling drug exposure data in electronic medical records: an application to warfarin. AMIA Annu Symp Proc. 2011;2011:815-823.
8. Meystre SM, Kim Y, Heavirland J, Williams J, Bray BE, Garvin J. Heart failure medications detection and prescription status classification in clinical narrative documents. Stud Health Technol Inform. 2015;216:609-613.
9. Alba PR, Patterson OV, Richman JS, DuVall SL. Automatic methods to extract prescription status quality measures from unstructured health records. Stud Health Technol Inform. 2019;264:15-19.
10. Fan Y, He L, Zhang R. Classification of use status for dietary supplements in clinical notes. Proceedings (IEEE Int Conf Bioinformatics Biomed). 2016;2016:1054-1061.
11. Fan Y, Zhang R. Using natural language processing methods to classify use status of dietary supplements in clinical notes. BMC Med Inform Decis Mak. 2018;18(Suppl 2):51.
12. Liu F, Pradhan R, Druhl E, et al. Learning to detect and understand drug discontinuation events from clinical narratives. J Am Med Inform Assoc. 2019;26(10):943-951.
13. Sohn S, Murphy SP, Masanz JJ, Kocher JP, Savova GK. Classification of medication status change in clinical narratives. AMIA Annu Symp Proc. 2010;2010:762-766.

14. Lerner I, Jouffroy J, Burgun A, Neuraz A. Learning the grammar of prescription: recurrent neural network grammars for medication information extraction in clinical texts. arXiv preprint arXiv:2004.11622, 2020.

15. Pakhomov SV, Ruggieri A, Chute CG. Maximum entropy modeling for mining patient medication status from free text. Proc AMIA Symp. 2002;587-591.

16. Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. J Biomed Inform. 2009;42(5):839-851.

17. Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc. 2011;18(5):552-556.

18. Mowery DL, Jordan P, Wiebe J, Harkema H, Dowling J, Chapman WW. Semantic annotation of clinical events for generating a problem list. AMIA Annu Symp Proc. 2013;2013:1032-1041.

19. Rumeng L, Abhyuday N J, Hong Y. A hybrid neural network model for joint prediction of presence and period assertions of medical events in clinical notes. AMIA Annu Symp Proc. 2018;2017:1149-1158.

20. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform. 2001;34(5):301-310.

21. Szarvas G, Vincze V, Farkas R, Csirik J. The bioscope corpus: annotation for negation, uncertainty and their scope in biomedical texts. Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing. 2008;2008:38–45.

22. Morante R. Descriptive analysis of negation cues in biomedical texts. LREC. 2010;2010:1429-1436.

23. Albright D, Lanfranchi A, Fredriksen A, et al. Towards comprehensive syntactic and semantic annotations of the clinical narrative. J Am Med Inform Assoc. 2013;20(5):922-930.

24. Kumar V, Stubbs A, Shaw S, Uzuner Ö. Creation of a new longitudinal corpus of clinical narratives. J Biomed Inform. 2015;58 Suppl(Suppl):S6-S10.

25. Stubbs A, Kotfila C, Uzuner Ö. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. J Biomed Inform. 2015;58 Suppl(Suppl):S11-S19.

26. Stubbs A, Kotfila C, Xu H, Uzuner Ö. Identifying risk factors for heart disease over time: Overview of 2014 i2b2/UTHealth shared task Track 2. J Biomed Inform. 2015;58 Suppl(Suppl):S67-S77.

27. Mahajan D, Liang JJ, Tsou C-H. Extracting daily dosage from medication instructions in EHRs: an automated approach and lessons learned. arXiv preprint arXiv:2005.10899, 2020.

28. Shoolin J, Ozeran L, Hamann C, Bria W 2nd. Association of medical directors of information systems consensus on inpatient electronic health record documentation. Appl Clin Inform. 2013;4(2):293-303.

29. Liu S, Ma W, Moore R, Ganesan V, Nelson S. RxNorm: prescription for electronic drug information exchange. IT Professional. 2005;7(5):17-23.

30. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825–2830.

31. Neumann M, King D, Beltagy I, Ammar W. ScispaCy: fast and robust models for biomedical natural language processing. ACL Proceedings of the 18th BioNLP Workshop and SharedTask. 2019;2019:319–327.

32. U.S. National Library of Medicine. RxNorm API. Available from: https://rxnav.nlm.nih.gov/RxNormAPIs.html.

33. Hanauer DA, Liu Y, Mei Q, Manion FJ, Balis UJ, Zheng K. Hedging their mets: the use of uncertainty terms in clinical documents and its potential implications when sharing the documents with patients. AMIA Annu Symp Proc. 2012;2012:321-330.

34. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018 Oct 11.

35. Alsentzer E, Murphy JR, Boag W, Weng WH, Jin D, Naumann T, et al. Publicly available clinical BERT embeddings. arXiv preprint arXiv:1904.03323. 2019 Apr 6.

36. Wolf T, Chaumond J, Debut L, Sanh V, Delangue C, Moi A, et al. Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 EMNLP: System Demonstrations. 2020;2020:38-45.