Contextualized Medication Event Extraction

Dibyendu Das

Department of Computer Science
Ramakrishna Mission Vivekananda Educational and Research Institute
Belur Math, Howrah
Pin - 711 202, West Bengal

A thesis submitted to
Ramakrishna Mission Vivekananda Educational and Research Institute
in partial fulfillment of the requirements for the degree of
MSc in Big Data Analytics
2022

Dedicated to ...

This Project is dedicated to my mom whose unconditional love and support inspires me to always do my best, be kind, laugh often and take nothing for granted .

Acknowledgements

I would like to express my sincere gratitude to several individuals for supporting me throughout my Internship. First, I wish to express my sincere gratitude to my supervisor, Professor Dr. Sriparna Saha, for her enthusiasm, patience, insightful comments, helpful information, practical advice and unceasing ideas that have helped me tremendously at all times in my research and writing of this thesis. Her immense knowledge, profound experience and professional expertise in Data Science has enabled me to complete this research successfully. Without her support and guidance, this project would not have been possible. I could not have imagined having a better supervisor in my study. I also wish to express my sincere thanks to Ramakrishna Mission Vivekananda Educational and Research Institute for guiding me in all scenarios.
Finally, last but by no means least; also to everyone in the Research institute for Data Science it was great sharing premises with all of you during last Two years.

Thanks for all your encouragement!

Ramakrishna Mission Vivekananda Educational and Research Institute, Belur Math, West Bengal

Dibyendu Das

June 30, 2022 (replace with correct date)

CERTIFICATE FROM THE SUPERVISOR

This is to certify that the thesis entitled *'Contextualized Medication Event Extraction'* submitted by *Mr.Dibyendu Das*, who has been registered for the award of MSc in Big Data Analytics degree of Ramakrishna Mission Vivekananda Educational and Research Institute, Belur Math, Howrah, West Bengal is absolutely based upon his own work under the supervision of *Dr. Sriparna Saha* of Department of Computer Science and Engineering Indian Institute of Technology Patna and that neither his thesis nor any part of the thesis has been submitted for any degree/diploma or any other academic award anywhere before.

Dr. Sriparna Saha
Emeritus Professor
Department of Computer Science and Engineering
Indian Institute of Technology Patna
Bihta, Patna -801103 (Bihar)

Abstract

Since the development of deep learning models, the field of clinical natural language processing has improved tremendously. Many natural language processing applications have adopted the transfer learning paradigms, particularly in contexts where high-quality manually annotated data is scarce. Electronic health record systems are widely used, and the bulk of patient data is now captured electronically, particularly as free text. Medical concept identification and information extraction is a difficult process, but it is an essential component of parsing unstructured data into an organised and tabulated format for downstream analytical operations.To have a thorough view of a patient's medication history, it's crucial to understand medication events in clinical notes. While previous research has looked into identifying medication changes in clinical notes, the longitudinal and narrative nature of clinical documentation means that extracting medication changes without the necessary clinical context is insufficient for use in real-world applications like medication timeline generation and reconciliation.Here, we present a framework to capture multi-dimensional context of medication changes documented in clinical notes .

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The ability to assess the appropriateness of current therapies, recognise potential drug-related diseases or symptoms, and deliver effective medical care requires an accurate medication history and help to guide future treatment options Although medication information is recorded in a variety of clinical data sources.Due to improved access in electronic health record systems, clinicians frequently depend on standardised drug orders in practise.Many pharmaceutical incidents, on the other hand, are only reported in unstructured clinical notes which offer more information. However, they are difficult to look through, particularly at the point of treatment. A medication change may be captured in clinical narratives but not in organised medication data in numerous instances in clinical practise. If the patient already has a drug, for example, the physician may urge the patient to alter the dosage or temporarily hold the medication without notifying the patient. Patient-initiated medication changes are also infrequently reported in organised medication data, but would be in unstructured clinical narratives. To obtain a comprehensive and complete picture of the patient's medication history, pharmaceutical event information must be captured from unstructured data inside the patient medical record. Because of the longitudinal and narrative character of clinical documentation, it is crucial to consider the surrounding contextual information when extracting drug changes from clinical text. The longitudinal quality of clinical text, in particular, comes in the documenting of occurrences throughout the patient's medical history, from past and current events through future probable events. Furthermore, while documenting a clinical contact with a patient, physicians may include their clinical reasoning behind any medical decisions, including why some treatment alternatives were postponed,

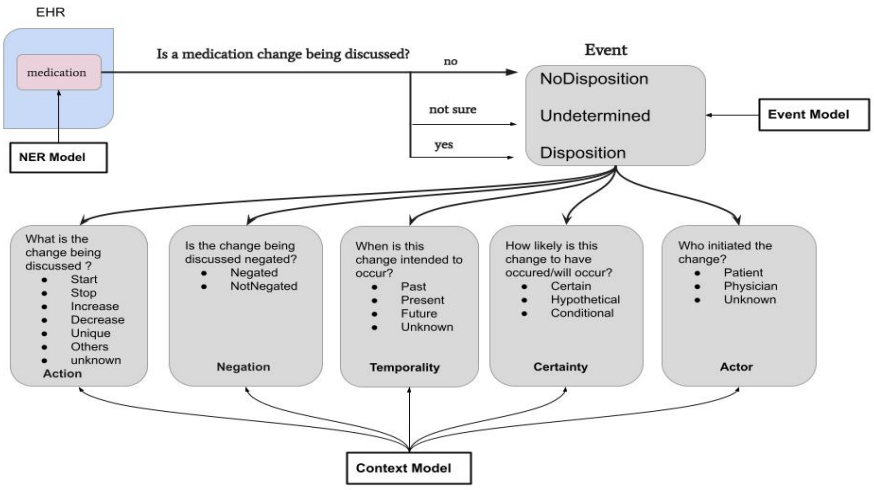


Figure 1.1: Problem statement of N2C2 2022 share task track-1

in addition to recording what has transpired in the patient and the plan forward. Complex clinical occurrences would be poorly recorded by extracting drug changes alone without consideration of the surrounding clinical context due to such features of clinical recording. This is especially true when creating a medication change extraction system for real-world applications like medication timeline development and reconciliation. To create a medicine timeline, for example, a system must extract not only the dose adjustment action, but also the right moment in time, i.e. when the activity is taking place. Medication reconciliation, meanwhile, necessitates information of not just what prescriptions were given by healthcare practitioners, but also whether the patient is taking (or not taking) a medication, i.e. patient-initiated activities.

In this work we offer a conceptual framework to organise multi-dimensional context for medication events in clinical narratives to meet the requirement for contextual information to be considered during medication change extraction.

1.1 Problem Statement :

Using the Contextualized Medication Event Dataset (CMED), the overall task of this track is to identify all medication mentions within a clinical note, indicate whether a change has been/is being discussed, and classify change events along 5 contextual dimensions. The overall structure is shown in the above diagram.

Chapter 2

Literature Review

Various use cases drove previous attempts to characterise drug change occurrences in clinical notes. Some previous studies focused solely on individual prescriptions, while others focused on specific types of pharmaceutical modifications but applied to all medications.Work on warfarin (labels on or stop) and other drugs has been done in the past. stop)[1], heart failure drugs (active, discontinued, or negative labels)[2], beta blockers (active, discontinued, or negative labels)[3],dietary supplements (labels continuing, discontinued, started, unclassified)[4][5].In contrast to earlier studies that looked at medication changes across the board, Liu et al. focused solely on drug discontinuation occurrences [6], Sohn et al. looked into a broader variety of drug modifications (labels start, stop, increase,decrease, no-change)[7], Lerner et al. additionally incorporated labels for sequential changes (start, start+stop, halt, continue, switch, drop, increase)[8]. Pakhomov et al. added temporal information to one of their labels, and (labels past, continuing, stop, start, not classified) [9]. Note that these studies have resulted in a mixed set of labels, some of which may not cover all types of changes (e.g., Pakhomov et allabels .'s - past, continuing, stop, start)[9], do not differentiate between increase and decrease) or do not cover all aspects of a medication event (e.g., Sohn et allabels .'s - start, stop, increase, decrease, no-change)[7], cover all types of changes but do not provide temp . Other researchers have attempted to understand context or statements for medical concepts like issues and tests[10]. Attempts to discover denied medical ideas in clinical text [11] have also been made.Although several of these studies highlight features of medical events such as certainty and denial, none of them have been applied to drug change events.Furthermore, no previous research has attempted to identify

the actor responsible for an incident, which is particularly essential in the case of drug changes because of the potential for patient noncompliance. As a result, a more well-organized schema of label definitions is required to better contextualise drug occurrences. In various areas, our research varies from past studies. First, in addition to recognising drug changes in clinical narratives, we collect pertinent contextual information.

Chapter 3

Methodology

3.1 Text pre-processing and Sentence Segmentation

Electronic health records (EHR) are becoming more widely acknowledged as a significant source of data that may be used for medical research and quality assurance. Clinical information for the patient's diagnostic and treatment pathways is gathered in the EHR. This contains logistical information like appointment times and dates, as well as laboratory test results, prescription lists, and, perhaps most importantly, the physicians' notes on the patients' visits. Basic text cleaning and pre-processing measures were conducted to standardise texts in order to compare the performance of the proposed pharmaceutical extraction model with n2c2 2022 data. Converting data into a suitable form for Machine Learning or Deep Learning to be applied to clinical data is a difficulty.For Segmenting sentence in such a way that the sentences contain most relevant information We used Spark-NLP as a tool .

3.2 Token Classification

A sub-task of information extraction is named-entity recognition, which involves detecting ideas of relevance in unstructured text and categorising them into predetermined categories, such as medicine names, doses, and administration frequency (NER). NER systems are implemented in a variety of ways, from rule-based string matching[12] to complicated Transformer models or hybrid mixes of these models[13]. The goal of the named entity recognition challenge is to forecast the text's medical

mentions. The outcomes are assessed by comparing the set of annotated mention spans inside the document to the set of predicted mention spans by the model. We use the stringent form of precision, recall, and F1-score to assess the findings. When the original datasets did not include such information, we utilised Spark-NLP to break the text into a sequence of tokens to create the dataset. We fin-tuned Bio-Bert for this task . It is a large-scale biomedical corpus-based domain-specific language representation model. BioBERT[14] successfully translates the information of a large number of biomedical texts into biomedical text mining models using the BERT architecture.

3.3 Event and Context Classification :

After finding the medication mentions from the text , we constructed sentences based on context around the medication mentions using Spark-NLP .As in our dataset there are total 7230 medication mentions in this stage we have 7230 sentences for event classification and as there are only 1413 medications of Disposition type we only have 1413 sentences for Context Classification . We divided the work of medication change categorization into five classification subtasks and grouped them in a two-step method to automate it:

- One subtask is to categorise each medicine reference into one of three categories: disposition, no disposition, or undetermined.
- Five subtasks to classify each given Disposition medication along four context dimensions: Action, Temporality,Certainty, and Actor,Negation.

For each subtask, we train a classification model based on two approaches: (1) Machine learning and (2) transformer-based language models.

3.3.1 Machine Learning Approach and Grid Search

In our tests, we chose to employ a feature-based method with a discriminative classification algorithm to examine the dataset and multiple dimensions. Most machine learning algorithms, in general, will not provide optimal results if their parameters are not appropriately set. It's critical to select a powerful machine learning algorithm and fine-tune its parameters if you want to develop a high-accuracy classification model.Originally, a grid search was an exhaustive search based on a defined

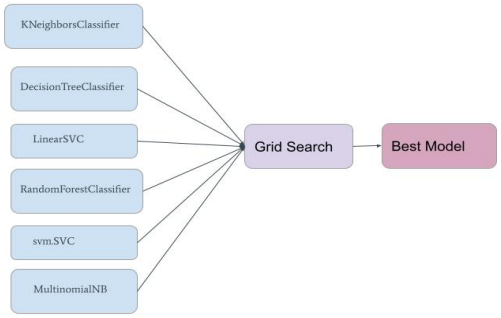


Figure 3.1: Grid Search approach for optimal parameter settings

subset of the hyper-parameter space. Minimal value (lower bound), maximum value (upper bound), and number of steps are used to specify the hyper-parameters.Grid search uses a cross validation (CV) approach as a performance indicator to adjust the Machine Learning Model parameters.We apply grid search method on six different classification algorithm . The Algorithms are K-Nearest Neighbour , SVM ,Dicision Tree , Random Forest, Multinomial NB,SGD Classifier.For Sentence Segmentation we used Spark-NLP as a tool and for transform each sentance into machine learinig input format we used TF-IDF for feature extraction(Scikit learn library).Using Grid Search method we choose best parmeters for our model performence .For handeling class imbalance problem we used oversampling method. We use Grid Search for both event classification and context classification task . The Results are shown in table II .

3.3.2 Deep Learning Approach :

Pre-training using language models has been shown to help people learn universal language representations. BERT (Bidirectional Encoder Representations from Transformers)[15] has produced excellent results in several language understanding tests as a state-of-the-art language model pre-training model. The challenge of text categorization is a well-known one in Natural Language Processing (NLP). The goal is to categorise a set of text sequences into predetermined groups. Text representation is a critical intermediary step.BERT-base model contains an encoder with 12

Transformer blocks, 12 self-attention heads, and the hidden size of 768. BERT takes an input of a sequence of no more than 512 tokens and outputs the representation of the sequence. The sequence has one or two segments that the first token of the sequence is always [CLS] which contains the special classification embedding and another special token [SEP] is used for separating segments. For text classification tasks, BERT takes the final hidden state h of the first token [CLS] as the representation of the whole sequence. A simple softmax classifier is added to the top of BERT to predict the probability of label c. In this work we first go for a simple and less complex model like Text-CNN , Bi-Lstm then we used Bert Based Complex Tranformer Models. We tested state-of-the-art Bidirectional Encoder Representations from Transformer (BERT34)-based language models with datasets from both the general domain (BERTbase34) and the clinical domain (ClinicalBERT)[16]. Each of the five tasks was reformulated as a sentence classification task, with the annotated medication's surrounding sentence serving as the context. We used the pre-trained transformer to get a distributed representation for each of the five subtasks in this procedure. We tuned our models using the train and development splits, then presented our findings on the test split, using the transformers package[17].

3.4 Class Imbalance :

Data imbalance is a typical problem in a number of NLP tasks, such as tagging and context classification, and it affects the objective function equally, although the F1 score is more concerned with positive cases at test time.The following two concerns arise as a result of the data imbalance: (1) the disparity between training and testing: Without balancing the labels, the learning process tends to converge to a point where the dominant label severely biases the learning process. This results in a mismatch between training and testing: Each training instance contributes equally to the objective function during training, but F1 gives equal weight to positive and negative instances during testing.(2) the dominance of simple negative instances. As Meng et al. (2019)[18] point out, a significant number of negative instances also entails a significant number of easy-negative examples. The large quantity of simple cases tends to overload the model, preventing it from learning to discriminate between positive and hard-negative situations. Form Table I we can see that the classes and their sub-classes having the issue of class imbalance problem. To handle this issues we did (i) Oversampling (ii) Text Augmentation (iii) Penalize Loss using

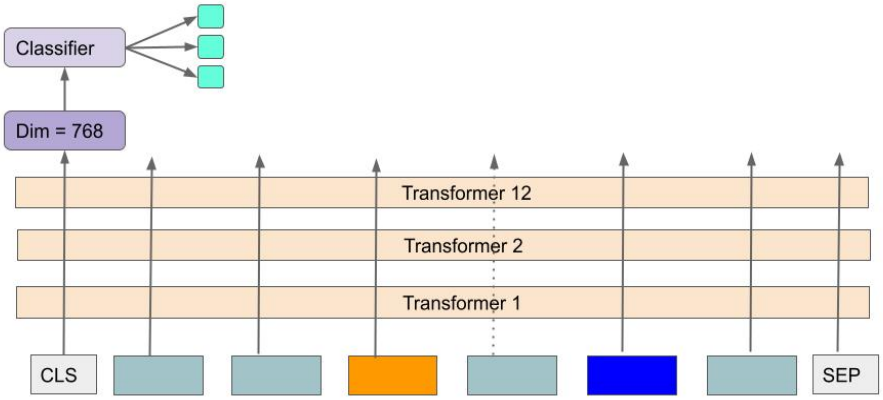


Figure 3.2: Bert Base Architecture for Classification

Reward . (iv) Different Loss Functions

3.4.1 Oversampling

The data level approach adjusts for unbalanced datasets by altering the minority class (oversampling) and the majority class (undersampling) (undersampling). Oversampling is the process of altering the data distribution so that samples appear depending on the determined cost. In other words, unless the data distribution is proportionate to their costs, this strategy replicates higher-cost training data. we used a WeightedSampler, so that the batches have all classes with equal probability. Give an equal sort of weight to the dataset.

3.4.2 Text Augmentation

Because text data is composed of words, we may perform operations on them, such as word replacement, word swapping, word deletion, and word insertion we also apply charcter level transformation. We primarily concentrate on word substitution

and deletion, as well as character insertion and deletion. We could synthesis fresh data based on the original data for word replacement and charecter replacement. More importantly, we want the newly created data to be distinct from the original while yet having the same meaning. We modify the structure of a sentence by eliminating part of the words, and then we have a new sentence. In a nutshell, enhanced data has the potential to boost data variety. We performed text augmentations in our NLP tasks, to generate a semantically invariant transformation of the textual data.We used NLPAUG Library for this task .we apply the following augmentation techniques to our text data - (i) Ocr Augmentation (ii) Keybord Augmentation (iii) Random Augmentation (iv) Spelling Augmentation (v) TfIdf Augmentation.

3.4.3 Diffrent Loss Function

Choosing how to weight the loss for different classes can be difficult when working with a long-tailed dataset (one in which the majority of the samples belong to a small number of classes and many other classes have very little support). The weighting is frequently set to the inverse of class support or the inverse of class support squared. here handling this Class Imbalance issue we used Dice Loss ,weighted Cross-entropy Loss , Focal Loss.Detailed Result is shown in table –

3.4.4 Penalize Loss

In this section we discuss how we penalize the loss for each batch in the dataloader . As our dataset having the issue of high class imbalance problem for handle this issue we calculate macro-F1 score for each bath in the train dataloader and we construct reward as inverse of macro-f1 score . As for some batch macro-F1 score may be zero for that reson we add small ϵ with it . The detailed algorithm is shown in **Algorithm 1**.

3.4.5 Architecture:

In this section we are going to describe our overall architecture for N2C2 2022 challenge. From a given Prescription we find the medication mentions using Bio-BERT model then after finding the medication mentions we construct sentences based on that medication mentions and pass those sentences into event model for event classification .we construct our event model using Clinical Bert. From the output of

Algorithm 1 Algorithm for Reward

Input: Batches from Train Dataloader
Output: Penalized Loss

Setting :
Class = $\{c_1, c_2, \cdots, c_T\}$, Batch_size = l, loss_fun = loss()
Training_Data = $\{(x_i, y_i) \ : \ i = 1(1)n, \ y_i \in Class\}$
Batch = $\{\{(x^{(j)}, y^{(j)}) \ : \ j = 1(1)l\}_k \ : \ k = 1(1)[n/l]\}$
model = model()

1: **for** for item in Batch **do**
2: input = $\{x^{(j)} \ : \ j = 1(1)l\}$
3: label = $\{y^{(j)} \ : \ j = 1(1)l, \ y^j \in Class\}$
4: $\{\hat{y}^{(i)} \ : \ j = 1(1)l\}$ = model(input)
5: calculate precison P_t, R_t for each $t \in Class$
6: Macro_Score = $\frac{1}{|Class|} \sum \frac{2P_tR_t}{P_t+R_t}$
7: Reward = $\frac{1}{Macro\_Score} + \epsilon(noise)$
8: loss = loss _fun (output, label)
9: loss = loss + Reward (Penalized Loss)
10: **end for**=0

clinical Bert model those medication which were classified as Disposition Events we again construct sentences based on that particular disposition medications. After that we pass those sentence to five clinical bert model for five different context classification.

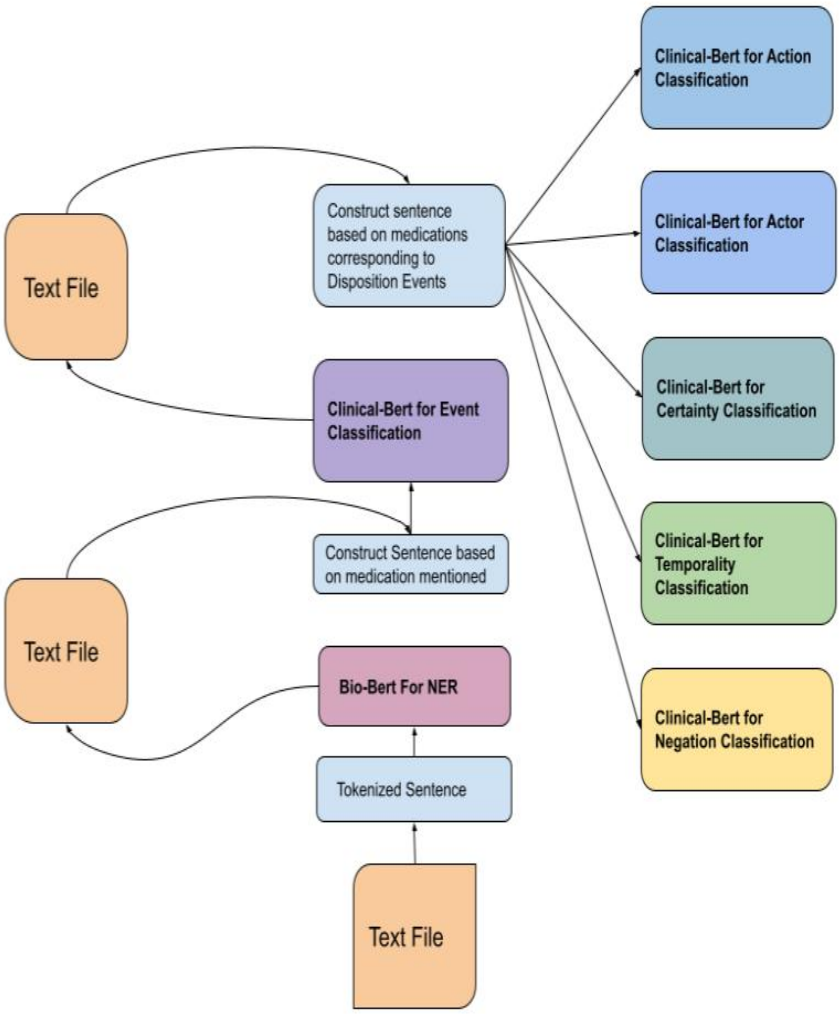


Figure 3.3: Submission Pipeline for N2C2 Task

Chapter 4

Experimental Evaluation

4.1 Description of Data

The annotated data set was sourced from 2014 i2b2/ UTHealth Natural Language Processing shared task corpus,which contain a total of 1,304 clinical notes over 296 patients as part of the Track 1 of The 2022 National NLP Clinical Challenges (n2c2) Shared Task on Contextualized Medication Event Extraction.The data set was made up of a collection of Intensive Care Unit discharge letters that included a wealth of information regarding the drugs used in therapy. The final corpus included 500 notes from 296 patients, 120 of which were double-annotated and adjudicated to determine inter-annotator agreement (IAA).The procedure for annotating is as follows: The annotator evaluates if a medication change is being discussed for each drug mention and gives one of the following medication change event labels:

- **NoDisposition :** no medication change is being discussed, e.g. "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)

The final dataset, has 9,013 annotated medicine references spread over 500 clinical notes. The training dataset includes 7,230 annotated medicine references spread across 400 notes, with the distribution of particular labels shown in Table 1.

As observed in Table 1, 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.

4.2 Evaluation Techniques

The F1 score (aka F-measure) is the metric for evaluating the performance of our NER and classification model.In the case of multi-class classification, we adopt aver-

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

Table 4.1: Label distribution for the training dataset

aging methods for F1 score calculation, resulting in a set of different average scores (macro,micro) in the classification report.

4.3 Experimental Setup

In table 4.2 we give information about Tools and Python Versions and in table 4.3 we describe our hyper parameter settings .

4.4 Analysis of Results

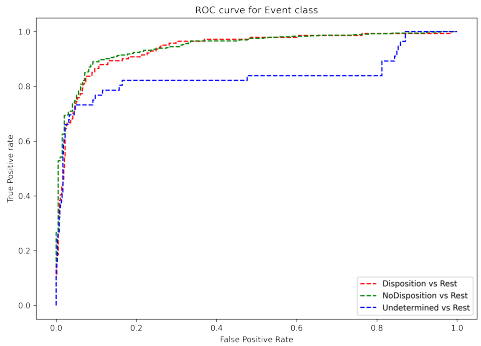
The experimental results may be presented in different tables followed by detailed analysis on these results.

Python: Spark-NLP, SciSpacy Python 3.8 , Pytorch 1.11.0
Device: NVIDIA Quadro GV100(32 GB) , NVIDIA A100 (80GB)

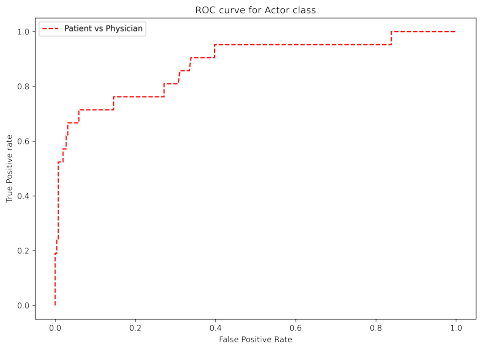
Table 4.2: Tools

Task	Hyper Parameter	Value
NER	Token Length	128
	Batch Size	32/64
	optimizer	AdamW
	Learning Rate	3e-5
	Epochs	5
EVENT Classification	Token Length	200
	Batch Size	32
	optimizer	AdamW
	Learning Rate	5e-5
	Epochs	20
Action Classification	Token Length	150
	Batch Size	32
	optimizer	AdamW
	Learning Rate	5e-5
	Epochs	30
Actor Classification	Token Length	200
	Batch Size	32
	optimizer	AdamW
	Learning Rate	5e-5
	Epochs	30
Temporality Classification	Token Length	200
	Batch Size	32
	optimizer	AdamW
	Learning Rate	5e-5
	Epochs	20
Certainty Classification	Token Length	100
	Batch Size	32
	optimizer	AdamW
	Learning Rate	5e-5
	Epochs	20

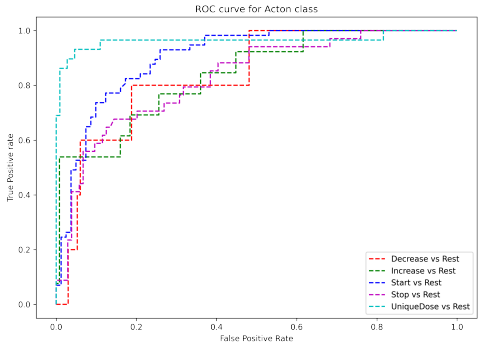
Table 4.3: Hyper Parameter settings



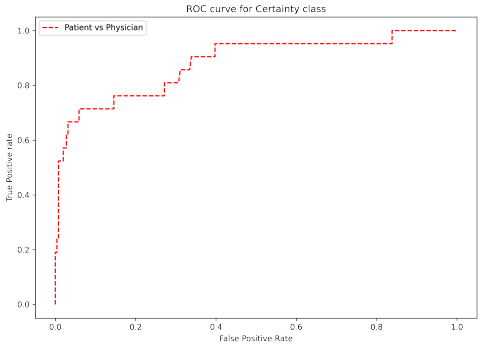
(a) 1a



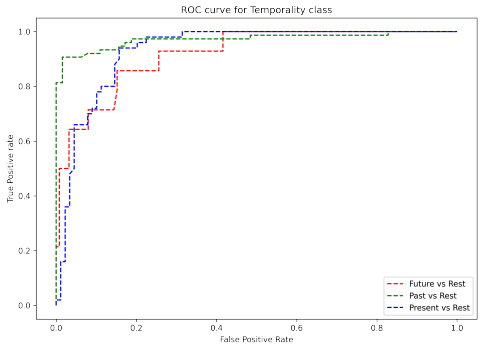
(b) 1b



(c) 1c



(d) 1d



(e) 1e

Figure 4.1: One Vs All Roc Curve (1a)Roc Curve for Event classification (1b) Roc Curve for Actor classification (1c) Roc Curve for Action classification (1d) Roc Curve for Certainty classification (1e) Roc Curve for Temporality classification

NER Result				
Experiment		Pre	Recall	F_1
Bio-Bert	Medicine-tag	0.98	0.98	0.98
	Other-tag	0.99	0.99	0.99
	Micro	0.99	0.99	0.00
	Macro	0.99	0.99	0.00
	Weighted	0.99	0.99	0.00

Table 4.4: NER result

Action				
Experiment		Pre	Recall	F_1
BenchMark-SVM	micro	0.59	0.59	0.59
	macro	0.50	0.51	0.50
Ours - RFC	micro	**0.67**	**0.68**	**0.67**
	macro	**0.63**	**0.57**	**0.58**
Actor				
Experiment		Pre	Recall	F_1
BenchMark-SVM	micro	0.88	0.88	0.88
	macro	0.63	0.68	0.65
Ours-KNeighbors	micro	**0.92**	**0.94**	**0.92**
	macro	0.83	0.62	0.65
Temporality				
Experiment		Pre	Recall	F_1
BenchMark-SVM	micro	0.71	0.71	0.71
	macro	0.60	0.59	0.59
Ours-RFC	micro	**0.79**	**0.79**	**0.78**
	macro	**0.74**	**0.65**	**0.67**
Certainty				
Experiment		Pre	Recall	F_1
BenchMark-SVM	micro	0.83	0.83	0.83
	macro	0.59	0.53	0.56
Ours-KNeighbors	micro	**0.87**	**0.88**	**0.87**
	macro	**0.74**	**0.62**	**0.66**
Negation				
Experiment		Pre	Recall	F_1
BenchMark-SVM	micro	-	-	-
	macro	-	-	-
Ours-RFC	micro	0.49	0.50	0.49
	macro	0.96	0.98	0.97

Table 4.5: Grid Search Result

Action				
Experiment		Pre	Recall	F_1
BenchMark-BERT	micro	0.75	0.75	0.75
	macro	0.75	0.62	0.64
Ours - BERT+Reward	micro	0.73	0.73	0.73
	macro	0.67	0.65	**0.65**
BenchMark-Clinical-BERT	micro	0.75	0.75	0.75
	macro	0.75	0.63	0.65
Ours - Clinical-BERT+Reward	micro	0.75	0.74	0.74
	macro	**0.68**	**0.66**	**0.67**
Actor				
Experiment		Pre	Recall	F_1
BenchMark-BERT	micro	0.92	0.92	0.92
	macro	0.79	0.72	0.75
Ours-BERT+Reward	micro	**0.94**	**0.94**	**0.94**
	macro	0.77	**0.75**	**0.76**
BenchMark-Clinical-BERT	micro	0.93	0.93	0.93
	macro	0.83	0.72	0.76
Ours-Clinical-BERT+Reward	micro	**0.94**	**0.95**	**0.94**
	macro	**0.86**	**0.73**	**0.78**
Temporality				
Experiment		Pre	Recall	F_1
BenchMark-BERT	micro	0.81	0.81	0.81
	macro	0.77	0.71	0.73
Ours-BERT+Reward	micro	**0.83**	**0.81**	**0.82**
	macro	0.73	0.72	0.72
BenchMark-Clinical-BERT	micro	0.83	0.83	0.83
	macro	0.80	0.74	0.75
Ours-Clinical-BERT+Reward	micro	**0.87**	**0.86**	**0.86**
	macro	**0.86**	**0.76**	**0.79**
Certainty				
Experiment		Pre	Recall	F_1
BenchMark-BERT	micro	0.90	0.90	0.90
	macro	0.83	0.74	0.77
Ours-BERT+Reward	micro	0.90	0.90	0.90
	macro	0.76	0.78	0.77
BenchMark-Clinical-BERT	micro	0.90	0.90	0.90
	macro	0.83	0.76	0.79
Ours-Clonical-BERT+Reward	micro	**0.92**	**0.92**	**0.92**
	macro	**0.84**	**0.78**	**0.81**
Event				
Experiment		Pre	Recall	F_1
BenchMark-BERT	micro	0.88	0.88	0.88
	macro	0.79	0.78	0.77
Ours-BERT+Reward	micro	0.87	0.86	0.87
	macro	0.75	0.77	0.76
BenchMark-Clinical-BERT	micro	0.88	0.88	0.88
	macro	0.79	0.79	0.79
Ours-Clinical-BERT+Reward	micro	0.88	0.88	0.88
	macro	0.78	0.79	0.79

Table 4.6: Results for Transfer Learning based approach

Chapter 5

Discussions

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.

5.1 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

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 De-

crease). 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.

Error Category	Example	Medication	Ground Truth	Prediction
Medicatio 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.	glyburide	Conditional	Certain

Table 5.1: Common error categories with examples, across five classification subtasks for the ClinicalBERT model.

5.2 Limitation

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[19],[20],[21]. 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.

Chapter 6

Conclusions and Scope of Further Research

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[22],[23]. 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 6.1 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.

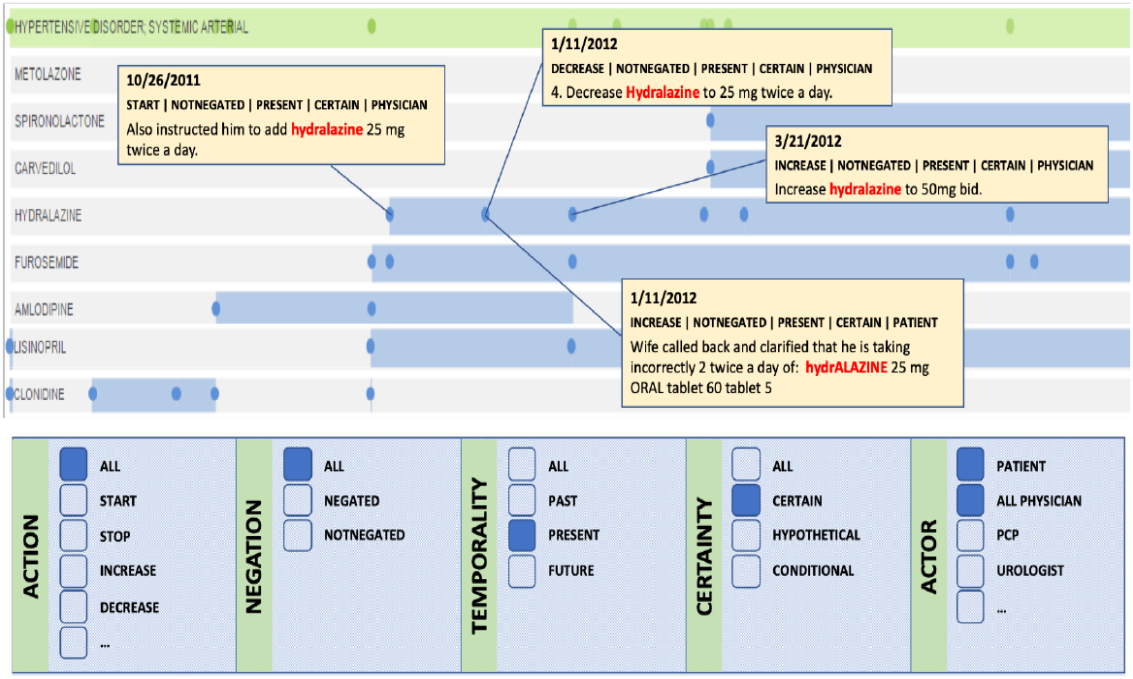


Figure 6.1: Prototype visualization incorporating extracted medication change events into structured EHR data showing all Present Certain events associated with the medication hydralazine.

Appendix A

Additional Work

Additionally since last two months I have been actively involved in another project on Skin Lesion Classification using Sugeno Fuzzy integral.

Here we are using ISIC 2019 Skin Lesion Classification Challenge Dataset, The objective of the challenge is to categorise skin lesions using dermoscopic pictures. For training, a heterogeneous dataset of 25000 photos from eight classes was made available. Additionally, a serious class imbalance is a concern with multi-class skin lesion categorization. By implementing a reward function, we attempt to solve this issue.In order to limit the model's emphasis to particular characteristics, we have additionally modified pre-processing approaches to eliminate noise.Convolutional neural networks (CNNs) are used as the basic model in this paper's ensemble approach, which is based on Sugeno and Choquet Fuzzy Integrals.

Bibliography

[1] Mei Liu, Min Jiang, Vivian K Kawai, Charles M Stein, Dan M Roden, Joshua C Denny, and Hua Xu. Modeling drug exposure data in electronic medical records: an application to warfarin. In *AMIA annual symposium proceedings*, volume 2011, page 815. American Medical Informatics Association, 2011.

[2] Stéphane M Meystre, Youngjun Kim, Julia Heavirland, Jenifer Williams, Bruce E Bray, and Jennifer Garvin. Heart failure medications detection and prescription status classification in clinical narrative documents. *Studies in health technology and informatics*, 216:609, 2015.

[3] L Ohno-Machado and B Séroussi. Automatic methods to extract prescription status quality measures from unstructured health records. In *MEDINFO 2019: Health and Wellbeing e-Networks for All: Proceedings of the 17th World Congress on Medical and Health Informatics*, volume 264, page 15. IOS Press, 2019.

[4] Yadan Fan, Lu He, and Rui Zhang. Classification of use status for dietary supplements in clinical notes. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1054–1061. IEEE, 2016.

[5] Yadan Fan and Rui Zhang. Using natural language processing methods to classify use status of dietary supplements in clinical notes. *BMC medical informatics and decision making*, 18(2):15–22, 2018.

[6] Yadan Fan and Rui Zhang. Using natural language processing methods to classify use status of dietary supplements in clinical notes. *BMC medical informatics and decision making*, 18(2):15–22, 2018.

[7] Sunghwan Sohn, Sean P Murphy, James J Masanz, Jean-Pierre A Kocher, and Guergana K Savova. Classification of medication status change in clinical nar-

ratives. In *AMIA Annual Symposium Proceedings*, volume 2010, page 762. American Medical Informatics Association, 2010.

[8] Ivan Lerner, Jordan Jouffroy, Anita Burgun, and Antoine Neuraz. Learning the grammar of prescription: recurrent neural network grammars for medication information extraction in clinical texts. *arXiv preprint arXiv:2004.11622*, 2020.

[9] Serguei V Pakhomov, Alexander Ruggieri, and Christopher G Chute. Maximum entropy modeling for mining patient medication status from free text. In *Proceedings of the AMIA Symposium*, page 587. American Medical Informatics Association, 2002.

[10] Henk Harkema, John N Dowling, Tyler Thornblade, and Wendy W Chapman. Context: an algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of biomedical informatics*, 42(5):839–851, 2009.

[11] Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5):552–556, 2011.

[12] P. Raghavan H. Schutze, C. D. Manning. Introduction to information retrieval. *Proceedings of the international ¨ communication of association for computing machinery conference*, 4:552–556, 2008.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[14] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[16] Emily Alsentzer, John R Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. Publicly available clinical bert embeddings. *arXiv preprint arXiv:1904.03323*, 2019.

[17] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*, 2019.

[18] Yijin Liu, Fandong Meng, Jinchao Zhang, Jinan Xu, Yufeng Chen, and Jie Zhou. Gcdt: A global context enhanced deep transition architecture for sequence labeling. *arXiv preprint arXiv:1906.02437*, 2019.

[19] Vishesh Kumar, Amber Stubbs, Stanley Shaw, and Özlem Uzuner. Creation of a new longitudinal corpus of clinical narratives. *Journal of biomedical informatics*, 58:S6–S10, 2015.

[20] Amber Stubbs, Christopher Kotfila, and Özlem Uzuner. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/uthealth shared task track 1. *Journal of biomedical informatics*, 58:S11–S19, 2015.

[21] Amber Stubbs, Christopher Kotfila, Hua Xu, and Özlem Uzuner. Identifying risk factors for heart disease over time: Overview of 2014 i2b2/uthealth shared task track 2. *Journal of biomedical informatics*, 58:S67–S77, 2015.

[22] Catherine Plaisant, Richard Mushlin, Aaron Snyder, Jia Li, Dan Heller, and Ben Shneiderman. Lifelines: using visualization to enhance navigation and analysis of patient records. In *The craft of information visualization*, pages 308–312. Elsevier, 2003.

[23] Jeffery L Belden, Pete Wegier, Jennifer Patel, Andrew Hutson, Catherine Plaisant, Joi L Moore, Nathan J Lowrance, Suzanne A Boren, and Richelle J Koopman. Designing a medication timeline for patients and physicians. *Journal of the American Medical Informatics Association*, 26(2):95–105, 2019.