Conceptualisation and Annotation of Drug Nonadherence Information for Knowledge Extraction from Patient-Generated Texts

Anja Belz
University of Brighton, UK
a.s.belz@brighton.ac.uk

Elizabeth Ford
Brighton and Sussex Medical School, UK
e.m.ford@bsms.ac.uk

Richard Hoile
Sussex Partnership NHS Foundation Trust, UK
Richard.Hoile@sussexpartnership.nhs.uk

Azam Mullick
University of Brighton, UK
a.mullick@brighton.ac.uk

Abstract

Approaches to knowledge extraction (KE) in the health domain often start by annotating text to indicate the knowledge to be extracted, and then use the annotated text to train systems to perform the KE. This may work for annotating named entities or other contiguous noun phrases (drugs, some drug effects), but becomes increasingly difficult when items tend to be expressed across multiple, possibly noncontiguous, syntactic constituents (e.g. most descriptions of drug effects in user-generated text). Other issues include that it is not always clear how annotations map to actionable insights, or how they scale up to, or can form part of, more complex KE tasks. This paper reports our efforts in developing an approach to extracting knowledge about drug nonadherence from health forums which led us to conclude that development cannot proceed in separate steps but that all aspects—from conceptualisation to annotation scheme development, annotation, KE system training and knowledge graph instantiation—are interdependent and need to be co-developed. Our aim in this paper is two-fold: we describe a generally applicable framework for developing a KE approach, and present a specific KE approach, developed with the framework, for the task of gathering information about antidepressant drug nonadherence. We report the conceptualisation, the annotation scheme, the annotated corpus, and an analysis of annotated texts.

1 Introduction

Depression is experienced by 1 in 4 people in the UK. More than two thirds of patients are mostly managed with antidepressant medication, yet nonadherence rates are very high. One study found that 4.2% of patients who were prescribed antidepressants did not take them at all, and 23.7% filled only a single prescription (van Geffen et al., 2009). Nonadherence is a major obstacle in the effective treatment of depression, but cannot currently be predicted or explained adequately (van Dulmen et al., 2007). An influential WHO report (Sabaté et al., 2003) concluded: "[i]ncreasing the effectiveness of adherence interventions may have a far greater impact on the health of the population than any improvement in specific medical treatments". Nonadherence is hard to investigate via controlled studies meaning alternative sources of information are needed. Recent results indicate a strong signal relating to usage of psychiatric medications on health forums and social media (Tregunno, 2017), and that social media users report nonadherence and reasons for it (Onishi et al., 2018).

The work reported in this paper aimed (i) to develop a conceptualisation of the information space around drug nonadherence, defining the relevant concepts, properties and relations; (ii) to develop an annotation scheme based on the conceptualisation; (iii) to annotate a corpus of depression health forum posts with the scheme; and (iv) to use the annotated data to examine the prevalence, and co-occurrence, of different kinds of nonadherence information (testing signal strength). We examine the interdependent and mutually constraining relationship between conceptualisation, annotation scheme and knowledge extraction processes.

The ultimate goal is to perform automatic knowledge extraction (KE) in order to provide valuable non-adherence information from a large sample about why and how non-adherence occurs. We hope this will in turn lead to better prescribing, better adherence and more informed discussions between patient and prescriber around medication.

2 Nonadherence and Health Forums

Different terms have been used to describe the "suboptimal taking of medicine by patients" (Hugtenburg et al., 2013). Among these, non-compliance and nonadherence both mean *not tak-*

Quitting your meds can be awful, for sure. I started trying to get off Zoloft by myself 5 years ago. I went cold turkey which was a complete disaster. Next I tried cutting down by large amounts which was slightly less of a disaster. Finally I tapered off very slowly. ONE WHOLE YEAR to get from 200mg to zero. With the tapering off the brain zaps were much less severe. Simply takes a long time to wean yourself off these drugs.

I reduced my mirtazapine from 45mg first to 30mg, and then to 15mg, then stopped it altogether. After two weeks I was feeling awful so I decided to restart it at 15mg.

So I stopped Fluoxetine about a month ago sort of by accident. After I missed a few doses I just decided to keep going. So far I've only had rather minor symptoms. One of my symptoms has been an electric shock sensation from my brain down my spine/body. This happens especially when I get up, sit down or move suddenly.

Last week my doctor increased my Lamictol dosage to 200 MG. I am beginning to notice serious cognitive deficits. For example constantly losing/misplacing things.. Has anyone else experienced this situation?

Figure 1: Four example depression forum posts created to closely resemble real posts.

ing a drug as instructed (the intended meaning in the present context), but nonadherence is the term now preferred as reflecting a more equal prescriber-patient relationship (Hugtenburg et al., 2013). Two types of nonadherence are distinguished, *intentional*, where a patient "actively decides" not to follow instructions, and *unintentional* nonadherence, including forgetting and not knowing how to take a drug (Hugtenburg et al., 2013).

Consider the four example posts from health forums for specific antidepressants in Figure 1. Some of the sentences contain explicit statements that the modifications described were instigated by the patient ("I started trying to get off Zoloft by myself 5 years ago"; "After two weeks I was feeling awful so I decided to restart it at 15mg"; "So I stopped Fluoxetine about a month ago sort of by accident"). Other modifications described in the first three posts (unlike in the fourth) are likely to also have been instigated by the patient, but clinician involvement cannot be ruled out.

These are typical examples of how patients talk about nonadherence in health forums: explicit statements ('my doctor told me to do one thing, but I did another') are rare (7% of posts on the depression forums we have been looking at, see also Section 5). More typically, a drug modification is described along with the side effect and/or other reason(s) that gave rise to it, but the extent to which the prescribing physician was involved in deciding to make the modification can only be inferred, with varying degrees of certainty, from the language, or on the basis of medical knowledge (e.g. a modification is known to be dangerous).

3 Conceptualisation and KE Task

Posts like the ones in Figure 1 clearly contain information about the *why* and *how* of drug nonad-

herence, but how can it be automatically extracted and rendered useful? In this section we discuss the main issues in developing a knowledge extraction (KE) approach for a specific domain and a specific KE task, nonadherence event extraction in our case. To introduce the different components in developing an approach to KE (overview see Figure 2, we use as an illustrative running example the simpler task of drug effect extraction which is—in contrast to the far more complex task of nonadherent event extraction we are addressing here—already a well established research task[1] (Leaman et al., 2010; Nikfarjam et al., 2015). Drug effect detection is an important subtask of nonadherence event extraction, because drug effects as perceived by the patient play an important role in nonadherence and are often the reason for it. The last post in Figure 1 is a typical example, containing a claim that a drug referred to as "Lamictol", is causing an effect described as "cognitive deficits".

3.1 From Text to Meaning

Suppose that we have a method capable of identifying drug and effect mentions in text and determining which ones are linked (i.e. which drug causes which effect), and that we extract a linked drug and effect pair from the fourth post in Figure 1 for which we could choose the following notation: *cause("Lamictol","cognitive deficits")*. Suppose from this and other posts we extract five such linked pairs as follows:

> *cause("Lamictol","serious cognitive deficits")*
> *cause("Lamictol","constantly losing/misplacing things")*
> *cause("Lamictal","uncoordinated")*   (1)
> *cause("lamotrigine","so forgetful")*
> *cause("Lamotrigene","keep losing my phone and going upstairs and forgetting what for")*

---

[1]See e.g. Task 2 at this year's SMM4H Shared Task event: https://healthlanguageprocessing.org/smm4h/challenge/

If this was the actual output of our method, it would be nothing more than a list of drug-effect mentions in a given set of texts, possibly with counts of multiple occurrences of identical mentions (none in Example 1 above). One important type of knowledge would be entirely inaccessible, namely that the above five pairs in fact all claim the same side effect for the same drug (lamotrigine, a mood-stabilising medication sometimes prescribed for depression).[2] In order to extract that knowledge (crucial to be able e.g. to act upon side effect reports depending on novelty or report frequency), the identified word strings need to be mapped to a more abstract level of representation where knowledge is encoded in terms of concepts, rather than word strings. It is only once this process, known as entity linking[3] (Han et al., 2011; Hachey et al., 2013; Rao et al., 2013; Smith et al., 2018), has been performed that the five linked pairs above can be interpreted as five mentions of the same drug effect. In this crucial step, we move from extracting word strings (surface representations) to extracting concept structures (meaning representations); from something that can be compared in terms of string similarity and counted, to something that can be incorporated into a knowledge graph and reasoned about.

3.2 Concept Model and KE Template

In the case of drug effect extraction it is clear what we want to extract, and how it is structured. A very simple conceptualisation suffices to express that understanding: a single parent concept, *drug_effect*, consisting of a *drug*, an *effect* and a causal relation, possibly implicit. Depending on application context, further concepts, such as duration or severity, could also be added (these would be required e.g. if the application task was automatic completion of Yellow Card reports). One of many possible notations for this conceptualisation is the following (for a complete concept model the range of possible values of the component concepts would also have to specified):

$$\text{DRUG\_EFFECT } [\textit{DRUG; EFFECT; SEVERITY; DURATION }] \quad (2)$$

The above can be seen as providing both a concept model representing a piece of domain knowledge,
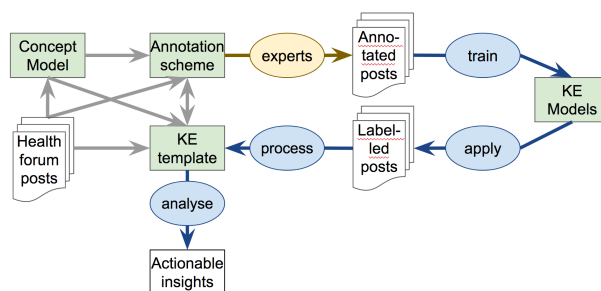

Figure 2: Components in developing KE approach.

and a template to be instantiated by a specific KE tool (depending on the application task, only subsets of concepts might be used for KE). A possible instantiation produced by a KE tool via entity detection and linking for the last post in Figure 1 is the following (initial underscores indicating terminal concepts as opposed to word strings):

$$\begin{aligned}
\text{DRUG\_EFFECT } [\\
\quad \text{DRUG= } \_lamotrigine; \\
\quad \text{EFFECT= } \_confusional\_state; \quad\quad (3) \\
\quad \text{SEVERITY= } \_moderate; \\
\quad \text{DURATION= } \_continuous; \quad\quad]
\end{aligned}$$

In order to produce the above we have to have created a suitable conceptualisation (concept model), a KE template, a KE task construal and methods for implementing it, here detecting word spans corresponding to the above concepts and for mapping the word spans to concepts. In order to be able to do the latter, we also need texts which have been labelled for entities such as drugs and effects and the links between them. All these aspects are shown to the left of Figure 2 which provides an overview of elements and steps involved in developing a KE approach. In the next sections we look at how conceptualisation and annotation scheme interact with possible KE tasks (Section 3.3), followed by issues in determining the details of the annotation scheme (Section 3.4).

3.3 Text Annotation and KE Task

Useful one-off analysis can be conducted on the basis of manual mark-up of text, but for knowledge extraction from *large* quantities of *new* texts, automatic processes are needed. Two common types of KE model are word sequence (post, sentence, phrase, etc.) classifiers and labellers. These tend to be supervised models, i.e. they require labelled training data the creation of which, especially initially, requires human annotation effort.

How the data is annotated limits what kind of KE tasks and models it can be used for. Con-

versely, aiming for a particular KE task has implications for the kinds of annotations that are needed. For the KE task described above, i.e. mapping from health forum posts (Figure 1) to instantiated KE templates (Example 3), the most straightforward way to interpret the annotation task would be to annotate each post (as a whole) with an instantiated KE template for every mention of a drug effect contained in the post. However, this would make the annotators' task cognitively extremely challenging (requiring multiple judgments to be made in conjunction), and very time-intensive (involving look up of concepts in databases and inventories). Moreover, it is not obvious how to define a corresponding KE modelling task and training regime. It is also not feasible to define one output class for each possible instantiated template, because that would lead to an unmanageable combinatorial explosion of classes.

In the relatively simple case of adverse drug effect extraction, the divide-and-conquer approach that tends to be used instead (Nikfarjam et al., 2015; Metke-Jimenez and Karimi, 2016) construes the task, as mentioned above, as a sequence of subtasks, first identifying all mentions of drugs and effects in the text, then linking them to concepts and to each other. For the first step to be possible, drug mentions and effect mentions need to be identified in the text, for which corresponding mark-up needs to be available in the training data. Similarly, any links between the marked up entities also need to be present in the annotations. The identified text strings can then be mapped (linked) in a separate step to drug (e.g. _lamictol) and effect (e.g. _confusional_state) concepts, potentially with separately retrained off-the-shelf tools.

The above discussion points to an annotation scheme involving a DRUG concept (but not concepts for individual drugs), and an EFFECT concept (but not concepts for specific effects), and for corresponding labels to be inserted into texts as mark up. However, more issues arise when mapping these conclusions to an annotation scheme.

3.4 Towards an Annotation Scheme

Some of the questions that arise in text annotation are (1) whether conceptually grounded annotations should attach to word strings (a) with the meaning 'these words together express the given concept', or (b) with the meaning 'somewhere in this text there is an occurrence of the concept';

(2) whether labels should be (a) terminal concepts (e.g. _lamictol) or (b) classes of terminal concepts (e.g. *Drug*); (3) how to treat instances where an entity or event is mentioned, but is not asserted to have occurred or have been observed, which happens e.g. with negation, questions or hypothetical considerations; and (4) how to present the task to annotators in such a way that the cognitive load is within manageable limits and annotations can be replicated with sufficient consistency.

Not all concepts can clearly and easily be associated with specific words in the text. While mentions of a drug or effect entity always have corresponding words in the text (entity annotations fall under (1a) above), this is not necessarily the case of concepts naturally seen as relations between other concepts. Consider again the last post in Figure 1: there are no substrings that can be associated with Lamictol *causing* cognitive deficits. Rather, it is the first two sentences in their entirety that imply (but do not state explicitly) the causal link. It tends to be considered not appropriate to mark up such relations in text, and they attach instead to one or more already marked up word strings (meaning they fall under (1b) above).

Regarding (2) above, aside from the issues raised in Section 3.3, available resources are a deciding factor: e.g. it would take annotators far longer to determine the specific drug concept label for a drug mention than it would to simply label each such mention with a generic drug label.

Regarding (3), while KE would typically aim to extract information with factual status from text (e.g. all drug effects patients claim to have experienced), far from all mentions of such information have factual status (drug effect mentions an be negated, part of a question, etc.). Simply treating e.g. a negated drug effect as not a drug effect is unlikely to be helpful in a machine learning context, because negated and non-negated versions will look identical except for a negation marker elsewhere in the text, potentially resulting in a large number of spurious negative examples. It is more likely to help generalisation to treat all drug effect mentions identically, and to additionally mark up in annotations, and subsequently learn, the characteristics of negation. The same holds for generalisations, questions and similar phenomena.

Regarding (4), it is virtually impossible to achieve perfect consistency between annotators, or even self-consistency, with mark-up annotation

```
DRUG_NONADHERENCE [
    +DRUG_MODIFICATION [ instigated_by = _patient_
            intentional || _patient_unintentional ]   ]

DRUG_MODIFICATION [
    drug : Drugs
    change : DOSE_CHANGE || FORM_CHANGE || PROD
            UCT_CHANGE
    +reason : DRUG_EFFECT || DRUG_PROPERTY ||
            Other_Reason
    instigated_by : Instigator
    +mod_effect : EFFECT || Other_Outcome   ]

DOSE_CHANGE [
    ?from_dose : ( Numbers, Units, Time_Span )
    ?to_dose : ( Numbers, Units, Time_Span )
    dose_mod_type : DoseModTypes   ]

FORM_CHANGE [
    ?from_form : Form_Types
    ?to_form : Form_Types   ]

PRODUCT_CHANGE [
    ?from_product : Drug_Products
    ?to_product : Drug_Products   ]

DRUG_EFFECT [
    drug : Drugs
    medDRA_preferred_term : MPTs
    duration : ( Numbers, Units )
    severity : Severity_Levels   ]

DRUG_PROPERTY [
    drug : Drugs
    property : Drug_Properties   ]

Drugs = {x|x is a drug entry in SIDER}
Drug_Products = {x|x is a drug product entry in SIDER}
MPTs = {x|x is an MPT entry in SIDER}
Drug_Properties = {_effective, _ineffective, _like, _dis-
            like, _cheap, _expensive, ...}
Numbers = {x|x ∈ ℝ}
Units = {_ml, _mg, ...}
Time_Span = {_halfday, _day, _week, ...}
Form_Types = {_tablet, _powder, _solution, _cream, ...}
DoseModTypes = {_start, _stop, _increase, _decrease}
Other_Reason = {_life_event, _insurance_financial,
            _view_belief, ...}
Instigator = {_patient_intentional, _patient_unintentional,
            _clinician, _both_together, _other}
Other_Outcome = {_negative, _positive, _neutral}
Severity_Levels = {_mild, _moderate, _severe}
```
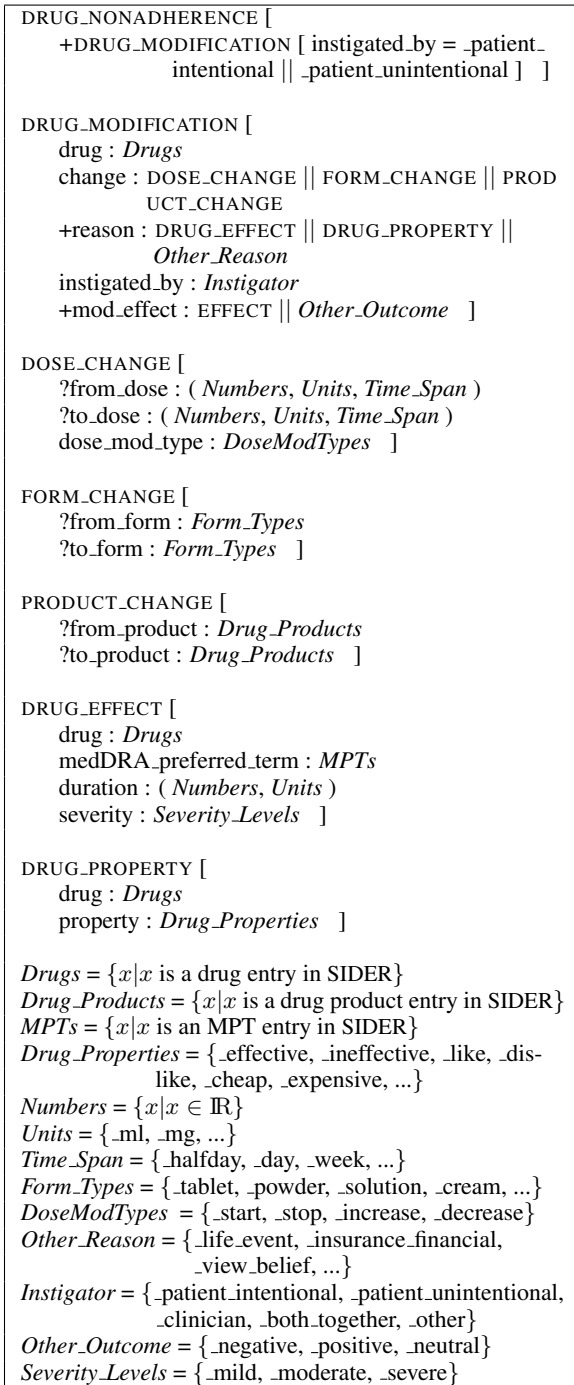
Figure 3: Nonadherence concept model (UML-like notation). + = 1 or more; || = xor; ? = at least 1 of two.

schemes. For simpler concepts corresponding to fewer possible word strings, such as named entities of type drug, issues are comparatively simple, and inter-annotator agreement (IAA, the extent to which annotators agree where drug mentions begin/end) would be high. As is apparent even in the simple example word strings in Example 1, there is comparatively higher variation in word strings describing side effects: if the last two examples

were extracted from the longer strings *I'm being so forgetful* and *I just keep losing my phone...*, respectively, in our experience there tends to be considerable variation among annotators about where to place the start of the effect mention.

One way to address this is to ensure that the concepts underlying annotation labels are highly coherent and crisply defined, so that there is high concurrence among annotators in how to interpret and apply them. This underlines the need to co-develop conceptualisation and annotation scheme, because they provide the formal grounding that can help ensure coherence and crispness. For high IAA, the tasks annotators are asked to perform need to be focused and homogeneous, and the visual interfaces as uncluttered as possible. If this is not the case it can make the task too difficult, and also quickly lead to frustration among annotators.

4   An Approach to Nonadherence KE

In this section, we scale up the insights from Section 3 to an approach to knowledge extraction (KE) in the nonadherence domain, a substantially more complex task than drug effect detection. We adopt the definition of nonadherence as *not taking a drug as instructed* (Section 2), and assume that the relationship between a patient and a drug starts with a prescription and instructions issued by a clinician. We see nonadherence as one or more modifications to the original prescription regimen, or to a previous modification, made without the approval of the prescribing clinician. Our first task then is to model the concept of modification, after which we can define nonadherence as modifications instigated by the patient. Our goal is to design a concept model, KE template and annotation scheme for nonadherence that support KE methods that extract information about the *how* and *why* of drug nonadherence.

4.1   Concept Model

Information about *how* and *why* nonadherent drug modifications occur will necessarily involve a specific **drug**. To address the *how* part, we need to know what **type of modification** was carried out. Nonadherent modifications in all examples we have encountered involve some change to the **dose** that is taken, if stopping, starting and forgetting to take a drug are considered dose modifications. Other types of modifications also apply to drug **form** (tablet, capsule, etc.), or drug **brand**

(e.g. generic vs. branded). To address the *why* part, we also need to know the **reason** for the modification and its **effect**, because it often becomes the reason for a further modification. Our nonadherence concept model is shown in Figure 3. Starting from the top, it defines drug nonadherence as one or more drug modifications instigated by the patient either intentially or unintentionally. A drug modification is composed of a drug, a change, optionally a reason and an effect, plus a specification of instigation. A reason is either a drug effect, drug property or another reason. A drug effect is as defined in Section 3; a drug property identifies the drug it relates to and covers non-effect properties such as cost, how well it works and whether the patient likes it. Other reasons include reasons relating to life events, insurance issues and beliefs held by the patient. Others are possible, and as indicated in the model, we have not defined a final set of terminal concepts for some of the preterminals (preterminal concepts are indicated by italics). In order to define the possible terminal concepts for the preterminal drugs, drug products and effects concepts, we use the SIDER knowledge base (Kuhn et al., 2015).

4.2 KE Template and Task

The concept model in Figure 3 is task-agnostic. For the specific task of nonadherence event extraction from health forum posts it needs to be mapped to a KE template, task specification and annotation scheme that are necessarily task-specific. Following inspection of about 200 random texts from our corpus of 150K posts (Section 5), and based on the concept model above, we construe the nonadherence KE task as follows:

1. Binary classification of posts into first person narration vs. others.

2. Binary classification of posts into containing modification mentions vs. others.

3. Anaphora resolution: replace drug, drug form and drug dose anaphora with full references.

4. Entity detection: drug, dose, effect, modification.

5. Entity linking: drug, effect, modification (yielding modification type); text normalisation: dose.

6. Topic segmentation: drug related, other.

7. NE relation detection via binary classification (applied to drug-related topic fields only):

   - drug, dose, context → dose_of?
   - drug, effect, context → drug_effect?
   - dose, effect, context → dose_effect?
   - modification, drug, context → drug_modification?
   - modification, dose, context → drug_modification?
   - drug_modification, effect, context → mod_reason?
   - drug_modification, context → clinician, patient_intentional, patient_unintentional, both_together, not_stated?
   - drug_modification, context → stop, start, increase, decrease, unclear?

Note that we do not currently include product or form changes in the task construal. Moreover, we are leaving identification of drug properties, and modification reasons other than effects to future work. While the first six tasks above would need to be implemented, in this order, in a pipeline, there is no intention to imply that the subtasks under (7) would be implemented separately and in a specific order. Rather, there is likely to be benefit from jointly modelling some or even all of them.

The above KE process is aimed at filling KE templates derived from the concept model in Figure 3 (and the annotation scheme in the next section). Initially, we are using the following template for each drug_modification identified by the KE process, here instantiated for the second and third sentences in the first post in Figure 1:

$$
\begin{aligned}
&\text{DRUG\_MODIFICATION [} \\
&\quad \text{DRUG= \_sertraline;} \\
&\quad \text{DOSE= \_unknown;} \\
&\quad \text{MOD\_REASON= \_unknown;} \\
&\quad \text{MOD\_TYPE= \_stop;} \\
&\quad \text{INSTIGATED\_BY= \_patient\_intentional;} \quad]
\end{aligned} \tag{4}
$$

Ideally we would also like to extract information about the severity and duration of drug effects, and the order in which they occur, but have had to exclude those for the time being as infeasibly hard from health forum posts.

4.3 Nonadherence Annotation Scheme

The annotation scheme we have devised to match the concept model, KE task and KE template above, consists of 8 entities and 3 events, as shown in Figure 4. The entities are annotated as labels attached to identified word strings in the text, with the meaning of (1a) in Section 3.4. Events are not associated with word spans, but link two or more entities; events also have sets of attributes.

In order to minimise cognitive load for our annotators, we made several implementational choices that are not reflected in Figure 4, partly influenced by the brat evaluation tool we are using for annotating texts.[4] E.g. we annotated DRUG_EFFECT and DRUG_PROPERTY events in one round, and DRUG_MODIFICATION events in another,

---

[4]brat.nlplab.org

```
Entities:
Drug, Drugs, Drug_Dose, Drug_Form, Drug_Brand, Drug_Effect, Drug_Property, Other_Reason.
Events:
DRUG_EFFECT:
    Argument 1: (Drug|Drugs|Drug_Form|Drug_Dose|Drug_Brand)+
    Argument 2: Drug_Effect.
    Attributes (binary, optional): Question, Negation, Generalisation, Speculation, ReducedCertainty.
DRUG_PROPERTY:
    Argument 1: (Drug|Drugs|Drug_Form|Drug_Dose|Drug_Brand)+
    Argument 2: Drug_Property
    Attributes (binary, optional): Question, Negation, Generalisation, Speculation, ReducedCertainty.
DRUG_MODIFICATION:
    Argument 1: Drug|Drug_Form|Drug_Dose|Drug_Brand
    Argument 2 (reason(s) for modification): (DRUG_EFFECT|DRUG_PROPERTY|Other_Reason)*
    Attributes (binary, optional): Question, Negation, Generalisation, Speculation, ReducedCertainty.
    Attributes (multiple valued, obligatory):
        Instigated_By = {not_stated, clinician, patient_intentional, patient_unintentional, both_together}.
        Mod_Type = {increase, decrease, start, stop, unclear}.
Relations:
antecedent: links pronouns and common noun references to their most recent antecedent (named reference).
dose_of, form_of, brand_of: link dose, form and brand mentions to the drugs they relate to.
```

Figure 4: Nonadherence annotation scheme.

separate round; we annotated antecedent links as chains of antecedent relations to the nearest full reference, using all intervening anaphoric references, in order to minimise clutter in the interface.

4.4 Agreement among annotators

The scheme was developed in several iterations of development /testing, each time improving concept model, scheme and Inter-Annotator Agreement (InterAA). For annotations with the final version of the scheme we allowed four hours per 50 posts (average post length is 83 words). InterAA for entities, as measured by averaged brateval scores (F1 on combined label/span matches) computed on 50 random posts, between our two main annotators, ranges from 0.74 for Drug_Effect, and 0.64 for Drug, to 0.39 for Drug_Property. The corresponding IntraAA scores are 0.85, 0.8 and 0.75 for one annotator, and 0.75, 0.73, and 0.81 for the other (numbers for Drug_Property indicate annotators are interpreting the guidance differently).

5 Data Collection and Analysis

We opted for arms-length data scraping and de-identification where a trusted third party scraped health forum posts, and de-identified the texts, making available to us the masked version of the dataset only. The partner accessed and downloaded all posts on the 11 drug-specific forums on www.depressionforums.org at the end of 20 Dec 2018, yielding 148,575 posts. The posts were processed and converted to text-only form, forum post IDs were removed and replaced with new dataset-specific post IDs, and personally identifiable information (PII) was masked, e.g. usernames were replaced by the token [USER] and person names by [NAME]. In addition, the partner performed adverse event (AE) (Nielsen, 2011) and sentiment scoring (Xu and Painter, 2016) for each post. AE scores express the probability p that the post contains mention of an adverse event (AE), thresholded at p=0.7. Sentiment scores range from -5 to 5, with negative sentiment thresholded at -1, and positive sentiment at 1, with scores from -1 to 1 indicating neutral sentiment.

Post were distributed over the 11 forums as shown in Table 1 in terms of number of posts and percentage of total (columns 2 and 3); also shown are median post length, percentage of posts with AEs, and percentages of posts with positive/negative/neutral sentiment. Some trends can be observed in Table 1. Post length tends to go up as forum size increases. Some forums contain substantially higher rates of AE mentions than others: prevalence ranges from 47.6% of posts for Abilify, to 63.9% for Citalopram. These correlate to some extent with sentiment scores: e.g. 22.5% of Abilify posts were classified as negative in sentiment, compared to 28.1% for Citalopram. There is no correlation (Pearsons r=0.17) between % AE mention and % positive sentiment; there is a strong inverse correlation (r=-0.74) between % AE and % neutral sentiment, and some correlation (r=0.3) between % AE and % negative sentiment.

We have so far annotated 2,000 posts in Phase 1. In the annotated posts, there are 3,882 individual

Forum	Posts	%-age	Median tokens	% AE mention	% pos sentim.	% neg sentim.	% neut sentim.
55-citalopram-celexa-lexapro-escitalopram	31,286	21.1%	87	63.9%	50.2%	21.7%	28.1%
46-other-depression-and-anxiety-medications	21,408	14.4%	80	56.3%	47.7%	24.7%	27.6%
56-wellbutrin-bupropion	20,098	13.5%	93.5	62.6%	51.0%	21.8%	27.2%
53-zoloftlustral-sertraline	19,597	13.2%	87	64.4%	50.2%	21.2%	28.5%
54-effexor-venlafaxine-pristiq-desvenlafaxine	15,014	10.1%	82	63.8%	46.5%	23.1%	30.4%
50-cymbalta-duloxetine	12,439	8.4%	85	62.1%	49.0%	22.0%	28.9%
52-prozac-fluoxetine	11,166	7.5%	80	54.3%	47.7%	23.2%	29.0%
51-remeron-mirtazapine	8,142	5.5%	80	61.4%	51.7%	22.6%	25.7%
57-paxilseroxat-paroxetine	4,476	3.0%	68	52.2%	45.9%	23.4%	30.7%
103-abilify-aripiprazole	3,223	2.2%	56	47.6%	52.2%	25.3%	22.5%
102-viibryd-vilazodone	1,726	1.2%	59	55.7%	42.9%	26.4%	30.7%
Total	*148,575*	*100.0%*	*83*	*60.9%*	*49.2%*	*22.6%*	*28.2%*

Table 1: Data set statistics, adverse drug event (AE) and sentiment scores.

Drug	Drugs	Drug_Dose	Drug_Form	Drug_Effect	Drug_Property	Que.	Neg.	Gen.	Spec.	Red.
2,320	216	615	51	2,845	1,037	137	360	313	169	308

Table 2: Occurrence counts for Phase 1 annotations (Que=Question, Neg=Negation, Gen=Generalisation, Spe= Speculation, Red=ReducedCertainty; see Section 4.3).

Drug_Effect and Drug_Property annotations altogether. 999 posts have at least one such annotation; 258 posts have exactly 1, 197 have 2, 151 have 3, 108 have 4, and 285 have 5 or more (up to 28). Table 2 presents occurrence counts for entities and events from Phase 1.

6 Related Research

Structured information resources in health informatics range from ordered lists of terms, glossaries and medical thesauri (MeSH[5], UMLS (Bodenreider, 2004)), to ontologies like SNOMED CT (Donnelly, 2006) and BioPax (Demir et al., 2010). Such resources have underlying concept models, from the very simple (e.g. in a drug list each entry is a member of the class drug) to the much more complex, e.g. ontologies incorporating complex relations, properties and structures.

KE work in health informatics involves implicit or explicit underlying concept models. Examples include adverse drug effect detection (Karimi et al., 2015; Yates et al., 2015), usually involving two main stages—entity identification and entity linking—although some simply classify posts as containing a drug-effect mention or not (Bollegala et al., 2018). Others have applied further layers of interpretation such as sentiment extraction, e.g. headache is negative (Cameron et al., 2013).

Conceptualisations have been developed for more complex health domains. Mowery et al. classify posts as containing evidence of depression to

yield a first layer of information which is then instantiated by either a specific symptom or psychosocial stressor (Mowery et al., 2015). Other studies have addressed suicide (Desmet and Hoste, 2014; Huang et al., 2017), flu avoidance (Collier et al., 2011), cyber-bullying (Van Hee et al., 2015), and rumours (Zubiaga et al., 2016).

7 Conclusion

In this paper our aim has been to pin down and clarify the interdependent and mutually constraining elements involved in developing an approach to knowledge extraction, encompassing the underlying concept model, KE task construal and corresponding KE template, as well as the annotation scheme. We have discussed the issues that arise when addressing each of the elements, the choices that need to be made and the trade-offs involved.

All this reflects our experience of developing an annotation scheme for drug nonadherence. While we have discussed the steps involved in developing a KE approach in the context of the nonadherence domain, we found that many of the steps and issues are not domain-specific, and are also applicable to KE in other domains.

Acknowledgments

---

[5] http://www.nlm.nih.gov/mesh/meshhome.html

References

Olivier Bodenreider. 2004. The unified medical language system (umls): integrating biomedical terminology. *Nucleic acids research*, 32(suppl_1):D267–D270.

Danushka Bollegala, Simon Maskell, Richard Sloane, Joanna Hajne, and Munir Pirmohamed. 2018. Causality patterns for detecting adverse drug reactions from social media: Text mining approach. *JMIR public health and surveillance*, 4(2):e51.

Delroy Cameron, Gary A Smith, Raminta Daniulaityte, Amit P Sheth, Drashti Dave, Lu Chen, Gaurish Anand, Robert Carlson, Kera Z Watkins, and Russel Falck. 2013. Predose: a semantic web platform for drug abuse epidemiology using social media. *Journal of biomedical informatics*, 46(6):985–997.

Nigel Collier, Nguyen Truong Son, and Ngoc Mai Nguyen. 2011. Omg u got flu? analysis of shared health messages for bio-surveillance. *Journal of biomedical semantics*, 2(5):S9.

Emek Demir, Michael P Cary, Suzanne Paley, Ken Fukuda, Christian Lemer, Imre Vastrik, Guanming Wu, Peter D'eustachio, Carl Schaefer, Joanne Luciano, et al. 2010. The biopax community standard for pathway data sharing. *Nature biotechnology*, 28(9):935.

Bart Desmet and Véronique Hoste. 2014. Recognising suicidal messages in dutch social media. In *9th international conference on language resources and evaluation (LREC)*, pages 830–835.

Kevin Donnelly. 2006. Snomed-ct: The advanced terminology and coding system for ehealth. *Studies in health technology and informatics*, 121:279.

Sandra van Dulmen, Emmy Sluijs, Liset Van Dijk, Denise de Ridder, Rob Heerdink, and Jozien Bensing. 2007. Patient adherence to medical treatment: A review of reviews. *BMC health services research*, 7(1):55.

Erica CG van Geffen, Helga Gardarsdottir, Rolf van Hulten, Liset van Dijk, Antoine CG Egberts, and Eibert R Heerdink. 2009. Initiation of antidepressant therapy: Do patients follow the gp's prescription? *Br J Gen Pract*, 59(559):81–87.

Ben Hachey, Will Radford, Joel Nothman, Matthew Honnibal, and James R Curran. 2013. Evaluating entity linking with wikipedia. *Artificial intelligence*, 194:130–150.

Xianpei Han, Le Sun, and Jun Zhao. 2011. Collective entity linking in web text: a graph-based method. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 765–774. ACM.

Xiaolei Huang, Linzi Xing, Jed R Brubaker, and Michael J Paul. 2017. Exploring timelines of confirmed suicide incidents through social media. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 470–477. IEEE.

Jacqueline G Hugtenburg, Lonneke Timmers, Petra JM Elders, Marcia Vervloet, and Liset van Dijk. 2013. Definitions, variants, and causes of nonadherence with medication: A challenge for tailored interventions. *Patient preference and adherence*, 7:675.

Sarvnaz Karimi, Alejandro Metke-Jimenez, Madonna Kemp, and Chen Wang. 2015. Cadec: A corpus of adverse drug event annotations. *Journal of Biomedical Informatics*, 55:73 – 81.

Michael Kuhn, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. 2015. The sider database of drugs and side effects. *Nucleic acids research*, 44(D1):D1075–D1079.

Robert Leaman, Laura Wojtulewicz, Ryan Sullivan, Annie Skariah, Jian Yang, and Graciela Gonzalez. 2010. Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. In *Proceedings of the 2010 workshop on biomedical natural language processing*, pages 117–125. Association for Computational Linguistics.

Alejandro Metke-Jimenez and Sarvnaz Karimi. 2016. Concept identification and normalisation for adverse drug event discovery in medical forums. In *BMDID@ ISWC*.

Danielle Mowery, Craig Bryan, and Mike Conway. 2015. Towards developing an annotation scheme for depressive disorder symptoms: A preliminary study using twitter data. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 89–98.

Finn Årup Nielsen. 2011. A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Azadeh Nikfarjam, Abeed Sarker, Karen Oconnor, Rachel Ginn, and Graciela Gonzalez. 2015. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association*, 22(3):671–681.

Takeshi Onishi, Davy Weissenbacher, Ari Klein, Karen O'Connor, and Graciela Gonzalez-Hernandez. 2018. Dealing with medication non-adherence expressions in twitter. In *Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop and Shared Task*, pages 32–33, Brussels, Belgium. Association for Computational Linguistics.

Delip Rao, Paul McNamee, and Mark Dredze. 2013. Entity linking: Finding extracted entities in a knowledge base. In *Multi-source, multilingual information extraction and summarization*, pages 93–115. Springer.

Eduardo Sabaté et al. 2003. *Adherence to long-term therapies: evidence for action*. World Health Organization.

Karen Smith, Su Golder, Abeed Sarker, Yoon Loke, Karen OConnor, and Graciela Gonzalez-Hernandez. 2018. Methods to compare adverse events in twitter to faers, drug information databases, and systematic reviews: Proof of concept with adalimumab. *Drug safety*, 41(12):1397–1410.

Phil Tregunno. 2017. Use of social media in pharmacovigilance. Report on WEB-RADR results to 2017 Yellow Card Centre Conference, Wales.

Cynthia Van Hee, Ben Verhoeven, Els Lefever, Guy De Pauw, Véronique Hoste, and Walter Daelemans. 2015. Guidelines for the fine-grained analysis of cyberbullying.

Yingzi Xu and Jeffery L Painter. 2016. Application of classification and clustering methods on mvoc (medical voice of customer) data for scientific engagement.

Andrew Yates, Nazli Goharian, and Ophir Frieder. 2015. Extracting adverse drug reactions from social media. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one*, 11(3):e0150989.