Patient Centered Identification, Attribution and Ranking of Adverse Drug Events

Ritwik Banerjee I. V. Ramakrishnan
Computer Science
Stony Brook University
Stony Brook, NY 11794-4400

Mark Henry Matthew Perciavalle
School of Medicine
Stony Brook University
Stony Brook, NY 11794-4400

***Abstract*—Adverse drug events (ADEs) trigger a high number of hospital emergency room (ER) visits. Information about ADEs is often available in online drug databases in the form of narrative texts, and serves as the physician's primary reference point for ADE attribution and diagnosis. Manually reviewing these narratives, however, is an error prone and time consuming process, especially due to the prevalence of polypharmacy. So ER health care providers, especially given the heavy volume of traffic in ERs, often either skip this step or at best do it rather perfunctorily. This causes ADEs to be missed or misdiagnosed, often leading to extensive and unnecessary testing and treatment, including hospitalization. In this paper, we present a system that automates the detection of ADEs and provides a list of suspect drugs, ranked by their likelihood of causing the patient's complaints and symptoms. The input data, i.e., medications and complaints, are obtained from triage notes that often contain descriptive language. Our application utilizes heterogeneous information sources (including drug databases) to refine and transform these descriptions as well as the online database narratives using a natural language processing (NLP) pipeline. We then employ ranking measures to establish correspondence between the complaints and the medications. Our preliminary evaluation based on actual ER cases demonstrates that this system achieves high precision and recall.**

I. Introduction

Adverse Drug Events (ADEs) are undesired reactions experienced due to the use, misuse or discontinuation of medications. Several studies have shown that a high fraction of emergency room (ER) visits are due to ADEs [1]–[3]. Reviewing the patient's medications for attribution of ADEs is necessary, but over 40% of such cases are overlooked by emergency physicians [4], [5]. Given that most ERs are overcrowded, such lapses in patient care are potentially an outcome of the increased workload [6]–[9]. Overcrowding also means that spending more time on a single patient may not be feasible. Indeed, it has been observed that under such circumstances, physicians increasingly restrict themselves to questions that can be instantly answered [10]. Unfortunately, current methods of ADE attribution requires a physician to manually read through narrative texts in online pharmaceutical databases (*e.g.* Lexicomp [11] or Micromedex [12]), and even while missing such a high percentage of ADEs, physicians already suffer from the "4000 click" syndrome – spending much of their time with electronic records rather than in direct patient care [13].

ER physicians thus need an *evidence-based* clinical decision support (CDS) tool that automatically detects possible ADE diagnoses and instantly *pushes* the suggestions to the physician without necessitating any clicks. Such a tool will not only improve patient safety by detecting ADEs, but also save a significant amount of time in crowded ERs by allowing physicians to quickly determine whether or not a patient's complaints should be attributed to an ADE.

This second aspect is particularly important due to the high number of potential drug interactions in ER patients. In research as well as in clinical practice, drug interactions have mainly been considered as pairwise events [14], and higher-order interactions have rarely been studied [15]. Due to the lack of available resources on higher-order drug interactions, such interactions are beyond the scope of this work. Even with this restriction, a patient taking 12 different drugs calls for the ER clinician to check up to 78 possible drug-related complications (12 drugs and 66 potentially interacting drug-pairs). Even if a clinician spends only a minute per potential complication to look up the pharmaceutical database and skim through the adverse effect narrative, it would be far beyond the feasible amount of time for this task, especially in ERs, which more often than not, tend to heavy traffic.

In this work, we present a CDS system for identification, attribution and ranking of ADEs (IATRO-ADE) that non-intrusively pushes ADE diagnosis suggestions to the physician. Further, in order to avoid *alert fatigue* (*i.e.*, the clinician becomes less responsive to alerts in general), it ranks the suggestions by the likelihood of a drug (or multiple drugs) causing the patient's symptoms. Our application features the simultaneous use of not just multiple information sources, but multiple *types* of sources to accomplish this. We make extensive use of natural language processing (NLP), template-based text mining and disproportionality analysis metrics to perform the following tasks:

- ($\mathcal{T}_1$) distill the potential adverse effects of a drug from unstructured or semi-structured narrative texts,
- ($\mathcal{T}_2$) perform *entity normalization* using structured ontologies and semi-structured narratives to resolve surface differences of symptom mentions, and
- ($\mathcal{T}_3$) measure the association between drugs and adverse events using structured and unstructured data sources.

978-1-4673-9548-9 2015

DOI 10.1109/ICHI.2015.8

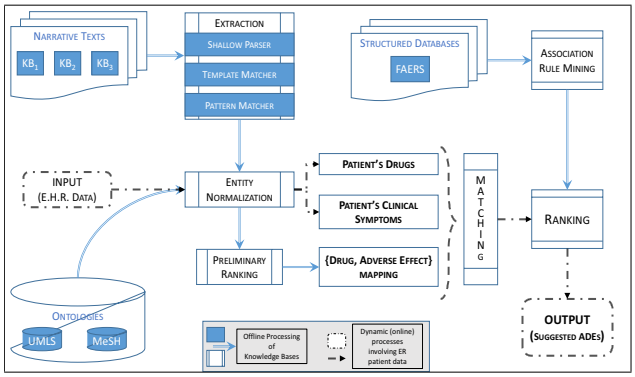

Fig. 1. **Process Flow:** The ADE detection system involves (i) offline information extraction from various knowledge bases (KBs) and (ii) online processes to handle patient data provided as dynamic input. The final output is a ranked list of ADE attribution suggestions provided to the ER physician.

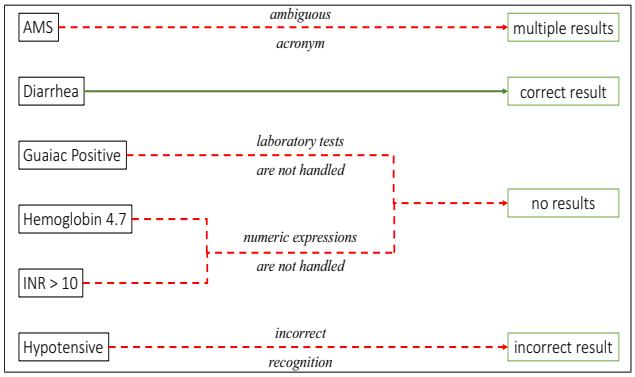

Fig. 2. **Entity Normalization:** Current ontology-based information extraction fails to map varied natural language expressions of patients' symptoms to unambiguous canonical medical entities. As a result, laboratory test results (*e.g. Guaiac positive*), ambiguous abbreviations (*e.g. AMS*), etc. are ignored.

In order to find the possible ADE diagnoses, we map the adverse effects (extracted by $\mathcal{T}_1$) of a patient's drugs to her symptoms (processed by $\mathcal{T}_2$). Subsequently, the ADEs are ranked based on the severity of the adverse effect and co-occurrence statistics as computed by the third task $\mathcal{T}_3$. The overall process flow is presented in Fig. 1.

Entity Normalization

The extraction module processes narrative texts through a combination of shallow parsing and template-based text mining methods. The standard information extraction approach is to use ontologies (*e.g.* MeSH [16], UMLS [17]) and tools to map natural language to concepts in these ontologies (*e.g.* Metamap [18]). This, however, is ill suited for extraction from triage notes because clinicians often use domain-specific expressions that are often not captured by existing methods. For example, a clinical symptom may be provided as a laboratory test result like "Hemoglobin 4.7". Existing methods address the *normalization* of medical expressions, *i.e.* mapping various linguistic expressions to an unambiguous canonical form, in a limited manner. This leads to non-recognition of ADE evidence. Fig. 2 provides an example where key evidence of an adverse reaction cannot be correctly detected without extensive entity normalization. Details of the information extraction and entity normalization processes are provided in Sections IV and V.

Matching and Relevance Ranking

The complete list of all those drugs that are being taken by a patient and whose adverse effects match the patient's complaints and symptoms is often a large fraction of the patient's entire drug regimen. In other words, simply matching the patient's complaints and symptoms to the side effects of her drugs will almost certainly detect an adverse effect (*i.e.* a very high recall), but is likely to suffer from low precision. This approach has been shown to cause *alert fatigue* because the CDS system provides too much information of low clinical significance [19]. To resolve this, our pipeline filters out spurious ADE attributions, and the remaining are ranked. Section VI explains this process in greater detail.

II. Related Work

Pharmacovigilance, the discipline pertaining to the detection, assessment and prevention of drug-related adverse events, has been a topic of significant interest in the healthcare community. Computer-aided approaches aimed at ADE prevention in patients has focused on creating tools and services that raise alerts by prompting clinicians about potential ADEs. A vast majority of these, however, are prevention systems designed to issue an alert at the moment of prescribing new medication [20]–[22]. Moreover, even though such CDS implementations are aimed at preventing adverse events, they have not always improved patient safety [23], [24]. The lack of a pervasive and tangible improvement is due to two seemingly conflicting factors: *coverage* and *alert fatigue*.

Coverage versus Alert Fatigue

In one line of work, alerts are rule-based, where the conditions for raising an alert are built from medical terms in a patient's symptoms and ADE descriptions. These terms are drawn from discharge summaries, ambulatory notes or ADE reports [25]–[29]. Entity extraction systems based on Metamap (*e.g.* MedLEE [30]) have also been used. But because the rules are manually curated, such approaches reportedly suffer from low coverage [25], [26], [31]. On the other hand, rule-based trigger systems that have attempted to generalize in order to improve coverage tend to raise too many alerts, thereby causing alert fatigue [32]–[35]. In some cases, clinicians were found to be ignoring upto 96% of the raised alerts [36]. Our work adopts a two-fold approach:

(a) To ensure maximum coverage, exploit multiple knowledge bases for each task.
(b) To produce relevant *evidence-based* ADE diagnosis suggestions, *normalize* entities and filter out ADE attributions not supported by adequate evidence. Finally, *rank* the suggestions that were retained.

Some prior studies have argued that ranking is a more suitable approach compared to filtering [37], [38], but more recent reports suggest that expert panels recommend filtering as well [39]. We thus adopt an *evidence based* filtering and

TABLE I
COMPONENTS OF THE IATRO-ADE PIPELINE

Offline	Dynamic (Online)
• Extraction from *semi-structured* and *unstructured* data (a) drug adverse effects (b) drug-drug interactions (c) disease characterization in terms of clinical symptoms (d) laboratory test information • Mining *structured* data (a) statistical association measures for ⟨drug, adverse-effect⟩ pairs	• symptom similarity resolution • abbreviation resolution and disambiguation • named entity normalization for patient symptoms • evidence-based ADE detection and attribution • ranking ADE diagnosis suggestions

ranking approach. Further, unlike the ranking algorithms in existing CDS systems, our simultaneous use of several types of information allows for the final suggestions to be ranked based on a combination of multiple factors like the severity of an ADE, the likelihood of its occurrence, etc.

Identification of ADEs using external knowledge sources

Previous work on identification of ADEs using external sources has largely not been patient-centered. Much of it has focused on discovering adverse reactions from retrospective data. Some have used a single source like the FDA Adverse Event Reporting System (FAERS) data [40], [41], while others have worked on information fusion [42]–[44]. Their techniques based on large amounts of retrospective data are well-suited for discovering new ADEs, but not applicable for *real-time* ADE detection and attribution.

A few examples do exist for patient-centered ADE-detection in hospital settings. For example, Duke and Friedlin [45] proposed a real-time decision support service for ADEs, but their system does not handle unstructured data of the type obtained from triage notes. These, however, do not perform any entity normalization beyond ontology-based mapping to UMLS concepts. As shown in Fig. 2, this leads to missing out on crucial evidence of adverse reactions.

III. METHOD OVERVIEW

We use unstructured, semi-structured and structured knowledge bases (KBs) for the complete pipeline, the key steps of which are divided into an offline component of extracting information from various KBs and a dynamic online component (shown with dashed lines in Fig. 1) that handles patient input data. Both are listed in table I. It is worth noting that the results obtained in the offline processes are required at various stages by the online component.

We use multiple *types* of KBs in IATRO-ADE . The first comprises of drug description repositories. The core knowledge of adverse effects of single drugs and drug-drug interactions (DDIs) is extracted from them. The second type of KBs are medical encyclopedias, which act as knowledge sources for symptoms, diseases and laboratory tests and procedures. Information extracted from these KBs forms the gold standard knowledge for laboratory test results and for characterization of diseases and syndromes in terms of clinical symptoms (*e.g. hypotension* is mapped to *dizziness*, *fainting*, etc.).

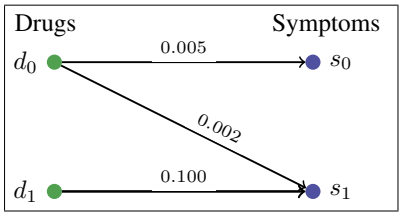

Fig. 3. The ADE attribution problem: a hypothetical case of a patient taking drugs $\{d_0, d_1\}$ exhibiting symptoms $\{s_0, s_1\}$. The directed edges represent causality between drug and symptom, with edge weights reflecting the likelihood of the cause. Even though d_0 *explains* all symptoms, the partial match $\langle d_1, s_1 \rangle$ has a higher likelihood.

Given a set of drugs $D = \{d_1, d_2, \ldots, d_m\}$ and symptoms $S = \{s_1, s_2, \ldots, s_n\}$ obtained from a patient record, IATRO-ADE identifies a subset $D_0 \subseteq D$ such that the adverse effects of drugs in D_0 explain all the symptoms in S. If a symptom $s_j \in S$ can be explained as a possible adverse effect of $d_i \in D$, we denote it as $d_i \xrightarrow{cause} s_j$. For ease of notation, we also extend it to sets of drugs and symptoms so that $D \xrightarrow{cause} S$ denotes that the set of drugs D may cause the symptoms S. It is possible that some symptoms are not known to be adverse effects of any $d \in D$. In such cases, our method attempts to find the subsets of D that can explain maximal subsets of S. Formally, this can be expressed as

$$\mathbb{D} = \{D_i \subseteq D\} = \underset{D' \subseteq D}{\operatorname{argmax}} \left\{ |S'| : S' \subseteq S, D \xrightarrow{cause} S \right\}. \quad (1)$$

Since the $argmax$ is not unique, (the patient's entire drug regimen, D, is trivially in $\mathbb{D}$), $\mathbb{D}$ is a set of subsets of D. We thus select the minimal such sets:

$$D_{min} = \underset{D_i}{\operatorname{argmin}} \{|D_i| : D_i \in \mathbb{D}\}. \quad (2)$$

Note that eq. 2 only attempts to find the minimal subsets with no regard for severity or likelihood of the adverse effects. This naïve approach is clearly not realistic. For instance, consider a patient taking two drugs d_0 and d_1 and complaining of headache and abdominal pain, where both are potential adverse effects of d_0, but d_1 can only cause the latter. If these adverse effects of d_0 are rare while d_1 causing abdominal pain is common, then it is more likely that the headache is caused by some non-iatrogenic factor and $\langle d_1, \text{abdominal pain} \rangle$ is the correct attribution. Fig. 3 illustrates this as a weighted bipartite graph. In light of such scenarios, instead of simply solving eq. 2, we use it to iteratively compute the list of all possible ADE attributions, as shown in Algorithm 1.

Data: Set of drugs D and symptoms S
Result: ADE attributions in descending order of coverage of S
$Q_D \leftarrow$ empty queue of sets of drugs;
$\mathcal{P}(D) \leftarrow \{D' \subseteq D\}$;
$i \leftarrow 0$;
while $P(D) \neq \emptyset$ **do**
  $D_i = \text{argmax}_{D_i \in P(D)} \left\{ |S'| : S' \subseteq S, D' \xrightarrow{cause} S' \right\}$
  $\mathcal{P}(D) \leftarrow \mathcal{P}(D) \setminus D_i$;
  $i \leftarrow i + 1$;
end
return *Minimal sets of* Q_D

Algorithm 1: Exhaustive search for ADE attributions

TABLE II
TEMPLATE-BASED TEXT MINING

TEMPLATE: "(side|adverse) (reactions|effects) .* $s_1, \ldots, s_k$"
• *The following additional* **adverse reactions** *have been identified during postapproval use of simvastatin: pruritus, alopecia, rhabdomyolysis, ...* [RxList.com]
• *If any of the following* **side effects** *occur while taking simvastatin, check with your doctor: dizziness, fainting, fast or irregular heartbeat ...* [Drugs.com]

Templates for mining adverse effects of drugs. Linguistic cues are in bold, while the extracted symptoms are italicized.

In order to automatically attribute a patient's symptoms to an ADE, the terms in the information extracted from the drug description repositories are *normalized*. This is the process of mapping varying linguistic expressions to unambiguous canonical forms by similarity resolution, abbreviation resolution and disambiguation, and finally, named entity normalization. For example, if the offline extraction has obtained "fatigue" as the adverse effect of a drug, while an ER patient taking that drug is complaining of "tiredness", similarity resolution identifies the two terms as *nearly equivalent*.

Symptoms expressed as acronyms or abbreviations are resolved by either looking up structured data tables, or by applying the abbreviation extraction algorithm of Schwarts and Hearst [46] on the third type of KBs: biomedical literature data. For this, we use the PubMed Central (PMC) repository. Ambiguous abbreviations are resolved (see section V), and the canonical forms are matched to see if a particular drug d_i can cause a symptom s_j.

Finally, a fourth type of KB, structured data from FAERS, is used in conjunction with the PMC data to compute statistical association measures. These measures are used to provide the final ranked list to the physician.

IV. INFORMATION EXTRACTION

In the scope of this work, we extract information to map drugs to their adverse effects, diseases to their symptoms and laboratory test results to diseases and symptoms. This section explains how these mappings are achieved.

A. Extracting adverse effects of drugs

We extract information about drugs from publicly available information on websites as well as proprietary services used by the Stony Brook University Hospital. DrugBank[47], consisting of $7,759$ drug entries, forms the basis of possible medications. Further, we extract information about these drugs from Drugs.com [48], RxList.com [49] and Micromedex [12]. These sources provide a mix of structured, semi-structured and unstructured information. The structured information can be directly obtained by table lookups. Semi-structured information can also be harnessed by rule- or template-based text mining. This requires identifying patterns that are indicative of the information being presented. Typically, they include cue words or phrases like "may cause", "side effects", etc. Table II presents an example template with two matching text snippets.

Finally, to extract adverse effect information from narrative texts, we resort to ontology-based entity extraction. This enables the identification of not just the medical entities, but also their *semantic type*. We use Metamap, a tool that maps phrases to UMLS semantic types [17], [18]. Metamap also handles chunking a sentence into smaller phrases, but experiments on development data showed that entity recognition was more accurate if it handled the natural language input one phrase at a time. We use the Genia tagger [50] for shallow parsing, and from its results, pass only the noun phrases to Metamap. This method of extracting the adverse effects of single drugs is also used to glean DDI information. Additionally, some KBs (*e.g.*, Drugs.com and Micromedex) also provide a 'severity' value for DDIs, which is used as a feature in our ranking algorithm (*cf.* Section VI).

B. Characterizing diseases in terms of their symptoms

As we saw in table II, an adverse effect of a drug may be given in terms of symptoms (*e.g.* "dizziness") or diseases/disorders (*e.g.* "rhabdomyolysis"). Similarly, the patient data may also have symptoms as well as diseases. Thus, a crucial piece of the IATRO-ADE pipeline is to understand the manifestation of a disease in terms of its clinical symptoms. Continuing with the example in table II, let us consider an ER patient on simvastatin and complaining of "muscle pain", "joint pain" and "tiredness". All three are symptoms of "rhabdomyolysis", an adverse effect of the drug. But since the ADE information extracted so far does not map "rhabdomyolysis" to the observable symptoms, this will not be detected. The technique of extracting the symptom-based characterization of a disease or disorder are identical to the template-based text mining and shallow parsing methods used in sec. IV-A. The KBs, however, are general-purpose medical encyclopedias instead of drug databases. We use two such repositories for this step: MedicineNet [51] and MedlinePlus [52].

C. Mapping laboratory test results to diseases and symptoms

Even though we usually think of symptoms as observable physical or mental states like seizures, headaches, etc., a wide range of symptoms are expressed in terms of medical tests [53]. The test result is often an indication or confirmation of a symptom or disease. For the purpose of this work, we use the term 'test' to mean laboratory tests as well as comparatively simple readings such as pulse rate, blood pressure, etc. As shown in Fig. 2, hospital records often use such evaluations to express the presence of a symptom or disease. For example, the symptom "hemoglobin 4.7" indicates a very low count. In order to infer this, however, IATRO-ADE needs to know the reference range for hemoglobin. This section explains how we extract information about tests and draw meaningful inferences from their results.

We use two KBs to map test results to symptoms: (i) the list of procedures and tests available on MedicineNet, (ii) the Laboratory Test Database at University of California, San Fransisco [54] and (iii) the health encyclopedia available from University of Rochester's Health Encyclopedia [55].

TABLE III
UMLS-BASED NAMED ENTITY NORMALIZATION

Entity Type	Recall	Error examples	Cause
Medications	0.98	*Amio* *Simva* *VPA*	non-standard abbrev.
Symptoms	0.69	*qtc* $>$ *460* *occult stool*	no knowledge of laboratory tests

Compared to the extraction of adverse effects, test results are harder to interpret. This is because reference ranges vary depending on the patient's age, gender and medication history. They may also vary from one laboratory to another. Our work errs on the side of caution, and if even one of the KBs claims a particular value to be abnormal, we consider it to be so. In other words, if different KBs disagree on a reference range, we take the most conservative estimate.

The extraction process is similar to the techniques used so far. For each test, we extract the structured data as provided, and perform shallow parsing using Genia tagger and Metamap on the natural language descriptions. This leads to the discovery of two kinds of knowledge:

(a) the reference range or value (*i.e.*, the normal value of the measurement), and
(b) what abnormal values may indicate

For example, from "hemoglobin 4.7" in a hospital record, IATRO-ADE is able to infer that (a) patient's hemoglobin count is lower than the normal value, and (b) the possible reasons are (to name a few) *anemia*, *bleeding*, *colon cancer*, *stomach ulcer*, *iron deficiency* or *folate deficiency*. Moreoever, in cases where multiple test results point toward a common reason, the ability to automatically infer these results enable our approach to distill the stronger signals from all the available data, and thus identify the most likely ADE. Fig. 4 illustrates how this step in the IATRO-ADE pipeline helps establish stronger causality between the drugs and the symptoms.

V. NAMED ENTITY NORMALIZATION

So far, we have described how we extract different types of information pertaining to diseases, symptoms, laboratory tests and adverse effects of drugs – using multiple information sources for each. Even though fusing such heterogeneous information ensures significantly higher coverage, it leads to a situation where vastly different linguistic expressions may refer to the same medical concept. In some cases, resolving the surface differences is simply an issue of identifying and linking synonymous (*e.g.*, the generic name "Warfarin" and its brand name "Coumadin") or nearly-synonymous (*e.g.*, the symptoms "breathlessness" and "shortness of breath") entities. Frequently, less obvious equivalences need to resolved, however. These arise from (a) the pervasive use of domain-specific abbreviations, and (b) expressions involving numeric (*e.g.* "hemoglobin 4.7", "INR 7.1") or nominal (*e.g.* +/-, high/low) values.

TABLE IV
LIMITATIONS OF ONTOLOGY-BASED ENTITY NORMALIZATION

Abbreviation	Expansion	UMLS concept*
NCRS	Nutrition-related chronic diseases	–
CB1	Cannabinoid-1	CNR1 gene
SGLT2	sodium glucose cotransport-2	SLC5A2 gene
vit. def.	Vitamin Deficiency	VIT gene; butyl phosphorotrithioate
GIB	gastrointestinal bleeding	(Gibraltar) [Geographic Area]

* The mapping to UMLS concepts was performed by Metamap.

A. Linking synonyms and near-synonyms

This is by far the simplest step in our pipeline due to the significant amount of prior research devoted to building ontologies of biomedical knowledge (*e.g.* SNOMED-CT, NDF-RT, RxNorm, MeSH) as well as general linguistic knowledge. Since the *Unified Medical Language System* (UMLS) integrates over a 100 such medical ontologies, we exclusively work with UMLS for identifying synonymous medical terms. For identifying equivalent or nearly-equivalent terms that are less domain-specific, we also engage WordNet [56].

The UMLS maps each medical term to a *unique concept identifier*, a unique *preferred* name and one or more *semantic types*, which are biomedical categories. The UMLS holds over 12 million concept names collated into 3.1 million unique concepts and categorized into 135 semantic types.[1] Exploiting this massive knowledge repository allows us to identify most variations in medication and disease names. To a large extent, it is also able to link synonymous symptoms like "shortness of breath" and "dyspnea". Table III presents the recall of UMLS-based entity identification on our data, showing nearly perfect identification of all drugs except when non-standard abbreviations, *e.g.* *simva* and *amio* for *simvastatin* and *amiodarone*, respectively). The ability to correctly identify patients' symptoms and complaints, however, suffers more. This is because a large fraction of symptoms are expressed in terms of laboratory test results, which the existing medical ontologies do not map to symptoms.

B. Abbreviation Resolution

A fairly comprehensive list of abbreviations is already present in UMLS, but as seen in table III, a better entity normalization process requires intelligent abbreviation resolution that goes beyond just looking up data tables. The first step we perform is to check whether an abbreviation matches a synonymous term instead of just the list of medications. This is done to normalize names such as 'VPA' and 'divalproex'.

1) Using unstructured data to discover non-standard abbreviations: Non-standard abbreviations, however, cannot be resolved in this manner. We thus implement the algorithm proposed by Schwartz and Hearst [46] to identify abbreviation definitions from unstructured medical texts. This algorithm

[1] http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/statistics.html

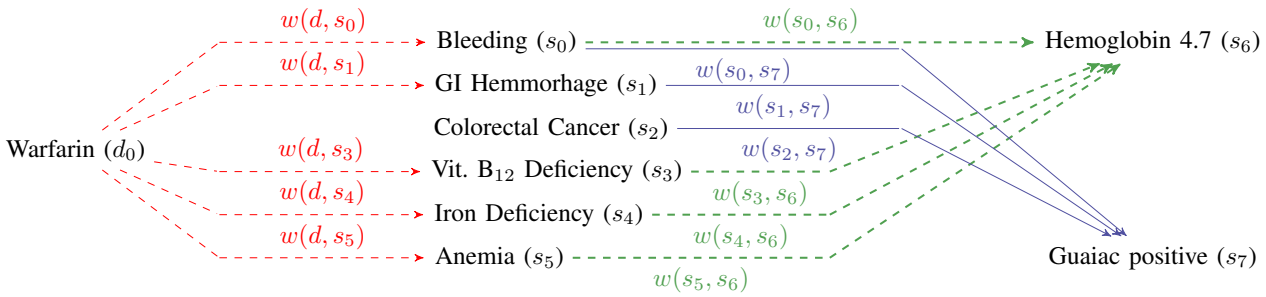

Fig. 4. A graphical model of ADE attribution for a single drug, Warfarin (d_0), based on two laboratory test results. Edge direction denotes causality, and the likelihood of a cause-effect relation between nodes m and n is encoded as edge weight functions $w(m, n)$. Note that multiple test results are capable of reinforcing ADE signals. In this example, most symptoms indicated by the hemoglobin count (s_6) and guaiac positive value (s_7) point to an adverse effect of Warfarin. In case of uniform likelihoods, the most probable adverse effect is "bleeding" (s_0).

is run on the PubMed Central (PMC) dataset, comprising of more than 2.7 million research articles. Table IV shows a few abbreviations we found by this method that were either absent in UMLS or were incorrectly labeled by Metamap.

It is often the case that a single abbreviation has multiple possible expansions. The observation made in such cases was that even though all expansions were sensible in their own right, there was, in each case, only one expansion that was most relevant to the patient's medications or symptoms. This clearly indicates that identifying the correct expansion, and subsequently, being able to perform entity normalization, it was imperative that the context of the abbreviation be taken into account. To learn the correct expansion in a context-dependent manner, we adopt a distributional semantic modeling approach built on the PMC dataset and the semi-structured datasets described in Section IV. The methodology of this step is described next.

Distributional Semantic Modeling Distributional semantics rely on the *distributional hypothesis*, which states that *the meaning of a word is the set of contexts in which it occurs across a large number of texts*. To illustrate this clearly, we present two examples in table V, one medical and one non-medical, to show how the meaning of a previously unknkown word may be inferred from its context. This approach has been used in non-medical recommender systems before [57], [58] and very recently, has also been applied to the entity normalization problem [59], [60].

Given an abbreviation, our goal is to identify the correct expansion out of multiple candidate expansions. In order to achieve this, we construct a context vector v_e for entity e. A *context vector* of a term is a vector designed to capture the set of contexts in which it occurs. For IATRO-ADE , we build these vectors by considering a fixed-size window around every mention of e. For documents in the PMC dataset, the paragraph in which e occurs is considered as the context window, and for documents from the semi-structured sources, the text contained under the relevant section heading (*e.g.* "adverse effects") is considered. All context vectors are normalized to unit length, and the angular distance based on cosine similarity is used to determine the distance between two vectors, *i.e.*

$$d(v_e, v_f) = 1 - cos^{-1}\left(\frac{v_e.v_f}{||v_e||.||v_f||}\right)\frac{1}{\pi} \tag{3}$$

where $||x||$ denotes the Euclidean norm of a vector. The candidate expansion whose context vector is closest to the context vector of the abbreviation is selected.

It should be noted that creating the context vectors is a resource-intensive process, and cannot be achieved in real-time scenarios. Context vectors of abbreviations in the KBs are thus constructed offline, and all the candidate expansions are retained. The final selection is done based on the context provided by patient data, *i.e.*the list of medications and symptoms. For a given abbreviation a_0, this final step can be expressed formally as

$$e_0 = \underset{e \in D \cup S}{\operatorname{argmin}}\, d(v_e, v_{a_0}) \tag{4}$$

where D and S are the sets of drugs and symptoms obtained from the patient record.

C. Normalizing expressions involving laboratory test results

As we saw before, a patient's symptoms may be expressed in terms of laboratory test results. This means that phrases like "hemoglobin 4.7", "guaiac positive" (Fig. 2) or "occult stool" (Table III) must be identified with the canonical name of the corresponding condition. Just like in abbreviation resolution, this mapping, too, generates multiple candidates.

Unlike abbreviation resolution, however, we do not select one candidate as the canonical expression. Instead, all the possible causes extracted from semi-structured KBs (using methods described in Sec. IV) are retained, and a probabilistic identification is induced. For example, if an abnormal test result indicates n distinct potential conditions (*e.g.* low hemoglobin may indicate gastrointestinal bleeding, iron deficiency, etc.), a probability is associated to each potential cause. The probability distribution is computed by extracting detailed syntactic and semantic information from the PMC dataset, and subsequently calculating co-occurrence statistics.

For each symptom expressed in terms of a laboratory test result, the lexical content is separated from the expression.

TABLE V
THE DISTRIBUTIONAL HYPOTHESIS

(I)	After 10mg of ____, *hyperglycemic patients* felt immediate *relief*. Metformin is known to provide *relief* to *patients* suffering from *hyperglycemia*.
(II)	The *tiny brown* ____ *scurried away under the branches*. The *little brown* rabbit *disappeared under the branches*.

The meaning of the words in the blank places may be inferred based on the context: (I) suggests a treatment that lower blood glucose levels, and (II) implies that the subject is most likely a small animal.

For example, from "hemoglobin 4.7", we retain "hemoglobin". Further, ubiquitous terms like "test", etc. are also removed in order to retain only the terms specific to the symptom. For example, if a symptom is expressed as "guaiac test was positive", we will only retain the term "guaiac". This filtering is done by tokenizing all the text from the laboratory test KBs, and ranking all terms by their *inverse document frequency* (IDF). For D being the set of all documents under consideration, the IDF of a term t is defined as $IDF(t) = \log(|D|/|\{d \in D : t \in d\}|)$. This measures whether t is common or rare across all documents. Common terms, clearly, are not significant for a specific test. This method allows for filtering out generic words like "test", "was", etc.

Let t_0 denote the term that remains after the above filtering is carried out. We then search the PMC dataset to find (i) the number of documents in which t_0 appears, and (ii) the number of documents in which a potential cause (*e.g.* iron deficiency) co-occurs with t_0. For our purpose, the first step calls for a significant amount of NLP techniques.

Computing the probability of causes In case the original symptom was a numeric expression, or contained an adjective modifier, we mark the semantic orientation of this expression. This is done for nominal expressions (*e.g.* +/-) by triggering binary values. For numeric expressions, we check whether the value is lower or higher than the reference range, and trigger a binary value indicating low/high. For example, for the expression "hemoglobin 4.7", we compare the value 4.7 with the reference range for hemoglobin test, and mark the fact that 4.7 is lower than normal. Let this semantic orientation be denoted by $S(t_0)$. When querying for t_0 in the KB, we filter out documents unless the same semantic orientation $S(t_0)$ is observed in the document.

This is done by splitting the text into sentences, and then generating the dependency parse tree of those in which t_0 appears. We then check if a word with a negative connotation, like "low", is a modifier of t_0. This check is performed after lemmatization so that inflected forms like "lower", "lowered", etc. are accounted for. Fig. 5 presents an example of such a sentence. In case there is a numeric value as a dependency, then the semantic orientation is checked by comparing it to the reference range for the test. Let the semantic orientation obtained from the dependency parse tree be denoted by $T(t_0)$. The probability of a cause c is then given by

$$P(c|t_0) = \frac{|\{d \in D : t_0 \in d,\, T(t_0) = S(t_0),\, c \in d\}|}{|\{d \in D : t_0 \in d,\, T(t_0) = S(t_0)\}|} \quad (5)$$

These probability values are used as shown in Fig. 4. They are also used to rank the patient's medications by their relevance with respect to the reported symptoms.

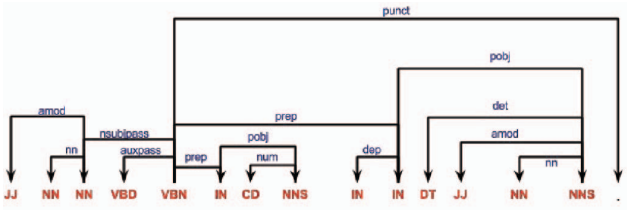

Fig. 5. Using dependency parse trees to identify modifiers. The connotation of a modifier ('low') in this figure indicates the semantic orientation of the object being modified ('hemoglobin levels').

VI. RELEVANCE RANKING

The ADE attributions are computed using the exhaustive search Algorithm 1. The extensive steps taken to extract relevant information from several KBS and subsequently normalize entity names is necessary for the matching algorithm to perform well, and exhibit high recall (*i.e.*it does not miss a potential ADE attribution). However, as prior work in this domain has shown, presenting the physician with all such potential cases leads to alert fatigue. IATRO-ADE thus employs a ranking algorithm as well. Next, we discuss this algorithm and show that the most plausible attributions are always ranked among the top diagnostic suggestions.

The first step is to rank the possible ADEs by severity. The severity ratings are extracted from semi-structured KBs as discussed earlier in Section IV-A. The next step is to obtain co-occurrence statistics for drugs and their adverse effects. To this end, we make use of the structured portions of the semi-structured KBs, where relevant statistics are often presented in tabular format. For example, for a majority of drugs, the percentage of reported cases where patients suffered from a particular adverse effect are reported in this manner. To bolster the association between drugs and their adverse effects, IATRO-ADE also performs simple disproportionality analysis (DPA) on the FAERS dataset.

The first DPA metric we use is a conditional probability value, also known as the *relative reporting ratio* (RRR). It computes the ratio between the probability of a symptom s given a drug d and the probability of s in the entire dataset:

$$RRR(s,d) = \frac{P(s|d)}{P(s)} = \frac{P(s,d)}{P(s).P(d)}. \quad (6)$$

The second metric is a relative risk metric, called the *proportional reporting ratio* (PRR), which measures the ratio of the frequency of s in patients exposed to d to the frequency of s in patients not exposed to d:

$$PRR(s,d) = \frac{P(s|d)}{P(s|\neg d)}. \quad (7)$$

Both metrics were computed only for those $\langle s, d\rangle$ pairs that occurred at least 5 times, as co-occurrences that are too rare are not deemed statistically meaningful. These frequency-based association scores are also computed on the PMC

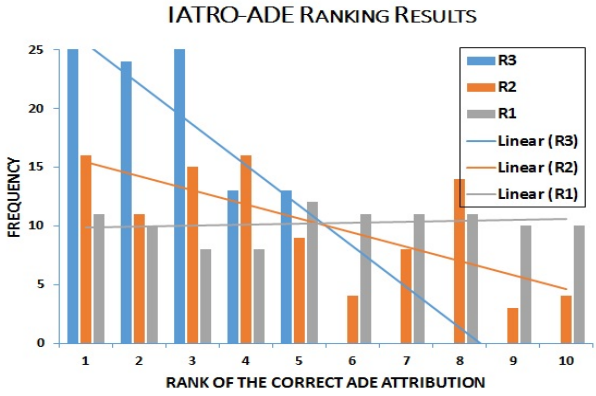

Fig. 6. Distribution of the ranks of the correct ADE attributions computed by IATRO-ADE . Shown here are R_1 (only severity rating, *i.e.*, $\beta = \gamma = 0$ in eq. 8), R_2 (severity plus DPA on structured and semi-structured KBs, *i.e.*, $\gamma = 0$), and R_3 (severity plus DPA on all types of KBs, *i.e.*, $\alpha, \beta, \gamma > 0$). The linear trends show that R_3 is successful in providing the correct ADE attributions at the top of the ranked list. In our data, IATRO-ADE is always able to provide the correct ADE attribution among the top 5 suggestions.

dataset. However, our initial experiments showed that if, while computing the joint probability score $P(s, d)$, we insist that s and d are syntactically connected, then the RRR and PRR scores tend to be negligible. Further, we noticed that a much stronger signal is obtained if the condition is relaxed to include every document where s and d co-occur. This relaxed notion of co-occurrence was used to compute the RRR and PRR metrics based on the PMC dataset. In addition to computing these two metrics to capture drug-symptom associations, we also used them to compute association metrics for DDIs. This was done with a simple extension of the formulae in eq. 6 and 7. Thus, the IATRO-ADE ranking scheme stands on three measures:

(m_1) severity
(m_2) RRR and PRR based on FAERS and tabular data from semi-structured KBs
(m_3) RRR and PRR based on PMC data

The likelihood $\mathcal{L}(d, s)$ of a drug d causing a symptom s is a weighted linear sum of these three factors whose coefficients are experimentally determined:

$$\mathcal{L}(d, s) = \alpha m_1 + \beta m_2 + \gamma m_3. \quad (8)$$

VII. Experimental Results

In this section, we present the experimental results of IATRO-ADE . The evaluation was done by human experts on a dataset of 100 ER patient records. As a baseline, we performed maximal matching (MM), where the output consists of all possible drugs that can cause the given symptoms. No entity normalization or ranking is done for the baseline, providing a naïve and imprecise system that returns every potential ADE that can be directly matched to the patient's drug(s), no matter how unlikely.

Note that even though the baseline yields every potential adverse effect, it still suffers from low recall because a large fraction of drugs and conditions cannot be identified using just

TABLE VI

Experiment	Precision	Recall	MRR		
Baseline	0.23	0.52	–		
MM_{EN}	0.39	0.94	–		
IA	0.65	0.52	–		
IA_{EN}	0.75	0.94	–		
$\text{IA}_{EN,Rank}$	0.75	0.94	R_1 0.30	R_2 0.37	R_3 0.59

medical ontologies. The missed out cases include laboratory test results, non-standard abbreviations and phrasal expressions. To see how these expressions can also be identified, we continued with maximal matching, but added entity normalization. This method, denoted MM_{EN}, achieves nearly perfect recall. But since it matches everything without performing any filtration, it suffers from very low precision. We then employ the iterative ADE attribution algorithm (Algorithm 1), denoted IA, to obtain the smallest set of drugs that can explain maximal sets of the patient's symptoms. Finally, we add entity normalization to the IA method, denoted by IA_{EN}.

The ranking algorithm is evaluated separately since precision and recall are set-based measures, computed over unordered collections. The precision and recall values over the entire list of ranked results will be identical to the unranked set obtained by IA_{EN}. To evaluate the ranking algorithm, we use *mean reciprocal rank* (MRR), a widely used metric in the information retrieval domain [61], [62]. It is the average of the reciprocal ranks of all the output lists. If Q is the set of data being evaluated (in our case, $|Q| = 100$), and r_i is the rank of the correct suggestion for the i^{th} patient, then

$$MRR = \frac{1}{|Q|} \sum_{i}^{|Q|} \frac{1}{r_i}. \quad (9)$$

The results of our experiments are shown in Table VI using (R_1) only the severity-based ranking, (R_2) severity plus DPA metrics from semi-structured and structured KBs, and (R_3) severity plus DPA metrics from all types of KBs.

VIII. Conclusion

We presented a pipeline to automatically identify adverse drug events, attribute them to patients' symptoms and rank the possible ADE diagnoses with respect to their relevance to the patient's medications and symptoms. This is done by extracting information from multiple *types* of knowledge sources. As a part of this pipeline, we utilize techniques of natural language processing and information extraction to present ways to reduce alert fatigue while retaining high precision and recall. Most notably, it involves sophisticated entity normalization and relevance ranking processes. Our evaluation on real ER patient data has shown that the approach taken by IATRO-ADE holds promise, and is capable of alerting clinicians with highly accurate ADE attribution notifications.

Acknowledgmment: This research was supported in part by NSF Award IIS 1447549.

REFERENCES

[1] G. Trifiro, G. Calogero, F. M. Ippolito, M. Cosentino, R. Giuliani, A. Conforti, M. Venegoni, G. Mazzaglia, and A. P. Caputi, "Adverse drug events in emergency department population: a prospective Italian study," *Pharmacoepidemiol Drug Saf*, vol. 14, no. 5, pp. 333–340, May 2005.

[2] P. J. Zed, R. B. Abu-Laban, R. M. Balen, P. S. Loewen, C. M. Hohl, J. R. Brubacher, K. Wilbur, M. O. Wiens, L. J. Samoy, K. Lacaria *et al.*, "Incidence, severity and preventability of medication-related visits to the emergency department: a prospective study," *Canadian Medical Association Journal*, vol. 178, no. 12, pp. 1563–1569, 2008.

[3] N. Jayarama, K. Shiju, and K. Prabahakar, "Adverse drug reactions in adults leading to emergency department visits," *Int J Pharm Pharm Sci*, vol. 4, pp. 642–646, 2012.

[4] C. M. Hohl, P. J. Zed, J. R. Brubacher, R. B. Abu-Laban, P. S. Loewen, and R. A. Purssell, "Do emergency physicians attribute drug-related emergency department visits to medication-related problems?" *Ann Emerg Med*, vol. 55, no. 6, pp. 493–502, Jun 2010.

[5] L. Roulet, F. Ballereau, J. B. Hardouin, A. Chiffoleau, G. Potel, and N. Asseray, "Adverse drug event nonrecognition in emergency departments: an exploratory study on factors related to patients and drugs," *J Emerg Med*, vol. 46, no. 6, pp. 857–864, Jun 2014.

[6] S. Trzeciak and E. P. Rivers, "Emergency department overcrowding in the United States: an emerging threat to patient safety and public health," *Emerg Med J*, vol. 20, no. 5, pp. 402–405, Sep 2003.

[7] J. S. Olshaker and N. K. Rathlev, "Emergency Department overcrowding and ambulance diversion: the impact and potential solutions of extended boarding of admitted patients in the Emergency Department," *J Emerg Med*, vol. 30, no. 3, pp. 351–356, Apr 2006.

[8] J. Collis, "Adverse effects of overcrowding on patient experience and care," *Emerg Nurse*, vol. 18, no. 8, pp. 34–39, Dec 2010.

[9] J. S. Weissman, J. M. Rothschild, E. Bendavid, P. Sprivulis, E. F. Cook, R. S. Evans, Y. Kaganova, M. Bender, J. David-Kasdan, P. Haug, J. Lloyd, L. G. Selbovitz, H. J. Murff, and D. W. Bates, "Hospital workload and adverse events," *Med Care*, vol. 45, no. 5, pp. 448–455, May 2007.

[10] K. Ramos, R. Linscheid, and S. Schafer, "Real-time information-seeking behavior of residency physicians," *Fam Med*, vol. 35, no. 4, pp. 257–260, Apr 2003.

[11] "Lexicomp." [Online]. Available: http://www.lexi.com/

[12] "Micromedex." [Online]. Available: http://micromedex.com

[13] R. G. Hill, L. M. Sears, and S. W. Melanson, "4000 clicks: a productivity analysis of electronic medical records in a community hospital ED," *Am J Emerg Med*, vol. 31, no. 11, pp. 1591–1594, Nov 2013.

[14] J. R. Horn, P. D. Hansten, and L.-N. Chan, "Proposal for a new tool to evaluate drug interaction cases," *Annals of Pharmacotherapy*, vol. 41, no. 4, pp. 674–680, 2007.

[15] B. Mannheimer, *Drug-related problems with special emphasis on drug-drug interactions*. Institutionen för klinisk forskning och utbildning, Södersjukhuset/Department of Clinical Science and Education, Södersjukhuset, 2009, pp. 7–8.

[16] F. B. Rogers, "Medical subject headings," *Bull Med Libr Assoc*, vol. 51, pp. 114–116, Jan 1963.

[17] D. Lindberg, B. Humphreys, and A. McCray, "The unified medical language system," *Methods of Information in Medicine*, vol. 32, no. 4, pp. 281–291, 1993.

[18] A. R. Aronson, "Effective mapping of biomedical text to the umls metathesaurus: the metamap program." in *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2001, p. 17.

[19] P. E. Beeler, D. W. Bates, and B. L. Hug, "Clinical decision support systems," *Swiss Med Wkly*, vol. 144, p. w14073, 2014.

[20] W. L. Galanter, R. J. Didomenico, and A. Polikaitis, "A trial of automated decision support alerts for contraindicated medications using computerized physician order entry." *J Am Med Inform Assoc*, vol. 12, no. 3, pp. 269–274, 2005.

[21] A. Schedlbauer, V. Prasad, C. Mulvaney, S. Phansalkar, W. Stanton, D. W. Bates, and A. J. Avery, "What evidence supports the use of computerized alerts and prompts to improve clinicians' prescribing behavior?" *J Am Med Inform Assoc*, vol. 16, no. 4, pp. 531–538, 2009.

[22] A. K. Jha, C. M. DesRoches, E. G. Campbell, K. Donelan, S. R. Rao, T. G. Ferris, A. Shields, S. Rosenbaum, and D. Blumenthal, "Use of electronic health records in U.S. hospitals," *N. Engl. J. Med.*, vol. 360, no. 16, pp. 1628–1638, Apr 2009.

[23] J. H. Gurwitz, T. S. Field, P. Rochon, J. Judge, L. R. Harrold, C. M. Bell, M. Lee, K. White, J. LaPrino, J. Erramuspe-Mainard, M. DeFlorio, L. Gavendo, J. L. Baril, G. Reed, and D. W. Bates, "Effect of computerized provider order entry with clinical decision support on adverse drug events in the long-term care setting," *J Am Geriatr Soc*, vol. 56, no. 12, pp. 2225–2233, Dec 2008.

[24] B. L. Strom, R. Schinnar, F. Aberra, W. Bilker, S. Hennessy, C. E. Leonard, and E. Pifer, "Unintended effects of a computerized physician order entry nearly hard-stop alert to prevent a drug interaction: a randomized controlled trial," *Arch. Intern. Med.*, vol. 170, no. 17, pp. 1578–1583, Sep 2010.

[25] G. B. Melton and G. Hripcsak, "Automated detection of adverse events using natural language processing of discharge summaries," *J Am Med Inform Assoc*, vol. 12, no. 4, pp. 448–457, 2005.

[26] X. Wang, G. Hripcsak, M. Markatou, and C. Friedman, "Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study," *J Am Med Inform Assoc*, vol. 16, no. 3, pp. 328–337, 2009.

[27] T. Botsis, M. D. Nguyen, E. J. Woo, M. Markatou, and R. Ball, "Text mining for the vaccine adverse event reporting system: medical text classification using informative feature selection," *J. Am. Med. Inform. Assoc.*, vol. 18, pp. 631–638, Jun 2011.

[28] N. P. Tatonetti, G. H. Fernald, and R. B. Altman, "A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports," *J. Am. Med. Inform. Assoc.*, vol. 19, no. 1, pp. 79–85, June 2011.

[29] M. N. Cantor, H. J. Feldman, and M. M. Triola, "Using trigger phrases to detect adverse drug reactions in ambulatory care notes," *Qual. Saf. Health Care*, vol. 16, no. 2, pp. 132–134, Apr 2007.

[30] C. Friedman, "A broad-coverage natural language processing system." in *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2000, p. 270.

[31] H. J. Murff, V. L. Patel, G. Hripcsak, and D. W. Bates, "Detecting adverse events for patient safety research: a review of current methodologies," *J Biomed Inform*, vol. 36, no. 1-2, pp. 131–143, 2003.

[32] J. Rozich, C. Haraden, and R. Resar, "Adverse drug event trigger tool: a practical methodology for measuring medication related harm," *Quality and Safety in Health Care*, vol. 12, no. 3, pp. 194–200, 2003.

[33] D. C. Classen, R. Resar, F. Griffin, F. Federico, T. Frankel, N. Kimmel, J. C. Whittington, A. Frankel, A. Seger, and B. C. James, "global trigger toolshows that adverse events in hospitals may be ten times greater than previously measured," *Health Affairs*, vol. 30, no. 4, pp. 581–589, 2011.

[34] R. Koppel, C. E. Leonard, A. R. Localio, A. Cohen, R. Auten, and B. L. Strom, "Identifying and quantifying medication errors: evaluation of rapidly discontinued medication orders submitted to a computerized physician order entry system," *J Am Med Inform Assoc*, vol. 15, no. 4, pp. 461–465, 2008.

[35] J. G. Baseman, D. Revere, I. Painter, M. Toyoji, H. Thiede, and J. Duchin, "Public health communications and alert fatigue," *BMC Health Serv Res*, vol. 13, p. 295, 2013.

[36] H. van der Sijs, J. Aarts, A. Vulto, and M. Berg, "Overriding of drug safety alerts in computerized physician order entry," *J Am Med Inform Assoc*, vol. 13, no. 2, pp. 138–147, 2006.

[37] H. van der Sijs, J. Aarts, T. van Gelder, M. Berg, and A. Vulto, "Turning off frequently overridden drug alerts: limited opportunities for doing it safely," *J Am Med Inform Assoc*, vol. 15, no. 4, pp. 439–448, 2008.

[38] E. K. Lee, A. F. Mejia, T. Senior, and J. Jose, "Improving Patient Safety through Medical Alert Management: An Automated Decision Tool to Reduce Alert Fatigue," *AMIA Annu Symp Proc*, vol. 2010, pp. 417–421, 2010.

[39] S. Phansalkar, H. van der Sijs, A. D. Tucker, A. A. Desai, D. S. Bell, J. M. Teich, B. Middleton, and D. W. Bates, "Drug-drug interactions that should be non-interruptive in order to reduce alert fatigue in electronic health records," *J Am Med Inform Assoc*, vol. 20, no. 3, pp. 489–493, May 2013.

[40] N. P. Tatonetti, P. P. Ye, R. Daneshjou, and R. B. Altman, "Data-driven prediction of drug effects and interactions," *Sci Transl Med*, vol. 4, no. 125, p. 125ra31, Mar 2012.

[41] B. Polepalli Ramesh, S. M. Belknap, Z. Li, N. Frid, D. P. West, and H. Yu, "Automatically Recognizing Medication and Adverse Event Information From Food and Drug Administration's Adverse Event Reporting System Narratives," *JMIR Med Inform*, vol. 2, no. 1, p. e10, 2014.

[42] A. Sarker and G. Gonzalez, "Portable automatic text classification for adverse drug reaction detection via multi-corpus training," *J Biomed Inform*, vol. 53, pp. 196–207, Feb 2015.
[43] G. Jiang, H. R. Solbrig, and C. G. Chute, "ADEpedia: a scalable and standardized knowledge base of Adverse Drug Events using semantic web technology," *AMIA Annu Symp Proc*, vol. 2011, pp. 607–616, 2011.
[44] S. Yeleswarapu, A. Rao, T. Joseph, V. G. Saipradeep, and R. Srinivasan, "A pipeline to extract drug-adverse event pairs from multiple data sources," *BMC Med Inform Decis Mak*, vol. 14, p. 13, 2014.
[45] J. D. Duke and J. Friedlin, "ADESSA: A Real-Time Decision Support Service for Delivery of Semantically Coded Adverse Drug Event Data," *AMIA Annu Symp Proc*, vol. 2010, pp. 177–181, 2010.
[46] A. S. Schwartz and M. A. Hearst, "A simple algorithm for identifying abbreviation definitions in biomedical text," in *Proceedings of the 8th Pacific Symposium on Biocomputing*, 2003, pp. 451–462.
[47] "Drugbank." [Online]. Available: http://www.drugbank.ca
[48] "Drugs.com." [Online]. Available: http://www.drugs.com/drug_information.html
[49] "Rxlist.com." [Online]. Available: http://www.rxlist.com
[50] Y. Tsuruoka and J. Tsujii, "Bidirectional inference with the easiest-first strategy for tagging sequence data," in *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 2005, pp. 467–474.
[51] "Medicinenet." [Online]. Available: http://www.medicinenet.com
[52] Bethesda, "MedlinePlus, National Library of Medicine (US)." [Online]. Available: http://www.nlm.nih.gov/medlineplus/
[53] R. Banerjee, Y. Choi, G. Piyush, A. Naik, and I. Ramakrishnan, "Automated suggestion of tests for identifying likelihood of adverse drug events," in *IEEE International Conference on Healthcare Informatics*, 2014, pp. 170–176.
[54] "Ucsf departments of pathology and laboratory medicine — sfgh lab manual — laboratory test database." [Online]. Available: http://labmed.ucsf.edu/sfghlab
[55] "Health encyclopedia – university of rochester medical center." [Online]. Available: http://www.urmc.rochester.edu/encyclopedia/
[56] G. A. Miller, "Wordnet: a lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.
[57] J. L. Herlocker and J. A. Konstan, "Content-independent task-focused recommendation," *Internet Computing, IEEE*, vol. 5, no. 6, pp. 40–47, 2001.
[58] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," in *Recommender systems handbook*. Springer, 2011, pp. 217–253.
[59] C. Musto, G. Semeraro, P. Lops, and M. de Gemmis, "Combining distributional semantics and entity linking for context-aware content-based recommendation," in *User Modeling, Adaptation, and Personalization*. Springer, 2014, pp. 381–392.
[60] M. Lipczak, A. Koushkestani, and E. Milios, "Tulip: lightweight entity recognition and disambiguation using wikipedia-based topic centroids," in *Proceedings of the first international workshop on Entity recognition & disambiguation*. ACM, 2014, pp. 31–36.
[61] E. M. Voorhees *et al.*, "The trec-8 question answering track report." in *TREC*, vol. 99, 1999, pp. 77–82.
[62] R. Baeza-Yates, B. Ribeiro-Neto *et al.*, *Modern information retrieval*. ACM press New York, 1999, vol. 463.