Extraction of Adverse Drug Effects from Medical Case Reports
Harsha Gurulingappa\textsuperscript{1}, Abdul-Mateen Rajput\textsuperscript{2}, and Luca Toldo\textsuperscript{2*}
\textsuperscript{1}Molecular Connections Pvt. Ltd., Basavanagudi, Bangalore 560004, India
\textsuperscript{2}Merck KGaA, Frankfurterstraße 250, Darmstadt 64293, Germany

ABSTRACT
A sheer amount of information about adverse effects of drugs are published in medical case reports that pose major challenges for drug safety experts to perform timely monitoring. Efficient strategies for identification and extraction of information about adverse drug effects from free-text resources are needed to support pharmacovigilance research and decision making. Therefore, this work focuses on the adaptation of a machine learning-based relation extraction system for the identification and extraction of drug-related adverse effects from MEDLINE case reports. It relies on a high quality corpus that was manually annotated, using ontology-driven methodology. Qualitative evaluation of the system show robust results.

1 INTRODUCTION
Adverse effects of drugs is a bothersome issue that confronts drug manufacturers, healthcare providers, and regulatory authorities. Stringent measures for capturing the risks associated with drug usage are established in forms of spontaneous reporting systems that are timely analyzed to ensure safe use of drugs (Hauben and Bate, 2009). Amongst various data sources used by drug safety experts to perform the safety monitoring, case reports published in the scientific biomedical literature represent an important resource due to their abundant existence, rapid rate of generation, and valuable information enclosed (Vandenbroucke, 2001). Due to their unstructured nature, however, manual analysis of the scientific literature is challenging, cumbersome, and labor intensive.

In recent years, development of automatic natural language processing (NLP) and information extraction (IE) techniques have gained immense popularity. They include identification of biomedical named entities, relations between the entities, or events associated with them. Noticeable efforts have been invested on mining the adverse effects in different forms of free-text data. Examples include Wang et al., 2009 who applied the MedLEE system on discharge summaries to identify medication events and entities that could be potential adverse entities that were detected using the strength of statistical association based on their co-occurrences. Leaman et al., 2010 proposed a lenient NLP model for extracting adverse effects of drugs from social media such as blogs. Gurulingappa et al., 2011 developed a machine learning-based system for classifying the sentences in MEDLINE case reports that assert adverse effects of drugs. However, according to the author’s knowledge, there is a limited focus on identification of semantic relationships between drugs and adverse effects in text. This is partly due to the unavailability of suitable public corpora that could be used for technology development and benchmarking. Extracting relations between drugs and adverse effects can facilitate appropriate indexing, precise searching, visualization, and faster information tracing. The use of ontology of adverse drug reactions for automated signal generation in pharmacovigilance has already been proposed (Henegar et al., 2006) and its application to information retrieval has been exploited by the same group few years later, in the VIGITERMES project (Dela mare et al., 2010), where the OntoEIM adverse event ontology have been used to extend the dictionary of adverse event entities, normalize queries, and consolidate annotations, delivering 29% precision and 67% recall of MEDLINE abstracts. Automatic extraction of adverse drug effects from clinical records is an active area of research (Aramaki et al., 2010). Mining social internet message boards to identify adverse drug reactions has been reported (Benton et al., 2011), whereby in that work the extraction of event - drug pairs was determined only using co-occurrence of terms within a window of 20 tokens apart, and the use of machine learning systems was only focused on deidentification for privacy protection. This work reports on the adaptation of a machine learning-based system for identifying the relations between drugs and adverse effects in MEDLINE case reports, that relies on an ontology-driven manually annotated corpus, that strictly follows semantic annotation guidelines developed for clinical text (Roberts et al., 2009). The system has been qualitatively evaluated and studied for its ability of support real time pharmacovigilance studies.

2 METHODS
2.1 Corpus Preparation
The data set used for training and validation of the relation extraction system is the ADE corpus (Gurulingappa et al., 2012). The ADE corpus contains 2972 MEDLINE case reports that are manually annotated and harmonized by three annotators. The corpus contains annotations of 5063 drugs, 5776 conditions (e.g. diseases, signs, symptoms), and 6821 relations between drugs and conditions representing clear adverse effect implications. All annotations are confined to sentence level i.e. drugs and conditions representing adverse effects co-occurring only within individual sentences are annotated. Drugs and conditions that do not fall into adverse effect relations are not annotated. This was done in accordance to the annotation guidelines.

The ADE corpus contains annotations of relations between drugs and conditions that represent \textit{True} relations. This represents a sparsely annotated dataset. For training a supervised classifier, it was essential to generate \textit{False} relations i.e. drugs and conditions that do not fall into adverse effect relations. For this purpose, ProMiner, a dictionary-based named entity recognition system (Hanisch et al., 2005) was employed. ProMiner was incorporated with DrugBank (Knox et al., 2011) and MedDRA (Merrill, 2008) dictionaries for the identification of drugs and conditions respectively in the ADE corpus that were previously not annotated by human annotators. As a result of named entity recognition, new instances encompassing 2269

\textsuperscript{*}Corresponding author: luca.toldo@merckgroup.com; Gurulingappa and Toldo have contributed equally

\textsuperscript{1}Molecular Connections Pvt. Ltd., Basavanagudi, Bangalore 560004, India
\textsuperscript{2}Merck KGaA, Frankfurterstraße 250, Darmstadt 64293, Germany

\textsuperscript{2*}Corresponding author: luca.toldo@merckgroup.com; Gurulingappa and Toldo have contributed equally
drugs and 3437 conditions were automatically annotated. Drug-condition pairs co-occurring within sentences that were previously not annotated by humans formed False relations. Altogether, 5968 False relations were automatically generated. The corpus enriched with machine annotated drugs, conditions, and relations between them is referred as ADE-EXT (indicating extended ADE corpus).

Figure 1 shows an illustration of True and False relations between drug and conditions co-occurring within a sentence.

In the ADE-EXT corpus, 120 manually annotated True relations were not suitable for the NLP task. Examples include overlapping inter-related entities such as acute lithium toxicity where lithium is related to acute toxicity. After removal of nested annotations, the ADE-EXT corpus was decomposed into a training set (ADE-EXT-TRAIN) and a test set (ADE-EXT-TEST). Counts of entities and relations in subsets of ADE-EXT corpora is shown in Table 1.

### Table 1. Counts of entities and relations in ADE-EXT corpus subsets.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>ADE-EXT-TRAIN</th>
<th>ADE-EXT-TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>1884</td>
<td>210</td>
</tr>
<tr>
<td>Drugs</td>
<td>6770</td>
<td>758</td>
</tr>
<tr>
<td>Conditions</td>
<td>8539</td>
<td>978</td>
</tr>
<tr>
<td>Sentences</td>
<td>5333</td>
<td>606</td>
</tr>
<tr>
<td>True Relations</td>
<td>6030</td>
<td>671</td>
</tr>
<tr>
<td>False Relations</td>
<td>4799</td>
<td>546</td>
</tr>
</tbody>
</table>

Fig. 1. Example of a sentence annotated with drug, conditions, and relations between them. True indicates presence of adverse effect relation and False indicates absence of adverse effect relation.

In the ADE-EXT corpus, 120 manually annotated True relations were not suitable for the NLP task. Examples include overlapping inter-related entities such as acute lithium toxicity where lithium is related to acute toxicity. After removal of nested annotations, the ADE-EXT corpus was decomposed into a training set (ADE-EXT-TRAIN) and a test set (ADE-EXT-TEST). Counts of entities and relations in subsets of ADE-EXT corpora is shown in Table 1.

### 2.2 Relation Extraction Workflow

For the identification and extraction of drug-condition entity pairs that fit into adverse effect relation, the Java Simple Relation Extraction (JSRE) system (Gurulingappa et al., 2007) was employed. JSRE provides a re-trainable and scalable supervised classification platform that uses Support Vector Machines (SVMs) (Burges, 1998) with different kernels specially designed for the NLP and relation extraction. All sentences in ADE-EXT-TRAIN and ADE-EXT-TEST containing drug-condition pairs labeled as either True or False were transformed into the SRE format before subjecting them to relation extraction. The SRE format is a unique way of representing data within the JSRE platform where tokens appearing in sentences are enriched with their parts-of-speech tags, lemmas, and flags indicating if a token is a part of named entity or not. Amongst different kernels available, the shallow linguistic kernel was thoroughly used since it has been widely applied and has shown success during similar relation extraction tasks (Tikk et al., 2010). The ADE-EXT-TRAIN was used as data for training and cross-evaluation of JSRE whereas the ADE-EXT-TEST was used an independent test set.

#### 2.3 Mapping annotation ontology against Ontology of Adverse Events

The CLEF initiative (Roberts et al., 2007) investigated how to generate semantically annotated medical corpora for information extraction. As described (Gurulingappa et al., 2012) we adopted the standard established by the CLEF framework for the annotation workflow (Roberts et al., 2009) however we reshaped the annotation schema by using only two of the original entities (CONDITION, DRUG) and extended it with a third one (DOSAGE). None of the relationships used by the CLEF annotation schema could be reused for our work, since the CLEF annotation schema did not consider adverse drug reactions, instead we created two relations: DRUG-CAUSE-CONDITION, DRUG-HAS-DOSAGE. In this work we focused only on automating the detection of DRUG-CAUSE-CONDITION thus DOSAGE will not be mentioned further. The ADE corpus has been created using the Knowtator plugin for Protégé (Ogren, 2006), an ontology-driven corpus annotation tool also used for the creation of the CLEF corpus. Although we adopted the same tool used in CLEF and also adopted the standard established by the CLEF framework for the annotation workflow, we could not adopt the same annotation ontology since the latter was not able to capture the adverse drug relation and the drug dosing relation. The annotation ontology described above was therefore used to create the ADE corpus. Subsequent to the corpus creation, the realism-based biomedical ontology for representation of adverse events (AEO) has been published (Yonggun et al., 2011). AEO has been developed following the principles of Ontological Realism, thus is aligned with the Basic Formal Ontology and the Relation Ontology, and with the Open Biological and Biomedical Ontologies (OBO) Foundry principles of openness, collaboration and use of a common shared syntax. AEO has 484 representational units, annotated by means of 369 terms with specific identifiers and 115 terms imported from existing ontologies. The use of ontologies has proven of great value in bio-medicine, also since it enable machine reasoning, abstraction and automatic hypothesis generation. We therefore had interest in investigating if the knowledge encoded in the annotations of the ADE
corpus could be semantically connected to the AEO. For doing this, we manually compared the definitions of the entities of AEO and of ADE annotation ontology. Figure 2 shows the basic design patterns of AEO, ADE and CLEF as from the original papers, emphasizing shared entities using green and red colours.

3 RESULTS

3.1 Performance Evaluation Criteria

The performance of relation extraction was evaluated by 10-fold cross-validation of the training data. During cross-validation of the training data and final evaluation over the test set, classification performances were assessed using the F-score over True-labeled relations since they denote adverse effect relations between drugs and conditions that denote a focussed relation class being studied.

3.2 Assessment of Relation Extraction

Baseline experiments began with training and cross-validation of JSRE over the ADE-EXT-TRAIN corpus. Results of system’s performances are shown in Table 2. The system achieved an overall F-score of 0.87 after cross-validation. Upon the final test over ADE-EXT-TEST, the system attained F-score of 0.87 indicating a consistency in classification. A subset of instances misclassified during the cross-validation and testing were manually investigated to understand the common sources of errors. Limited context appeared to be one reason for misclassification. For example, the title Niacin maculopathy (PMID:3174043) infers maculopathy as an adverse effect of niacin that lacks contextual description to support machine classification. Distantly co-occurring inter-related entities constituted couple of errors. For example, in the sentence CASE SUMMARY: A 65-year-old patient chronically treated with the selective serotonin reuptake inhibitor (SSRI) citalopram developed confusion, agitation, tachycardia, tremors, myoclonic jerks and unsteady gait, consistent with serotonin syndrome, following initiation of fentanyl, and all symptoms and signs resolved following discontinuation of fentanyl (PMID:17381671); the relation between confusion and the last appearing drug name fentanyl was not correctly classified.

3.3 Impact of Size of the Training Set on the Performance

In order to study the impact of size of the training data on performance of classification, the ADE-EXT-TRAIN was decomposed into random subsets containing 10, 20, 50, 100, 200, 500, 1000, and 2000 documents. The JSRE was trained on these subsets independently in different rounds and subsequently applied on the ADE-EXT-TEST for performance evaluation. Table 3 shows that already using 500 documents one could achieve performances in the 80% range. Whereby, to reach a classifier with a standard deviation of 1%, one needs a substantially large training data.

3.4 Mapping the ADE Annotation Ontology to the Ontology of Adverse Events

As clearly shown in Figure 2, both the ADE annotation ontology and the Ontology of Adverse Events represent adverse drug reactions using formal ontological methods. Inspite of this common goal, the two ontologies use different naming for the two core entities: a CONDITION in the ADE annotation ontology coincide with a DRUG ADVERSE EVENT in AEO; a DRUG in the ADE annotation ontology coincide with a DRUG-ADMINISTRATION in AEO. The ADE ontology additionally introduce the entity DOSAGE, not specified in AEO at the time of its development since AEO originally focused on vaccines for which dosing is not an essential medical concept. Both ADE and AEO model a causal relationship between CONDITION OR ADVERSE EVENT and DRUG OR MEDICAL INTERVENTION, with the latter being the causal source. The only entity shared by the CLEF annotation ontology with AEO and ADE is the DRUG-OR-DEVICE, that coincide with a DRUG OR MEDICAL INTERVENTION.

4 CONCLUSION

This work reports on the adaptation of a machine learning-based JSRE system for the identification and extraction of adverse effects of drugs in case reports. A methodology has been discussed to enrich a sparsely annotated corpus and its subsequent use to build a classification model. Evaluation of the system’s performance showed promising results. Performance of the system can be improved in several ways. In the current experiments, only the default features acceptable by JSRE were used. Optimization of feature representation to include additional features for instance from syntactic sentence parse trees may further improve the results. Development of additional strategies like post-processing to classify relations with missing contextual descriptions can help to recover more relations.

The reported experimental results denote research status on adverse drug effect identification from text. There are several strategies that will be immediately followed. The authors plan to benchmark the performances of several named entity taggers against the ADE corpus for the identification of drugs and conditions mentions in text.

The current experiments have been performed on the ADE corpus, since that was the only one available when this work was done, however while writing this report a new corpus has been published, namely the EU-ADR corpus (van Mulligen et al., 2012). It will be interesting to see if the performance of JSRE on the ADE corpus will be different compared to the EU-ADR corpus.
Similarly, benchmarking results of commercial and public relation extraction systems such as SemRep, Luxid MER Skill Cartridge®, ReLex, MedScan will be performed. The outcome of relation extraction from text to support signal detection and identify potentially novel or under-reported adverse effects will be studied.

The use of ontologies for driving information extraction has been reported (Wimalasuriya and Dou, 2010; Pandit and Honavar, 2010), we plan to explore the use of various available tools (e.g. ODIE, semantixs) using the AEO ontology and compare the performance of the ontology driven methods against the method presented here.

An outcome of the current work has demonstrated promising results and it has a potential to reduce the manual reading time, accelerate the signal tracking process, and therefore ensure safe use of drugs in the market.

5 ACKNOWLEDGEMENTS

This work has been partly supported by Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany and Bonn-Aachen International Center for Information Technology, Bonn, Germany.

REFERENCES


In 22nd IEEE International Conference on tools with artificial intelligence (ICTAI), pages 173–178.


