Extraction of Lactation Frames from Drug Labels and LactMed

Heath Goodrum, Meghana Gudala, Ankita Misra, Kirk Roberts
School of Biomedical Informatics
University of Texas Health Science Center at Houston
{heath.goodrum,kirk.roberts}@uth.tmc.edu

Abstract

This paper describes a natural language processing (NLP) approach to extracting lactation-specific drug information from two sources: FDA-mandated drug labels and the NLM Drugs and Lactation Database (LactMed). A frame semantic approach is utilized, and the paper describes the selected frames, their annotation on a set of 900 sections from drug labels and LactMed articles, and the NLP system to extract such frame instances automatically. The ultimate goal of the project is to use such a system to identify discrepancies in lactation-related drug information between these resources.

1 Introduction

Medical information about prescription drugs is publicly available in a variety of sources, including the biomedical literature, consumer-focused websites, and the drug labels mandated by the U.S. Food & Drug Administration (FDA). But the rapid advances in biomedicine—especially recently-approved drugs—threatens to make these sources discordant. Synchronizing such sources is difficult due to their unstructured nature and the wide variety of ways in which they are organized. This paper presents initial work in an effort to align two such sources—drug labels and a single consumer health website—for information particular to a single sub-population—nursing mothers.

This is a critical sub-population for providing validated health information to, especially for prescription drugs. Notably, randomized trials, the gold standard in drug evaluation, contain few if any nursing mothers in their trial populations. Thus, the information that a pharmaceutical substance has on such mothers and their children is scarce and of poor evidence quality, which only serves to promote misinformation and discourage mothers from taking needed medications. Authoritative guidance is critical in regards to what is supported or contradicted by the limited evidence, as well as what is simply unknown. Several public sources attempt to provide such authoritative information. Here, two such sources are studied: a section of drug labels specific to nursing mothers and a government website specific to drugs and lactation, LactMed. By identifying discrepancies in the free text narratives of these sources, further review can pinpoint information gaps, conflicting opinions, and out-of-date guidance.

The general strategy proposed in this paper involves (a) identifying seven key information types of drug information specific to nursing mothers, (b) utilizing linguistically-motivated frame semantic representations for these information types, (c) annotating instances of these frames on both lactation information sources, and (d) developing natural language processing (NLP) methods to extract this information automatically from these sources.

Our specific contributions include:

1. The first NLP method to focus specifically on drug information for nursing mothers.

2. Development of frame representations for lactation-specific drug information.

3. Application of a deep learning-based system on two separate lactation information sources, drug labels and LactMed.

4. Evaluation of cross-corpus similarity in terms of important lactation information.

While this paper's scope is quite narrow, just lactation information from two sources, we posit the techniques described here are generalizable to other lactation information sources (with minimal annotation/training) as well as to other important pharmaceutical sub-populations (with, albeit, considerable annotation effort).

2 Related Work

The existing work related to that proposed here is broken down into information extraction efforts on drug labels (§2.1), maternal health in particular (§2.2), and frame semantics in biomedicine (§2.3).

2.1 Drug Label Information Extraction

Drug labels contain a wealth of unstructured information relating to FDA-approved pharmaceuticals, and thus have proven to be a consistent target of NLP-based systems interested in automatically creating knowledge bases (KBs) (Harpaz et al., 2014). For instance, SIDER (Kuhn et al., 2010, 2016) is a well-used KB, constructed from drug labels, for adverse drug reaction (ADR) information (i.e., side effects). The 2017 TAC ADR task (Roberts et al., 2017) utilized a corpus of 200 drug labels with sections specific to ADR information (Demner-Fushman et al., 2018b). On the other hand, Duke et al. (2013) demonstrated the dangers of using drug labels as an ADR KB by identifying numerous inconsistencies between the labels for bioequivalent drugs. Meanwhile, drug indications (i.e., the medical condition the drug is intended to treat) have also been well-studied (Névéol and Lu, 2010; Fung et al., 2013; Li et al., 2013; Khare et al., 2014), as have drug interactions (Demner-Fushman et al., 2018a). All of these focus on general aspects of a drug, while hardly any work has focused on the information in drug labels related to specific populations, though both the TAC task as well as Culbertson et al. (2014) identified ADR-population relations.

2.2 Maternal Health Information Extraction

A few NLP methods have been applied to support maternal health. This includes processing biomedical literature to support evidence-based review of maternal mortality (de Groot et al., 2015) and identifying genes associated with placenta-mediated maternal diseases (Rodriguez et al., 2017). Electronic health record (EHR) data has been used to identify important maternal health information (Borra et al., 2013; Abhyankar and Demner-Fushman, 2013) and screen for suicide (Zhong et al., 2018, 2019). Social media has been used to identify pregnant women (Chandrashekar et al., 2017; Sarker et al., 2017). Finally, only one known work focuses drug labels for maternal health, focusing on the identification of pregnancy risk categories (Rodriguez and Fushman, 2015).

2.3 Frame Semantics in Biomedicine

Frame semantics (Fillmore, 1976, 1982) is a linguistic theory that postulates the meaning of most words is understood in relation to a conceptual frame in which entities take part. E.g., the meaning of sell in the "*Jerry sold a car to Chuck*" evokes a frame related to COMMERCE, which includes four elements: BUYER, SELLER, MONEY, and GOODS, though not all elements are required (as with MONEY here). Frames also include a lexical unit that triggers the frame ("*sold*" in the example). Frames provide a good connection between an abstract information representation and the actual text that specifies that information, and is thus a natural choice for a task such as identifying detailed lactation information in drug labels. Most notably, frame semantics have been operationalized in the large-scale resource FrameNet (Baker et al., 1998, 2003), though this resource is not specific to biomedicine.

Several works have explicitly extended FrameNet for biomedical tasks. This includes frame for molecular biology information (Dolbey et al., 2006; Dolbey, 2009; Tan, 2014), cancer information from EHRs (Roberts et al., 2018; Si and Roberts, 2018; Datta et al., 2017), and general medical information for Swedish (Kokkinakis, 2013). Many other works have implicitly used representations that are similar to frames, including the TAC ADR task data on drug labels (Roberts et al., 2017; Demner-Fushman et al., 2018b).

3 Data

Two different datasets were used to create the text corpus for frame annotation. Section 3.1 describes the drug labels dataset and Section 3.2 describes the LactMed dataset.

3.1 Lactation Information in Drug Labels

Drug labels were downloaded in August 2018 from the full release collection made available by DailyMed[1]. DailyMed is a public website operated by the National Library of Medicine (NLM) and is the official provider of FDA label information. These labels are maintained in a document markup standard approved by Health Level Seven (HL7) referred to as Structured Product Labeling (SPL), which specifies various drug label sections.

---

[1]https://dailymed.nlm.nih.gov/dailymed/

Figure 1: Example "Lactation" section of a drug label

Figure 2: Example "Use During Lactation" section from LactMed

For this work, only the lactation section was extracted. An example of this section is shown in Figure 1.

3.2 LactMed

LactMed[2] is a database created by the National Library of Medicine under the collection of TOXNET databases. LactMed provides information about various drugs and chemicals that nursing mothers may be exposed to that may then be passed to their infant through breast feeding. Information provided in LactMed includes the amount of a substance that may be excreted into breast milk, the absorption rate of an infant, and any potential adverse effects to a nursing infant. Data in LactMed is derived from reviews of the scientific literature, with each entry including references. Additionally, all records are peer-reviewed by a panel of experts. For this work, only the "Summary of Use During Lactation" section was extracted for each LactMed article. An example of this section is shown in Figure 2.

3.3 Preprocessing

Each individual drug label is stored in the Daily-Med collection as a zip compressed folder that includes the drug label as an XML file and scanned images of the label. We extracted the folders and parsed each XML document to identify the relevant lactation information. While the drug labels provide additional information regarding the use of the drug, only section 8.2, "Lactation", was extracted into individual documents for each label. Prior to a specification change in June 2015,

---

[2]https://toxnet.nlm.nih.gov/newtoxnet/lactmed.htm

this section was labeled as section 8.3, "Nursing Mothers". There were 37,005 separate drug labels parsed. Of those, lactation information was identified in 31,309 drug labels. Additionally, since many drug labels exist for the same drug, due to multiple manufactures and dosage amounts, only the lactation information from the most recent label for a drug was extracted. After this process of selecting only unique drug labels based on name, a dataset of 4,486 documents was created.

The entirety of LactMed is made available as a single XML document. This file was parsed to identify the drug name and the Summary of Use During Lactation section. Each LactMed article was already unique, therefore no de-duplication process is required. In total, 1,151 documents were created from LactMed.

4 Lactation Frames

Section 4.1 describes the frames annotated for both the drug labels and LactMed. Section 4.2 describes the annotation process.

4.1 Frame Descriptions

Since one of the primary purposes of annotating these two datasets is to compare information between them, a standard set of frames was chosen that would be applicable to both datasets. Seven lactation-related frames were chosen based on an initial review of sample drug labels and LactMed entries. These frames, detailed in Table 1, are: INFORMATION_AVAILABILITY, EFFECT_ON_MILK_SUPPLY, EXCRETION_INTO_MILK, ABSORPTION, ADVERSE_REACTION, ALTERNATIVES, and VERDICT. For each of these frames, elements

Table 1: Frames and Frame Elements

Element	Description
Non-Core Elements - Elements that are common across all/most frames.	
ANIMAL	Marks non-humans to which the frame applies. Frequently information is only available in animals studies and as not been verified/observed in human studies.
CONDITION	A condition (specific circumstance) under which the rest of the frame applies.
DRUG	The name of the drug or the class of drugs to which the frame applies.
INFORMATION	Any reference to how the information was obtained/published or the information quality that results in the frame's information.
LIKELIHOOD	Any expression that suggests the frame is less than 100% positive, including hedging ("possible"), infrequency ("sometimes"), and negation ("no evidence").
???	Marks any span that the annotator feels is important but does not currently have an annotation to match.
INFORMATION_AVAILABILITY - The quantity/quality of lactation information for the drug.	
QUALITY	A reference to the quality of information available. (e.g., observational studies, randomized controlled trials)
QUANTITY	A reference to the quantity of information available (e.g., a large number of studies, minimal information)
SOURCE	The source of information (e.g., journal article, post marketing surveillance)
EFFECT_ON_MILK_SUPPLY - The impact the drug has on the overall milk supply.	
QUALITY	A reference to the change in quality of the breast milk due to the drug.
QUANTITY	A quantitative expression of the impact of the drug on the milk supply.
TREND	The generalized trend (e.g., increases, decreases) in milk supply due to the drug.
EXCRETION_INTO_MILK - Information that the drug is excreted into the breast milk.	
QUANTITY	A quantitative expression of how much of the drug (or other substance) is excreted into the breast milk.
TIMEFRAME	Either the span of time from taking the medication till initial excretion (e.g., "2 hours after taking") or the span of time (possibly half-life) until the drug will no longer be excreted (e.g., "within 4 days")
ABSORPTION - Information that the nursing infant absorbs the drug from the breast milk.	
QUANTITY	A quantitative expression of how much of the drug (or other substance) is actually absorbed by the infant from the breast milk.
TIMEFRAME	Some span of time related to the absorption of the drug/substance by the infant.
ADVERSE_REACTION - Reactions the infant may have from being exposed to the drug.	
REACTION	The adverse reaction resulting from the drug.
ALTERNATIVES - Alternative drug options for breastfeeding mothers.	
ALTERNATIVE	The name of the alternative drug, drug class, or agent.
PREFERENCE	A statement about the preference for the alternative, which can be positive ("preferred") or negative ("not recommended").
VERDICT - Recommendations for nursing mothers using the drug.	
POLARITY	Positive or negative verdict.
DECISION	What the nursing mother taking the drug should do (or not do).
MONITOR	Statement that the mother/child should be monitored (e.g., for adverse reactions).
REASON	The particular reason leading to the verdict decision.

	DL	LM	Total	Length
FRAMES				
EXCRETION_INTO_MILK	631	222	853	1.24
VERDICT	492	360	852	1.10
ADVERSE_REACTION	376	351	727	1.84
EFFECT_ON_MILK_SUPPLY	132	52	184	1.70
INFORMATION_AVAILABILITY	15	113	128	1.09
ABSORPTION	30	96	126	1.09
ALTERNATIVES	3	111	114	1.33
ELEMENTS				
DRUG	1452	901	2353	1.31
CONDITION	603	639	1242	4.66
REACTION	606	554	1160	1.92
DECISION	669	357	1026	2.84
INFORMATION	563	341	904	2.53
REASON	310	283	593	5.57
QUANTITY	202	315	517	2.15
MONITOR	37	118	155	1.97
ANIMAL	128	3	131	1.09
PREFERENCE	2	101	103	1.62
TREND	45	42	87	1.31
TIMEFRAME	21	12	33	4.79
POLARITY	3	26	29	1.41
QUALITY	16	1	17	1.08
SOURCE	3	14	17	1.50
ALTERNATIVE	0	17	17	2.67

Table 2: Frame and Element Frequency and average number of tokens. DL: drug labels, LM: LactMed, Length: Average length of each lexical unit or frame element (in tokens)

	LM	DL	DL + LM
FRAMES			
INFORMATION_AVAILABILITY	73.64	28.57	64.24
EFFECT_ON_MILK_SUPPLY	62.30	92.00	83.41
EXCRETION_INTO_MILK	85.14	72.47	76.39
ABSORPTION	74.84	63.64	72.36
ADVERSE_REACTION	66.24	63.10	64.84
ALTERNATIVES	30.25	0.00	27.69
VERDICT	56.77	64.56	61.09
ELEMENTS			
DRUG	88.22	81.92	84.79
ANIMAL	66.67	77.91	77.71
QUANTITY	72.66	35.16	60.67
INFORMATION	56.77	62.92	60.40
TREND	60.38	57.63	58.93
MONITOR	63.77	40.00	58.43
LIKELIHOOD	67.52	31.19	53.52
REASON	29.31	61.84	47.76
DECISION	41.78	48.63	46.30
QUALITY	0.00	50.00	46.15
REACTION	41.94	39.31	45.77
ALTERNATIVE	38.71	0.00	36.36
CONDITION	43.29	24.62	32.62
TIMEFRAME	47.62	12.90	26.92
PREFERENCE	29.21	0.00	26.53
POLARITY	40.00	0.00	24.49
SOURCE	31.58	13.33	23.53

Table 3: Inter-Annotator Agreement, measured by F1.

were selected that describe the individual attributes and relations for the frame. These elements are where the detailed semantic information is located. Certain elements were selected that exist across all frames, these are referred to as Non-Core Elements.

4.2 Annotation

A random subset of equal amounts of documents from the drug labels dataset and LactMed dataset were selected for manual annotation. Example annotations from LactMed articles and the drug labels are shown in Figure 3.

The annotation process was completed by three individuals using BRAT (Stenetorp et al., 2013). Documents were double-annotated with a pair of individuals first annotating a collection of documents independently and then meeting to reconcile any differences. In cases where annotations could not be easily reconciled, the case was presented to two other individuals to help establish rules which could be used in similar situations moving forward. Annotation guidelines, which included frequently occurring lexical units for a given frame and example annotations, were devel-

oped in order to identify and ensure consistency.

After annotation of each subset of documents was completed and reconciled, a final review was performed by one of the annotators to ensure that any newly-established guidelines were consistent throughout all documents.

In total, 900 documents were double-annotated, 450 drug labels and 450 LactMed entries. Within these 900 documents a total of 2,984 frames and 8,384 frame elements were annotated. The frequency breakdown for each frame and frame element type is shown in Table 2. The most frequently identified frames were EXCRETION_INTO_MILK with 853 frames, VERDICT with 852 frames, and ADVERSE_REACTION with 727 frames.

Table 3 shows the inter-annotator agreement for each frame and frame element. When determining the inter-annotator agreement, only exact matches are considered, though partial disagreements were quite common. For example if one annotator choose the lexical unit "breast milk" for an EXCRETION_INTO_MILK frame and the second annotator choose "milk", this would be considered a mismatch. (The annotation guidelines specify that "breast milk" is the correct lexical unit in such a case.)
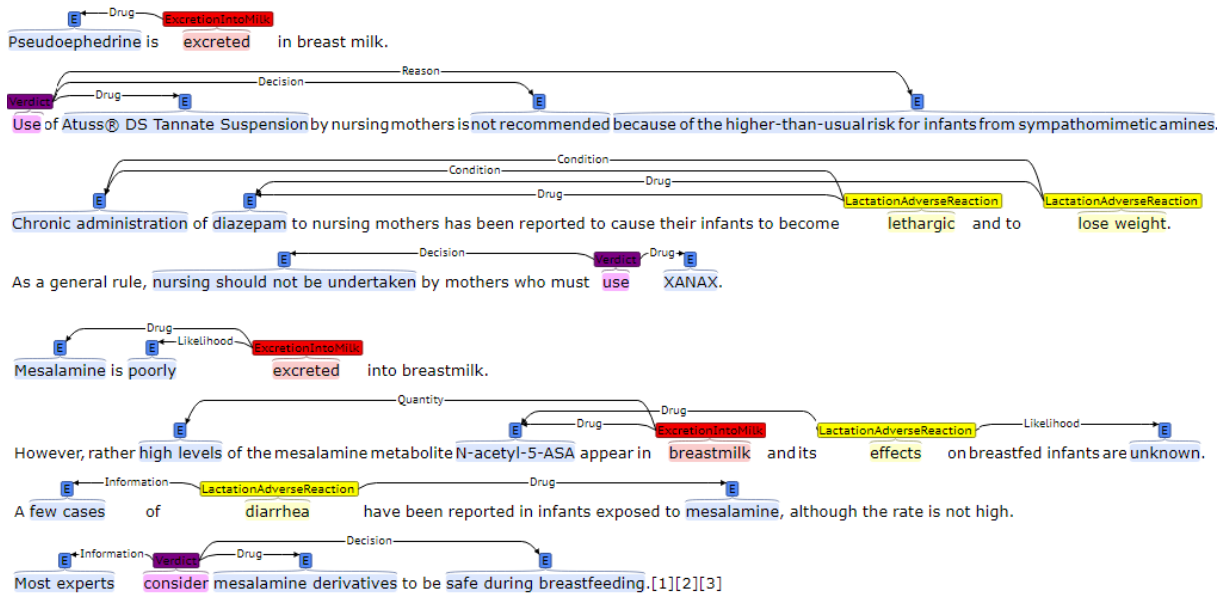
Figure 3: Annotation Examples

5 Extraction

A standard bi-directional Long Short-Term Memory (Bi-LSTM) Conditional Random Field (CRF) was utilized to extract lactation frames. The Bi-LSTM utilizes both character embeddings (dynamic) and word embeddings (static, described below). Specifically, a pipeline approach was used that extracts frames and frame elements are identified in two separate steps. The first step identifies lexical units in a sentence for all potential frames, essentially equivalent to a named entity recognition approach. The second step performs relation extraction for each identified lexical unit, identifying frame elements associated for the frames identified in the first step.

We experimented with four different embedding corpora: (i) pre-built 300-dimension GloVe (Pennington et al., 2014) embeddings built from Wikipedia; (ii) pre-built 100-dimension word2vec (Mikolov et al., 2013) embeddings built from MIMIC-III (Johnson et al., 2016); (iii) pre-built 200-dimension word2vec embeddings built from PubMed, PMC, and Wikipedia (Pyysalo et al., 2013); and (iv) specially-built 300-dimension GloVe embeddings on the lactation data (all drug labels and LactMed articles combined). Section 7 describes experiments with how these embedding combinations perform.

6 Evaluation

For our evaluation we created three separate collections of training, test, and validation sets by

Embeddings	Frame	FE
MIMIC + GloVe (Wikipedia)	84.53	71.33
MIMIC + GloVe (DL + LM)	83.89	70.93
W2V (PubMed + PMC + Wikipeida)	82.57	73.31
GloVe (DL + LM)	80.78	68.67

Table 4: Experiment with different word embeddings, measured by F1. DL: drug labels; LM: LactMed; FE: frame elements.

splitting the documents of the drug labels (DL), LactMed (LM), and LactMed and drug label combined (DL+LM). 80 percent of each dataset was used for training, 10 percent for testing, and 10 percent for validation. We trained and tested on various combinations of datasets.

We also experimented with training and testing on the different combinations of datasets, for example training on drug labels and LactMed (DL+LM) and testing on LactMed (LM), or training on LactMed (LM) and testing on the drug labels (DL).

Finally, to determine the effect that creating more manual annotations may improve the results of our model we generated a learning curve, using the full LactMed and drug label combined datasets. For generation of the learning curve, the same testing set was maintained and documents were added to the training set 50 documents at a time to generate a new model and evaluate against the test set.

Train	Test	Frame	FE
DL + LM	DL + LM	84.53	71.33
DL + LM	DL	90.07	78.32
DL + LM	LM	77.18	65.49
DL	DL	88.52	76.7
DL	LM	51.49	12.15
DL	DL + LM	68.9	56.24
LM	DL	57.41	27.9
LM	LM	71.54	52.03
LM	DL + LM	67.81	30.81

Table 5: Experiment with different train/test combinations, measured by F1, using best system from Table 4. DL: drug label; LM: LactMed; FE: frame elements

7 Results

Table 4 shows the results of our different embedding experiments. The GloVe embeddings generated from the drug labels and LactMed articles combined with the MIMIC embeddings, which had an F1 measure for the frames of 83.89 and the elements of 70.93, performed slightly worse than the top performing embeddings. The GloVe (Wikipedia) embeddings combined with MIMIC embeddings had an F1 of 84.53 for frames and 71.33 for frame elements. The W2V (PubMed + PMC + Wikipeida) embeddings performed best on frame element extraction with an F1 of 73.31.

Table 5 shows the different combinations of training and testing on various datasets. This data shows that training on the drug labels and LactMed together does improve the prediction performance on a single dataset opposed to just training on one dataset alone. For example, the model that was trained on drug labels and LactMed performed better on the LactMed test set (frame F1 of 77.18) than the model that was trained only on the LactMed dataset (frame F1 of 71.54). This effect is likely caused by an increase in training data overcoming the differences between the datasets.

Table 6 shows the breakdown of the results by each frame and frame element for the model that was created using the embeddings that performed best on frame identification (MIMIC + GloVe (Wikipedia)) and the combined drug labels and LactMed dataset for training and testing.

Figure 4 shows the learning curve as additional documents were added to the training set and the effect it has on the overall F1-measure. For both the frame and frame element the curve is beginning to level off, however it does seem to show that additional training data may continue to have a positive effect on the overall F1 for both cases.

	P	R	F1
FRAME			
OVERALL	85.52	83.55	84.53
ABSORPTION	71.43	35.71	47.62
ALTERNATIVES	86.67	86.67	86.67
EFFECT_ON_MILK_SUPPLY	100.0	85.71	92.31
EXCRETION_INTO_MILK	86.21	83.33	84.75
INFORMATION_AVAILABILITY	100.0	91.67	95.65
ADVERSE REACTION	89.71	85.92	87.77
VERDICT	78.02	87.65	82.56
ELEMENT			
OVERALL	69.74	72.99	71.33
ALTERNATIVE	100.0	50.00	66.67
ANIMAL	76.47	92.86	83.87
CONDITION	54.19	67.20	60.00
DECISION	76.84	71.57	74.11
DRUG	74.52	83.55	78.78
INFORMATION	92.59	82.42	87.21
LIKELIHOOD	68.12	74.60	71.21
MONITOR	44.19	76.00	55.88
POLARITY	100.0	75.00	85.71
PREFERENCE	81.25	100.0	89.66
QUALITY	100.0	100.0	100.0
QUANTITY	62.75	55.17	58.72
REACTION	92.86	50.00	65.00
REASON	54.39	58.49	56.36
SOURCE	100.0	100.0	100.0
TIMEFRAME	33.33	16.67	22.22
TREND	66.67	76.92	71.43

Table 6: Frame and Element Breakdown

8 Discussion

This paper addresses a critical component for assessing the consistency of drug information for nursing mothers, namely the information extraction techniques to extract semi-structured information from two drug information sources: manufacturer-supplied drug labels and expert-sourced LactMed. A frame-based approach was devised utilizing seven frames dealing with the availability/quality of lactation information, the effects the drug has on a mother's milk supply, the degree to which the drug is excreted into the milk, the degree to which that drug is absorbed into the child's body, any potential adverse reactions the child may experience due to breastfeeding, recommended alternative drugs while nursing, and any general statements or verdicts on what nursing mothers should do as it relates to the particular drug. Each of these seven frames was double-annotated on a corpus of 450 drug label sections and 450 LactMed article summaries. A standard Bi-LSTM-CRF combining character and word embeddings is trained to extract these frames automatically. Experiments were performed to assess the best set of embeddings to use, the transferability of drug label and LactMed annotations, and
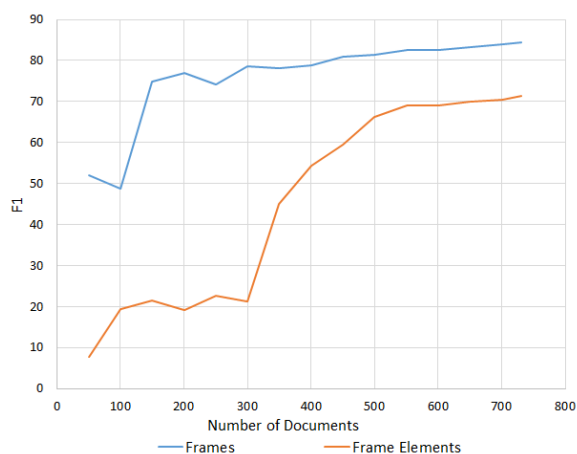
Figure 4: Learning curve of test performance with increasing amounts of training data.

whether sufficient annotated data exists to maximize frame extraction performance. These experiments yield several observations that have implications on further development of such a frame extraction system.

First, the fact that open-domain embeddings outperformed embeddings trained on the drug labels and LactMed (see Table 4) can be considered a negative, but not entirely conclusive, result. In our initial error analysis it was clear that the lack of embedding information for common terms in the dataset (such as particular drug names) resulted in numerous errors. We did not experiment with additional embedding combinations, such as concatenating separate embeddings for Wikipedia, MIMIC, and drug label/LactMed, though this concatenation strategy has shown promise in other biomedical tasks (Roberts, 2016).

Second, the experiments demonstrated that training on both drug labels and LactMed improves performance over training on each individually (Table 5). This improvement is despite the fact that the drug labels and LactMed data appears to be quite different, as can be seen by comparing the result of training on one and testing on the other. For instance, LactMed results are quite poor when only training on drug labels (51.49 F1) and improve significantly when training on LactMed (71.54), but improve further still when training on both drug labels and LactMed (77.18). This would suggest that there is sufficient similarity to train on both, but perhaps domain adaptation methods could be employed to gain the benefits of larger training datasets while still identifying source-specific differences in the data.

Third, the amount of data available for training (Figure 4) suggest small gains are still likely to be expected given more data. Our error analysis, however, suggested that many of the "errors" could in fact be considered legitimate frame instances. This is typically a result of inconsistent frame annotation, which is of course quite common in complex semantic annotation tasks. However, it is clear that further quality control on the existing annotations will likely be a more promising effort prior to adding further annotations.

Beyond the work described in this paper, there is still a good distance to go before an automatic method exists for detecting inconsistencies in lactation information sources. Notably, this work only extracts the basic frame instances from each of these sources, but does nothing to compare frames. Future work will thus be necessary to compare frame instances from a drug label and its corresponding LactMed article (not to mention that there are often multiple labels per drug). Comparing frames is certainly much easier than comparing full documents, but not without its challenges. Comparing individual frames requires both frame element-specific comparisons (e.g., is "*safe during breastfeeding*" equivalent to "*no major concerns*") as well as comparing to null frame elements (e.g., if one frame has a QUANTITY of "*high levels*" but the other frame has no QUANTITY at all). It is unlikely simple rule-based procedures can be used to identify equivalent or contradictory frames with high accuracy. However, this need not be the goal. Instead of providing a complete list of all inconsistent labels/articles, a likely application for the use of such a system is to provide a ranked list of labels/articles that are most likely to be incongruous. This approach may have greater robustness to errors in frame matching, and is a likely direction of future work.

9   Conclusion

This paper described a frame-based approach for lactation information extraction from drug labels and LactMed. Seven lactation-related frames were identified, manually annotated, and automatically extracted using a standard NLP approach. Future work will involve utilizing this system in order to identify discordant information present in drug information sources for nursing mothers.

References

Swapna Abhyankar and Dina Demner-Fushman. 2013. A simple method to extract key maternal data from neonatal clinical notes. In *Proceedings of the AMIA Annual Symposium*, pages 2–9.

Collin F Baker, Charles J Fillmore, and Beau Cronin. 2003. The Structure of the Framenet Database. *International Journal of Lexicography*, 16(3):281–296.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project.

Allan Borra, Karla Delos Santos, Daniel Gonzales, and Alden Reyes. 2013. Maternal Medication Information Extraction (MaMIE) System. In *Proceedings of the 9th National Natural Language Processing Research Symposium*, pages 25–29.

Pramod Bharadwaj Chandrashekar, Arjun Magge, Abeed Sarker, and Graciela Gonzalez. 2017. Social media mining for identification and exploration of health-related information from pregnant women. *CoRR*, abs/1702.02261.

Adam Culbertson, Marcelo Fiszman, Dongwook Shin, and Thomas C. Rindflesch. 2014. Semantic Processing to Identify Adverse Drug Event Information from Black Box Warnings. In *Proceedings of the AMIA Annual Symposium*, pages 442–448.

Surabhi Datta, Elmer V Bernstam, and Kirk Roberts. 2017. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. *CoRR*, abs/1904.01655.

Christianne J. M. de Groot, Thed van Leeuwen, Ben Willem J. Mol, and Ludo Waltman. 2015. A Longitudinal Analysis of Publications on Maternal Mortality. *Paediatric and Perinatal Epidemiology*, 29(6):481–489.

Dina Demner-Fushman, Kin Wah Fung, Phong Do, Richard D. Boyce, and Travis R. Goodwin. 2018a. Overview of the TAC 2018 Drug-Drug Interaction Extraction from Drug Labels Track. In *Proceedings of the Text Analysis Conference*.

Dina Demner-Fushman, Sonya E. Shooshan, Laritza Rodriguez, Alan R. Aronson, Francois Lang, Willie Rogers, Kirk Roberts, and Joseph Tonning. 2018b. A dataset of 200 structured product labels annotated for adverse drug reactions. *Scientific Data*, 5:180001.

A. Dolbey. 2009. *BioFrameNet: A FrameNet Extension to the Domain of Molecular Biology*. Ph.D. thesis, UC Berkeley.

A. Dolbey, M. Ellsworth, and J. Scheffzyk. 2006. BioFrameNet: A Domain-specific FrameNet Extension with Links to Biomedical Ontologies. In *Proceedings of the Biomedical Ontology in Action Workshop at KR-MED*, pages 86–94.

Jon Duke, Jeff Friedlin, and Xiaochun Li. 2013. Consistency in the safety labeling of bioequivalent medications. *Pharmacoepidemiology Drug Safety*, 22(3):294–301.

Charles J Fillmore. 1976. Frame Semantics and the Nature of Language. *Annals of the New York Academy of Sciences*, 280(1):20–32.

Charles J Fillmore. 1982. Frame semantics. *Linguistics in the Morning Calm*, pages 111–137.

Kin Wah Fung, Chiang S Jao, and Dina Demner-Fushman. 2013. Extracting drug indication information from structured product labels using natural language processing. *Journal of the American Medical Informatics Association*, 20(3):482–488.

Rave Harpaz, Alison Callahan, Suzanne Tamang, Yen Low, David Odgers, Sam Finlayson, Kenneth Jung, Paea LePendu, and Nigam H. Shah. 2014. Text Mining for Adverse Drug Events: the Promise, Challenges, and State of the Art. *Drug Safety*, 37(10):777–790.

Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, , and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3:160035.

Ritu Khare, Jiao Li, and Zhiyong Lu. 2014. LabeledIn: Cataloging Labeled Indications for Human Drugs. *Journal of Biomedical Informatics*, 52:448–456.

Dimitrios Kokkinakis. 2013. Medical Event Extraction using Frame Semantics – Challenges and Opportunities. *International Journal of Computational Linguistics and Applications*, 4(2):121–133.

Michael Kuhn, Monica Campillos, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. 2010. A side effect resource to capture phenotypic effects of drugs. *Molecular Systems Biology*, 6:343.

Michael Kuhn, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. 2016. The SIDER database of drugs and side effects. *Nucleic Acids Research*, 44(D1):D1075–D1079.

Qi Li, Louise Deleger, Todd Lingren, Haijun Zhai, Megan Kaiser, Laura Stoutenborough, Anil G Jegga, Kevin Bretonnel Cohen, and Imre Solti. 2013. Mining FDA drug labels for medical conditions. *BMC Medical Informatics and Decision Making*, 13(53).

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, pages 3111–3119.

Aurélie Névéol and Zhiyong Lu. 2010. Automatic Integration of Drug Indications from Multiple Health Resources. In *Proceedings of the 1st ACM International Health Informatics Symposium*, pages 666–673.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1532–1543.

Sampo Pyysalo, Filip Ginter, Hans Moen, Tapio Salakoski, and Sophia Ananiadou. 2013. Distributional Semantics Resources for Biomedical Text Processing. In *Proceedings of Languages in Biology and Medicine*.

Kirk Roberts. 2016. Assessing the Corpus Size vs Similarity Trade-off for Word Embeddings in Clinical NLP. In *Proceedings of the COLING Workshop on Clinical Natural Language Processing*.

Kirk Roberts, Dina Demner-Fushman, and Joseph Tonning. 2017. Overview of the TAC 2017 Adverse Reaction Extraction from Drug Labels Track. In *Proceedings of the Text Analysis Conference (TAC)*.

Kirk Roberts, Yuqi Si, Anshul Gandhi, and Elmer Bernstam. 2018. A FrameNet for Cancer Information in Clinical Narratives: Schema and Annotation. In *Proceedings of the Language Resources and Evaluation Conference*, pages 272–279.

Laritza M. Rodriguez and Dina Demner Fushman. 2015. Automatic Classification of Structured Product Labels for Pregnancy Risk Drug Categories, a Machine Learning Approach. In *Proceedings of the AMIA Annual Symposium*, pages 1093–1102.

Laritza M. Rodriguez, Stephanie M. Morrison, Kathleen Greenberg, and Dina Demner Fushman. 2017. Mining the literature for genes associated with placenta-mediated maternal diseases. In *Proceedings of the AMIA Annual Symposium*, pages 1498–1506.

Abeed Sarker, Pramod Chandrashekar, Arjun Magge, Haitao Cai, Ari Klein, and Graciela Gonzalez. 2017. Discovering Cohorts of Pregnant Women From Social Media for Safety Surveillance and Analysis. *Journal of Medical Internet Research*, 19(10):e361.

Yuqi Si and Kirk Roberts. 2018. A Frame-Based NLP System for Cancer-Related Information Extraction. In *Proceedings of the AMIA Annual Symposium*.

Pontus Stenetorp, Sampo Pyysalo, Goran Topic, Tomoko Ohta, Sophia Ananiadou, and Junichi Tsujii. 2013. BRAT: a Web-based Tool for NLP-Assisted Text Annotation. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107.

He Tan. 2014. A System for Building FrameNet-like Corpus for the Biomedical Domain. In *Proceedings of the 5th International Workshop on Health Text Mining and Information Analysis (Louhi)*, pages 46–53.

Qiu-Yue Zhong, Elizabeth W. Karlson, Bizu Gelaye, Sean Finan, Paul Avillach, Jordan W. Smoller, Tianxi Cai, and Michelle A. Williams. 2018. Screening pregnant women for suicidal behavior in electronic medical records: diagnostic codes vs. clinical notes processed by natural language processing. *BMC Medical Informatics and Decision Making*, 18:30.

Qiu-Yue Zhong, Leena P. Mittal Margo D. Nathan, Kara M. Brown, Deborah Knudson Gonzlez, Tianrun Cai, Sean Finan, Bizu Gelaye, Paul Avillach, Jordan W. Smoller, Elizabeth W. Karlson, Tianxi Cai, and Michelle A. Williams. 2019. Use of natural language processing in electronic medical records to identify pregnant women with suicidal behavior: towards a solution to the complex classification problem. *European Journal of Epidemiology*, 34(2):153–162.