Finding Structure in Figurative Language: Metaphor Detection with Topic-based Frames

Hyeju Jang, Keith Maki, Eduard Hovy, Carolyn Penstein Rosé
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{hyejuj, kmaki, hovy, cprose}@cs.cmu.edu

Abstract

In this paper, we present a novel and highly effective method for induction and application of metaphor frame templates as a step toward detecting metaphor in extended discourse. We infer implicit facets of a given metaphor frame using a semi-supervised bootstrapping approach on an unlabeled corpus. Our model applies this frame facet information to metaphor detection, and achieves the state-of-the-art performance on a social media dataset when building upon other proven features in a nonlinear machine learning model. In addition, we illustrate the mechanism through which the frame and topic information enable the more accurate metaphor detection.

1 Introduction

Computational work on metaphor has largely focused on metaphor detection within individual sentences, for the purpose of identification of literal meaning, with an eye towards improvement of downstream applications like Machine Translation. This limited conceptualization of metaphor within these restricted contexts has allowed prior work to leverage local indicators to identify metaphorical language, such as the violation of selectional preferences (Martin, 1996; Shutova et al., 2010; Huang, 2014) or the use of abstract vs concrete descriptors (Turney et al., 2011; Brysbaert et al., 2014; Tsvetkov et al., 2013). When detecting metaphor in an extended discourse, and especially for the purpose of modeling the use of metaphor in interaction, however, a broader conceptualization of metaphor is needed in order to accommodate the many places where these simplifying assumptions break down (Jang et al., 2015, 2016). Detection of metaphors in naturalistic discourse remains an open problem.

To begin to address this gap, this paper suggests adopting a concept of *framing* in discourse (Tannen, 1993; Tannen and Wallat, 1987; Gee, 2014; Minsky, 1975; Schank and Abelson, 1975; Fillmore, 1976; Fauconnier and Turner, 1998). *Framing* is a well-known approach for conceptualizing discourse processes, with variants that have arisen in linguistics, cognitive psychology, and artificial intelligence. This approach stands in contrast to conceptualizations of metaphor as a violation of narrowly defined linguistic rules such as selectional restrictions, instead adopting a softer, Gricean notion that an expectation of coherence broadly construed has been flouted. Specifically, a metaphor occurs when a speaker brings one frame into a context governed by another frame, and explicitly relates parts of each, so that the original frame's expectations are extended or enhanced according to the new frame.

We propose a novel and highly effective method for induction and application of metaphor frame templates as a step toward detecting metaphor in an extended discourse. Our contributions are three-fold. (1) We computationally induce frames, which can be either metaphorically or literally used, from unannotated text. Our approach infers the facets of a given frame through template induction using a semi-supervised bootstrapping approach. Then, (2) we evaluate the obtained template in an established metaphor detection task which distinguishes whether a target word from the given frame is used metaphorically or literally in text. We demonstrate that this frame information is effective in metaphor detection in combination with features from Jang et al. (2016) in a nonlinear machine learning model, which significantly outperforms Jang et al. (2016), the state-of-the-art baseline on a social media dataset. Ad-

Proceedings of the SIGDIAL 2017 Conference, pages 320–330,
Saarbrücken, Germany, 15-17 August 2017.

ditionally, (3) through error analysis, we illustrate the mechanism through which the frame and topic information that are germane to our approach enable the more accurate metaphor detection it achieves. Frame switching can occur not only for metaphor but also for other reasons e.g., topic switches. Our model provides more fine-grained information about what pieces of the frame make the frame metaphorical or literal. Specifically, in our model, semantically-related words from the same frame that co-exist around a target word aid metaphor detection whereas they confuse metaphor detection in other prior approaches.

The remainder of the paper is organized as follows. Section 2 relates our work to prior work. Section 3 shows how adopting the concept of a *frame* may be useful for studying metaphor in discourse from a social perspective. Section 4 explains our semi-supervised approach of template induction to model a metaphor frame in detail. Section 5 presents the effectiveness of the frame information through metaphor detection experiments. Section 6 analyzes the results and identifies when the frame information is beneficial. Section 7 concludes the paper.

2 Relation to Prior Work

In this section, we discuss previous computational work on metaphor that is most relevant to our study. (For more thorough review, refer to (Shutova, 2015).) Next, Section 2.1 introduces approaches to metaphor detection by modeling metaphorical mapping patterns instead of relying on the idea of violation of linguistic expectations. Section 2.2 reviews work that specifically aims to address problems of metaphor detection in discourse. As a direction related to metaphor detection, Section 2.3 introduces computational work that extracts properties of similes, which provides inspiration for our template induction approach used to induce properties (facets) of a metaphor frame.

2.1 Modeling Metaphorical Mapping

There are many different types of metaphor including metaphors that do not violate any local linguistic expectations (Jang et al., 2015, 2016). In order to find other patterns not predicated on the assumption of constraint violation, one might investigate which domains are frequently mapped metaphorically, or what target and source domains are frequently used together in metaphors.

Within these approaches that model frequent target and source domain mappings, Shutova et al. (2010) identified new metaphors by expanding seed metaphors. The idea in this approach is that target concepts that are frequently used with the same source concept occur in similar lexico-syntactic settings. They cluster nouns (target domain) and verbs (source domain), and search the corpus for metaphors that use the verbs in the source domain lexicon to represent the target domain concepts. Extending Shutova et al. (2010), (Shutova and Sun, 2013) find metaphorical mappings by building and traversing a graph of concepts. Then, they generate lists of salient features for the metaphorically connected clusters, and search the corpus for metaphors that use the verbs in the salient features to represent the target domain concepts.

Another approach, Hovy et al. (2013) detected metaphors using certain semantic patterns appearing in metaphor manifestations. For example, "sweet" with *food* is literal, but is metaphorical with *people*. By finding these patterns on different levels, they extended the application of this mapping information from a narrow focus on verb relations to other syntactic relations.

Along the same lines, Mohler et al. (2013) presented a domain-aware semantic signature to capture source and target domains for a text. A semantic signature represents the placement of a text on a semantic space by using a set of related WordNet senses, and it includes source concept dimensions and target concept dimensions. The primary idea is that the signature of a known metaphor is used to detect the same conceptual metaphor.

These approaches are effective for capturing frequent domain specific metaphorical mappings, and in appropriate contexts are helpful for metaphor detection. They also provided valuable insight to our approach. Nevertheless, they may overgeneralize in cases where frequent mappings are metaphorical when applied to an extended discourse.

2.2 Metaphor Detection in Discourse

Other approaches, which share more conceptually with our approach, use context information above the clause level to more directly address problems related to metaphor detection in discourse.

In these contexts, using only local indicators has less predictive power since metaphor is not always confined to a single clause in discourse.

In detection of metaphor in running discourse, coherence in context is an important ingredient. For example, Jang et al. (2015) detected metaphor in discourse focusing on modeling the context of a target word that may or may not have been used metaphorically. They modeled context as global and local, using lexical categories and topic distributions to detect whether cohesion in context was disrupted. In addition, within a sentence, they used the idea that interplay between the target words category and that of other words is indicative of the non-literalness of the target word. Jang et al. (2016), building on the work of Jang et al. (2015), more aggressively tackle the problem that distinguishes metaphorical/literal usage when there has been a recent topic transition. They do so by modeling topic transitions in conjunction with situational context. These approaches begin to grapple with the challenges of leveraging context, but encounter problems when related metaphors co-exist around a target word i.e., extended metaphor. In contrast, in our approach, nearby related words are strategically used to assist rather than obfuscate metaphor detection.

Detecting extended metaphor is important for modeling the use of metaphor in communication. Beigman Klebanov and Beigman (2010) offers an example of studying extended metaphor, showing that extended metaphors can reveal motivations behind metaphor use and the effect of metaphor use on social dynamics in political communication. However, this study was conducted using manually-annotated extended metaphors on a small dataset, and to our knowledge there has been no computational work on detecting extended metaphor. In this paper, we demonstrate promising improvement over prior approaches by leveraging frame facet information on an established metaphor detection task. There is no existing corpus for extended metaphor detection; however, our error analysis suggests that the broad conceptualization of metaphor we employ will be applicable to extended metaphor.

2.3 Extraction of Properties

So far very little computational work has focused on facets, or properties, of metaphor specifically. However, the Qadir et al. (2016) approach automatically infers implicit properties evoked by similes. They generate candidate properties from different sources using a vehicle and an event. Then, properties are evaluated based on the influence of multiple simile components: using PMI or similarity between a candidate property and the second component of a simile, and aggregate ranking of the properties from different sources. This work is similar to our work in that it extracts properties related to the source domain. However, this work only focuses on similes, which have more formulaic structural patterns compared to metaphors, e.g. *He's as cold as ice*. In addition, the grammatical patterns used in their work are fixed manually by human intuition whereas we automatically infer the patterns in our work.

3 Metaphor Frames

A metaphor occurs when a speaker brings one frame into a context/situation governed by another. In this section, we offer a qualitative analysis of the data from this standpoint, and the technical approach described in Section 4 will build on this understanding.

The same or related metaphors from the frame may be used repeatedly. For example, EX(1) compares people to a gun and bullets, and EX(2) compares the world and people to a stage and players. Related metaphors can be used not only within a sentence, but also beyond a sentence. For instance, EX(3) compares the author's imagination to a circus and imagination-related things to circus-related things throughout the paragraph.

EX(1) "He is the pointing gun, we are the bullets of his desire."

EX(2) "All the world's a stage and men and women merely players." (Shakespeare, Twelfth Night)

EX(3) "Bobby Holloway says my imagination is a three-hundred-ring circus. Currently I was in ring two hundred and ninety-nine, with elephants dancing and clowns cart wheeling and tigers leaping through rings of fire. The time had come to step back, leave the main tent, go buy some popcorn and a Coke, bliss out, cool down." (Dean Koontz, Seize the Night. Bantam, 1999)

In the breast cancer discussion forum we use in our work, community participants frequently bring in *journey* and *battle* frames when talking about their cancer experience. Depending on what aspects of the cancer experience they choose to focus on, they invoke different frames accordingly even within the same text. For example, in EX(4), the *journey* and *road* metaphors are used to say that the speaker is having a similar experience with the hearer. Further on, *weapons* from the *battle* frame are used to emphasize the power of faith and prayer in cancer treatment. In this way, metaphor introduces specific facets for specific communicative purposes.

EX(4) "I know, the age thing struck me too when I read about **your bc journey — we have been going down the same road at the same time, only in another part of the country!** It does help to know you are not alone! How amazing with the size of your tumor, that you did not have positive nodes. That is a miracle in itself. I do believe faith and prayer are our most powerful **weapons against this disease**. It is what gets me thru each day."

While metaphor provides resources for the speaker to use in communication, it also creates corresponding resources for the hearer. For example, EX(5)–EX(8) from the same thread in the breast cancer discussion forum shows how conversational participants repeat and expand one another's metaphors. The speaker in EX(5) starts using the *falling off the wagon* metaphorical idiom to convey her opinion that failing to stay on a controlled diet is okay. EX(6) relays the *falling off* part, and connects it to *journey*. EX(7) and EX(8) carry the *wagon* part of the initial post, and use *on the wagon* to describe her status (EX(7)) and her wish to the other person with the extension of *get back on after you fall*. Although *falling off the wagon* and *on the wagon* are metaphorical idioms, *get back on after you fall* is a novel metaphor created by the following speaker. This novel metaphor is drawn from the *wagon* frame that has been brought into this conversation. In this way, a metaphor that is taken up by multiple speakers may increase empathetic understanding as well as add creative opportunities (e.g., for "fun") to the conversation.

EX(5) "**falling off the wagon** is no big thing in my opinion, the psychological good feelings of enjoyment weigh in big for feeling good."

EX(6) "Tina **falling off is part of this journey**, it is stupid to deny yourself everything."

EX(7) "I am **on the wagon** so far today . . . ongoing battle."

EX(8) "Tina — hope you **stay on the wagon**, or at least **get back on after you fall**!"

As shown in the above examples, metaphor performs social functions through the switching of frames. In other words, observing frame switches offers insight into the ways in which people use metaphor to achieve social goals. The goal of our work is to lay a computational foundation for detection of such switches so that social strategies regarding metaphor use in interaction can be accomplished as follow-up work. Thus, in this paper, we empirically construct a metaphor frame, and model the linguistic signals of frame switches.

4 Our Approach

To investigate how a metaphor frame appears in discourse, we computationally model frames that can be either metaphorically or literally used. A frame characterizes a conceptual domain, a "world" that is defined by a number of co-occurring facets. For example, the *journey* domain in *"life is a journey"* or *"he took a journey to Sweden"* could have facets such as *origin*, *destination*, *path*, *vehicle*, *companion*, and *guide*. Using a *journey*-related metaphor activates this domain and its facets, which become available as conversational resources in communication. In our work, we identify facet "slots" of a frame such as the *origin* and *destination* of the *journey* frame, and discover linguistic manifestations of the facets that fill the slots. We later use this frame information for metaphor detection, and observe how the same frame is used metaphorically or literally depending on its facets. We will call the facet slots *facets*, *facet categories*, or *facet slots*, and the linguistic manifestations *facet instances*.

In order to obtain both facets (template slots) and facet instances (slot instances), we propose a simple bootstrapping algorithm (Figure 1) which expands on the number of the facet instances, inspired by earlier bootstrapping approaches such

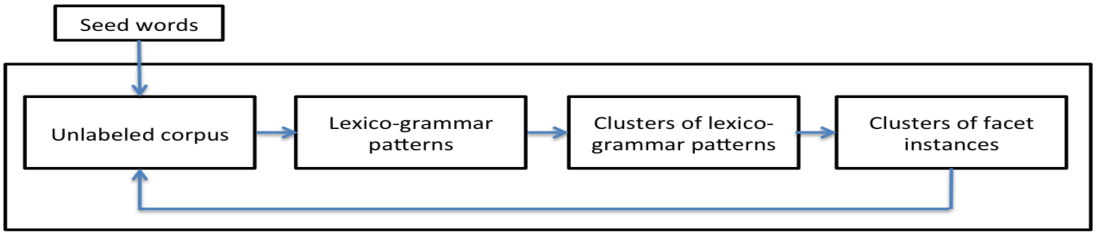

Figure 1: System flow diagram.

as (Riloff et al., 1999, 2003; Qadir and Riloff, 2013). In our model, we assume that a sentence tends to contain more than one important facet of a metaphor frame. In other words, if a sentence contains one facet of a metaphor frame, the sentence is likely to contain additional facets. Additionally, we assume that facets and dependency relations have some relationship. There are certain grammatical patterns that represent semantic relations that connect facets in context. Note that we disregard frame facet instances that do not co-occur with a keyword (e.g., *journey*) within the same sentence. This can be considered as a limitation of this approach.

Our bootstrapping process begins with several seed words (Section 4.1) that specify the domain and provide seed facet instances. Using the seed words, we collect lexico-grammatical patterns (Section 4.2) in unannotated texts and cluster them to find facets (template slots) (Section 4.3). Next, the induced patterns are used for identifying facet instances which comprise a facet cluster (Section 4.4). Then, the most representative facet instances for each cluster are identified and added to the seed word set. Repeating this process expands the seed facet instances and lexico-grammatical patterns into larger sets. The overall sequence is illustrated in Table 1.

1. Harvest sentences containing the seed words from the unannotated texts.
2. Parse the harvested sentences, and obtain lexico-grammatical patterns of the sentences.
3. Cluster the lexico-grammatical patterns.
4. Extract candidate facet instances from the lexico-grammatical patterns in each cluster.
5. Compute the score of each candidate facet instance.
6. Top ranked candidate facet instances of each cluster are added to the original seed words.
7. Repeat starting with step 1.

Table 1: The bootstrapping process.

4.1 Seed Words

The mutual bootstrapping process begins with predefined seed words and a text corpus. The seed words are the frame related words including the domain (e.g. *journey*) and a few examples of representative facet instances (e.g. *train*, *long*) for one or more unspecified facets. The corpus is then filtered for sentences that contain the frame (e.g. *journey*) and at least one example seed facet instance. Note that the sentences in the corpus are not annotated metaphorical or literal. Since we are building a frame that can be used either metaphorically or literally, we do not require sentences where the seed words are used in a desired sense. For this reason, any general corpus that contains sufficient amount of sentences that include frame-related words can be used.

4.2 Collect Lexico-Grammar Patterns

We collect lexico-grammatical patterns using the seed words to represent relations between the domain and its facets. Representing relations in this way is a common approach in event extraction where relations often appear in text within a verb relation. For example, in a Bombing event, *perpetrator* can be represented as a *person/org* who detonates, blows up, plants, hurls, stages, launches, or is detained, suspected, or blamed for the bombing (Chambers and Jurafsky, 2009). However, representing a relation for a domain and its facets for our purpose is not as straightforward as it is in event extraction because facets appear in more diverse ways than merely as verb relations. In particular, facets appear in a diversity of syntactic contexts.

As a solution, we propose using lexico-grammatical patterns generated from dependency paths between a domain word and facet words via the *ROOT*. The lexico-grammatical patterns are defined as the shortest path that passes through the *ROOT* in dependencies between the domain name and seed facet instances. For example, StanfordCoreNLP (Manning et al., 2014) outputs the dependencies in Table 2 for the sentence *She resumed her journey through the city.* The lexico-grammatical pattern that connects *journey* with other candidate property words such as *she* and *city* is defined as the reverse path from *journey* to *ROOT* combined with the path from *ROOT* to *journey*. The paths for the example are shown in Table 3. Words are lemmatized to reduce sparsity.

This lexico-grammatical pattern representation has advantages. First, it allows representing patterns connecting pairs of words in a position invariant manner. For example, in our baseline bootstrapping model, it is difficult to represent the pattern *reach __ of my journey* because *reach* is not located between the slot for a property instance and *journey*. However, using the lexico-grammatical pattern enables formalization of this pattern. Second, the lexico-grammar pattern is not affected by modifiers in the path. For example, the patterns representing the relationships between *journey* and *she*, and between *journey* and *city* do not change even for the sentence "She resumed her long journey through the city", in which *long* has been added.

4.3 Cluster Lexico-Grammar Patterns

Using the idea that lexico-grammar patterns can approximate semantic relations, we first cluster collected lexico-grammar patterns so that each cluster may represent a different relation (facet slot).

The feature representation of each pattern is based on all arguments (e.g., *origin* and *destination* in Table 3) the pattern has in the corpus. For example, the pattern "dobj_r(*origin*, resume), root_r(resume, root), root(root, resume), nmod:through(resume, *destination*)" in Table 3 may have other *origins* and *destinations* in the corpus in addition to many occurrences of "city". We use all arguments appearing with the pattern as features for the pattern, with the feature space size of the whole vocabulary. This is based on the idea that patterns with similar arguments would have similar roles that can be facet slots, which is similar to the distributional hypothesis (Harris, 1954).

For the clustering algorithm, we use Nonnegative Matrix Factorization (NMF) (Lin, 2007). We adopt this algorithm because our feature space is greatly sparse and NMF is effective for sparse data. We use the scikit-learn (Pedregosa et al., 2011) implementation of NMF.

4.4 Identify Representative Facet Instances

After obtaining pattern clusters, we extract tokens that match patterns in each pattern cluster. Tokens extracted for each pattern cluster are facet instance candidates.

Although we have clusters of similar facet instance candidates, there are many noisy instances in each cluster. To determine which instances are most reliable, we score each instance based on how far its generating patterns are from the center of the cluster. Specifically, an instance is scored high if it is found in more patterns in the cluster, and in patterns with higher within-cluster scores. We also take into account how semantically close each instance is to the other words in the same cluster. We use the GloVe vector representations (Pennington et al., 2014) to compute cosine similarity between two words. The scoring formula is shown below, where N_i is the number of different patterns that extracted $word_i$, Sim is the average cosine similarity with all other words in the same cluster, $score\_pattern_k$ is within-cluster score computed by NMF.

$$score(word_i) = Sim * \sum_{k=1}^{N_i} 1 + (.01 * score\_pattern_k) \tag{1}$$

Once the best facet instances are identified in this ranking step, the new instances are added to the original seed words, and the process repeats. The lexico-grammar patterns and property instances are clustered again and rescored after each iteration. The process stops after a specified number of iterations. For our experiments, we found five iterations to be sufficient. We leave an exploration of more heuristic stopping criteria to future work.

5 Evaluation

We evaluate our learned facet clusters, which define a particular metaphor frame template, with

Sentence	She resumed her journey through the city.
Dependencies	nsubj(resumed-2, She-1) root(ROOT-0, resumed-2) nmod:poss(journey-4, her-3) dobj(resumed-2, journey-4) case(city-7, through-5) det(city-7, the-6) nmod:through(resumed-2, city-7)

Table 2: Dependencies from parsed result

origin	destination	pattern
journey	she	dobj_r(*origin*, resume), root_r(resume, root), root(root, resume), nsubj(resume, *destination*)
journey	city	dobj_r(*origin*, resume), root_r(resume, root), root(root, resume), nmod:through(resume, *destination*)

Table 3: Examples of lexico-grammar patterns. _r represents a reverse dependency.

Model	κ	F1	P-L	R-L	P-M	R-M	A
Frame	.204	.602	.381	.369	.826	.833	.732
Unigram	.446	.720	.707	.434	.858	.950	.837
Unigram + Frame	.485	.742	.665	.520	.874	.927	.838
Jang et al. (2016)	.618	.808	.789	.615	.899	.954	.880
Jang et al. (2016) + Frame***	.655	.827	.814	.648	.907	.959	.891

Table 4: Performance on metaphor detection. (**Metrics**) κ: Cohen's kappa, F1: average F1 score on M/L, P-L: precision on literals, R-L: recall on literals, P-M: precision on metaphors, R-M: recall on metaphors, A: accuracy, ***: highly statistically significant ($p < 0.01$) improvement over Jang et al. (2016) by Student's t-test.

respect to how well they perform for an application, metaphor detection. In so doing, we assess the performance of the represented frame information and compare to state-of-the-art models for the same task. The evaluation results are presented in Table 4. The results show that our model performs significantly better than the state-of-the-art model, which indicates that modeling metaphor in terms of frames is promising for distinguishing metaphorical and literal usage of words.

Section 5.1 explains our evaluation task, and which datasets we have used for the evaluation. Section 5.2 describes baseline systems we compare our model with. Section 5.3 illustrates how we model the frame information as features for classification, and explains the classification settings used in our experiments. Finally, Section 5.4 provides the experiment results.

5.1 Evaluation Task

For our experiments, we use the metaphor detection task as in Jang et al. (2016). The task is to decide whether a given target word is metaphorically or literally used. Because there is a set of pre-determined target words, this task is beneficial to see whether the applied model has disambiguating power.

We conducted our metaphor detection experiments on a subset of the breast cancer metaphor dataset annotated by Jang et al. (2015). We chose to work on this dataset because this dataset contains conversational texts so that we can observe how people use metaphor in discourse. In addition, more importantly, this dataset has multiple target metaphors from a single frame, *journey*. From the cross-validation and development datasets used in (Jang et al., 2016), we select the journey-related words *road*, *train*, and *ride* to evaluate the journey frame template we built. We exclude other target words, *spice*, *boat*, *light*, and *candle* for our experiments because they do not belong to the journey frame. After filtering out these target words that are not relevant to the journey frame, the development dataset contains 488 instances, and the cross-validation dataset contains 1,119 instances.

To learn templates for the *journey* frame, we use unannotated data from the BookCorpus (Zhu et al., 2015). The corpus contains 11,038 books in 16 different genres. Particularly for our experiments, we use 74,004,228 sentences from the books, which are provided together with the original book files in the corpus. We use this data instead of more conversational data in order to minimize errors from detecting sentence boundaries and parsing, and to ensure broad topical coverage.

5.2 Baselines

First, we compare our model with a baseline Context Unigram Model that uses all the words in a post as features. Additionally, we compare our model with (Jang et al., 2016), a state-of-the-art model on this dataset. Their model uses sentence-level topic transition features and emotion and cognition related features. We use their best configuration of features, which includes unigram, lexical contrast between a target word and its global and local context (Jang et al., 2015), and topic transition surrounding the target word and emotion and cognition features (Jang et al., 2016). For comparison to approaches using only local indicators, see (Jang et al., 2015).

5.3 Features and Classification Settings

We extract a vector of binary features for each target word to indicate which of the learned facets of the journey frame appear in its immediate context. The presence of each cluster in the same sentence, preceding sentence, and following sentence relative to the target word; as well as the presence of each cluster in any of those three contexts, is indicated respectively by features in a vector of length four times the number of clusters.

We used the support vector machine (SVM) classifier provided in the `LightSIDE` toolkit (Mayfield and Rosé, 2010) with sequential minimal optimization (SMO) and a polynomial kernel of exponent 2. This enables the model to make use of contingencies between features. We expect that in order for a frame to be meaningfully identified, an appropriate topic shift coupled with identification of associated slot fillers in the nearby context is needed. The nonlinearity in this model enables this. For each experiment, we performed 10-fold cross-validation. We also trained the baselines with the same SVM settings.

5.4 Results

The results of our classification experiment are shown in Table 4. We tested our frame features alone (Frame), with context unigram features (Unigram + Frame), and with features from the previous state of the art ((Jang et al., 2016) + Frame).

Adding our frame features to the baselines improved performance in predicting metaphor detection. We see that our features combined with the unigram features slightly improved over the Unigram baseline. However, when our features are combined with the features from Jang et al. (2016), we see large gains in performance, which suggests that there is an synergistic interaction between our frame features and the features from Jang et al. (2016).

6 Discussion

Our experiments show that frame facets that appear in surrounding sentences can be strong indicators of metaphor detection. This is promising, and suggests that observing frame facets can be crucial key to understanding how metaphor is used in discourse. However, the frame facets themselves are not as informative as when used with other features from the baseline. The improved performance when the frame facets are used with baseline features in the nonlinear model suggests that there are interactions among the features. In this section, we discuss the benefits of our model by examining prediction errors of our model and the (Jang et al., 2016) baseline.

The majority of the instances where the baseline model and our model do not agree is where our model improves on classifying literal instances as literal. In these cases, a topic shift is sufficient evidence of a metaphor, but the model without our template slots is not able to determine that. EX(9) and EX(10) show some specific examples where the baseline failed by incorrectly predicting metaphor. In both of these examples, a target word *road* is used literally, but the baseline classified it as metaphorical. Although their own topic transition features correctly captured that there is no topic transition in both cases, in combination with Jang et al. (2015) features, the baseline model did not make a correct prediction.

EX(9) ... Planning on having my right removed then reconstruction on both sides . I am an avid runner , road biker and downhill skier . Was looking at the tram flap. ...

EX(10) ... I did go to my son 's for Christmas , 500 miles away . My husband drove and we spent one night at our daughters to break up the time on the *road* .

When our frame features are added, however, the model correctly predicted that they are literal. This is probably because our frame features that picked up frame facet words surrounding the target word in combination with topic transition features strongly signaled literal usage of the target word. In EX(10), for example, our model picked up the distance word, *miles*, in the sentence prior to the sentence where the target word *road* resides.

From this, we can see that adding the frame facet information allows having more complete frame information for distinguishing metaphorical and literal usage of the topic frame. Our model seems to provide more fine-grained information about what pieces of the frame make it metaphorical or literal.

Conducting an error analysis on the instances where both baseline and our model failed reveals the limitations of using a topic frame based approach in general. EX(11) shows that *train* is used literally in the post. However, because there are different topical words around the target word and there is no other journey frame words, both (Jang et al., 2016) model and our model classify the target word as metaphorical by picking up the topic transition.

EX(11) ... I woke at 2 a.m. because it was so quiet . I could n't hear the frogs or crickets and then I heard a *train* getting louder and louder and then it threw us around . When we got out the giant trees looked like x-mas trees from all the clutter in the tops of them . ...

7 Conclusion

In this paper, we argued that a frame-based approach is useful for metaphor detection and may be useful in subsequent work for studying metaphor from a social perspective. In particular, we described a semi-supervised computational approach for constructing a metaphor frame from unlabeled text. We demonstrated the effectiveness of this frame information in metaphor detection when used together with other proven features in a nonlinear machine learning model, which suggests interactions among the features. We discussed the ways in which the frame and topic information anchor the classifier to allow for more accurate metaphor detection.

Although our approach showed promising results which suggest that how the frame facet information is used in text helps determine the frame's metaphorical usage, applying frame information to metaphor detection in this way has a limitation in scalability – we need to know which frame target words belong to in advance. Our contributions here demonstrated the potential of modeling metaphor through the lens of frame theory; we hope to address scalable ways to leveraging frame information in future work, for example, by automatically detecting primary frames that exist in text.

In addition, we hope to exploit this frame information for detecting extended metaphor, a series of related metaphors under the same frame. Obtaining a metaphor corpus that contains a sufficient amount of extended metaphors is a big challenge. However, once such a dataset becomes available, we believe that the findings from this paper will be applicable in that context.

Acknowledgments

We thank the reviewers for their helpful and insightful comments. This research was supported in part by NSF grant IIS 1546393.

References

Beata Beigman Klebanov and Eyal Beigman. 2010. A game-theoretic model of metaphorical bargaining. In *Proceedings of the 48th annual meeting of the association for computational linguistics*. Association for Computational Linguistics, pages 698–709.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods* 46(3):904–911.

Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*. Association for Computational Linguistics, pages 602–610.

Gilles Fauconnier and Mark Turner. 1998. Conceptual integration networks. *Cognitive science* 22(2):133–187.

Charles J Fillmore. 1976. Frame semantics and the nature of language. In *Annals of the New York*

Academy of Sciences: Conference on the origin and development of language and speech. volume 280, pages 20–32.

James Paul Gee. 2014. *An introduction to discourse analysis: Theory and method*. Routledge, fourth edition.

Zellig S Harris. 1954. Distributional structure. *Word* 10(2-3):146–162.

Dirk Hovy, Shashank Srivastava, Sujay Kumar Jauhar, Mrinmaya Sachan, Kartik Goyal, Huiying Li, Whitney Sanders, and Eduard Hovy. 2013. Identifying metaphorical word use with tree kernels. *Meta4NLP 2013* page 52.

Ting-Hao Kenneth Huang. 2014. Social metaphor detection via topical analysis. In *Sixth International Joint Conference on Natural Language Processing*. page 14.

Hyeju Jang, Yohan Jo, Qinlan Shen, Michael Miller, Seungwhan Moon, and Carolyn Rose. 2016. Metaphor detection with topic transition, emotion and cognition in context. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, pages 216–225. https://doi.org/10.18653/v1/P16-1021.

Hyeju Jang, Seunghwan Moon, Yohan Jo, and Carolyn Penstein Rosé. 2015. Metaphor detection in discourse. In *16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. page 384.

Chih-Jen Lin. 2007. Projected gradient methods for nonnegative matrix factorization. *Neural computation* 19(10):2756–2779.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. pages 55–60. http://www.aclweb.org/anthology/P/P14/P14-5010.

James H Martin. 1996. Computational approaches to figurative language. *Metaphor and Symbol* 11(1):85–100.

Elijah Mayfield and Carolyn Rosé. 2010. An interactive tool for supporting error analysis for text mining. In *Proceedings of the NAACL HLT 2010 Demonstration Session*. Association for Computational Linguistics, pages 25–28.

Marvin Minsky. 1975. A framework for representing knowledge .

Michael Mohler, David Bracewell, David Hinote, and Marc Tomlinson. 2013. Semantic signatures for example-based linguistic metaphor detection. In *Proceedings of the First Workshop on Metaphor in NLP*. pages 27–35.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*. pages 1532–1543. http://www.aclweb.org/anthology/D14-1162.

Ashequl Qadir and Ellen Riloff. 2013. Bootstrapped learning of emotion hashtags# hashtags4you. In *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*. pages 2–11.

Ashequl Qadir, Ellen Riloff, and Marilyn A Walker. 2016. Automatically inferring implicit properties in similes. In *Proceedings of NAACL-HLT*. pages 1223–1232.

Ellen Riloff, Rosie Jones, et al. 1999. Learning dictionaries for information extraction by multi-level bootstrapping. In *AAAI/IAAI*. pages 474–479.

Ellen Riloff, Janyce Wiebe, and Theresa Wilson. 2003. Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*. Association for Computational Linguistics, pages 25–32.

Roger C Schank and Robert P Abelson. 1975. *Scripts, plans, and knowledge*. Yale University.

Ekaterina Shutova. 2015. Design and evaluation of metaphor processing systems. *Computational Linguistics* 41(4):579–623.

Ekaterina Shutova and Lin Sun. 2013. Unsupervised metaphor identification using hierarchical graph factorization clustering. In *HLT-NAACL*. pages 978–988.

Ekaterina Shutova, Lin Sun, and Anna Korhonen. 2010. Metaphor identification using verb and noun clustering. In *Proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics, pages 1002–1010.

Deborah Tannen. 1993. *Framing in discourse*. Oxford University Press.

Deborah Tannen and Cynthia Wallat. 1987. Interactive frames and knowledge schemas in interaction: Examples from a medical examination/interview. *Social Psychology Quarterly* pages 205–216.

Yulia Tsvetkov, Elena Mukomel, and Anatole Gershman. 2013. Cross-lingual metaphor detection using common semantic features .

Peter D Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 Conference on the Empirical Methods in Natural Language Processing*. pages 680–690.

Yukun Zhu, Ryan Kiros, Richard Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *arXiv preprint arXiv:1506.06724*.