Metaphor Detection with Topic Transition, Emotion and Cognition in Context

Hyeju Jang, Yohan Jo, Qinlan Shen, Michael Miller,
Seungwhan Moon, Carolyn P. Rosé
Language Technologies Institute
Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213, USA
{hyejuj,yohanj,qinlans,millerm,seungwhm,cprose}@cs.cmu.edu

Abstract

Metaphor is a common linguistic tool in communication, making its detection in discourse a crucial task for natural language understanding. One popular approach to this challenge is to capture semantic incohesion between a metaphor and the dominant topic of the surrounding text. While these methods are effective, they tend to overclassify target words as metaphorical when they deviate in meaning from its context. We present a new approach that (1) distinguishes literal and non-literal use of target words by examining sentence-level topic transitions and (2) captures the motivation of speakers to express emotions and abstract concepts metaphorically. Experiments on an online breast cancer discussion forum dataset demonstrate a significant improvement in metaphor detection over the state-of-the-art. These experimental results also reveal a tendency toward metaphor usage in personal topics and certain emotional contexts.

1 Introduction

Figurative language is commonly used in human communication ranging from literature to everyday speech. One of the most common forms of non-literal language is *metaphor*, in which two dissimilar concepts are compared. In the utterance, "Time is money" (Lakoff and Johnson, 1980), for example, the concept of "time" is compared to "money" to emphasize that time is valuable. Bringing in information from another domain allows more effective ways of expressing thoughts, feelings, and ideas than only using literal language.

Previous approaches to modeling metaphor have used either the semantic and syntactic information in just the sentence that contains a metaphor (Turney et al., 2011; Tsvetkov et al., 2014), or the context beyond a single sentence (Broadwell et al., 2013; Strzalkowski et al., 2013; Schulder and Hovy, 2014; Klebanov et al., 2015; Jang et al., 2015) to detect topical discrepancy between a candidate metaphor and the dominant theme (See Section 2 for more detailed literature review).

Although previous approaches were effective at capturing some aspects of the governing context of a metaphor, the space of how to best use the contextual information is still wide open. Previous context-based models tend to overclassify literal words as metaphorical if they find semantic contrast with the governing context. These cases manifested in the work by Schulder and Hovy (2014) and Jang et al. (2015) as high recall but low precision for metaphorical instances.

We present a new approach that uses lexical and topical context to resolve the problem of low precision on metaphor detection. To better capture the relevant context surrounding a metaphor, we approach the problem in two directions. First, we hypothesize that topic transition patterns between sentences containing metaphors and their contexts are different from that of literal sentences. To this end, we incorporate several indicators of sentence-level topic transitions as features, such as topic similarity between a sentence and its neighboring sentences, measured by Sentence LDA. Second, we observe that metaphor is often used to express speakers' emotional experiences; we therefore model a speaker's motivation in using metaphor by detecting emotion and cognition words in metaphorical and literal sentences and their contexts.

To demonstrate the efficacy of our approach, we

216

evaluate our system on the metaphor detection task presented by Jang et al. (2015) using a breast cancer discussion forum dataset. This dataset is distinct in that it features metaphors occurring in conversational text, unlike news corpora or other formal texts typical in computational linguistics.

Our contributions are three-fold: (1) We extend the previous approaches for contextually detecting metaphor by exploring topic transitions between a metaphor and its context rather than only detecting lexical discrepancies. In addition, (2) we propose to capture emotional and cognitive content to better uncover speakers' motivation for using metaphors. Lastly, (3) through our empirical evaluation, we find that metaphor occurs more frequently around personal topics.

2 Relation to Prior Work

Research in automatic metaphor detection has spanned from detecting metaphor in limited sets of syntactic constructions to studying the use of metaphor in discourse, with approaches ranging from rule-based methods using lexical resources to statistical machine learning models. Here, we focus in particular on approaches that use context wider than a sentence for metaphor detection. For a more thorough review of metaphor processing systems, refer to Shutova (2015).

The main idea behind using context in metaphor detection is that metaphorically used words tend to violate lexical cohesion in text. Different methods, however, approach the problem of detecting semantic outliers in different ways.

Li and Sporleder (2009; 2010) identify metaphorical idioms using the idea that non-literal expressions break lexical cohesion of a text. Li and Sporleder (2009) approached the problem by constructing a *lexical cohesion graph*. In the graph, content words in a text are represented as vertices, which are connected by edges representing semantic relatedness. The intuition behind their approach was that non-literal expressions would lower the average semantic relatedness of the graph. To classify a word as literal or metaphorical, Li and Sporleder (2010) use Gaussian Mixture Models with semantic similarity features, such as the relatedness between this target word and words in its context.

Broadwell et al. (2013) and Strzalkowski et al. (2013) base their approach on the idea that metaphors are likely to be concrete words that are not semantically associated with the surrounding context. Broadwell et al. (2013) implemented this idea using *topic chains*, which consist of noun phrases that are connected by pronominal mention, repetition, synonym, or hyponym relations. Strzalkowski et al. (2013) build on this idea by taking nouns and adjectives around the target concept as candidate source relations. They filtered out candidate sources that were in the same topical chain as the target concept or were not linked to the word being classified by a direct dependency path.

Schulder and Hovy (2014) also hypothesize that novel metaphors are marked by their unusualness in a given context. They use a domain-specific *term relevance* metric, which measures how typical a term is for the domain associated with the literal usage of a word, and *common relevance*, which measures how common a word is across domains. If a term is neither typical for a text's domain nor common, it is taken as a metaphor candidate. A particular strength of this approach is its accommodation of common words without discriminative power, which often confuse context-based models.

Jang et al. (2015) model context by using both *global context*, the context of an entire post, and *local context*, the context within a sentence, in relationship to a word being classified as metaphorical or literal. They used word categories from FrameNet, topic distribution, and lexical chain information (similar in concept to the topic chain information in (Broadwell et al., 2013)) to model the contrast between a word and its global context. To model the contrast between a word and its local context, they used lexical concreteness, word categories and semantic relatedness features.

Mohler et al. (2013) built a domain-aware semantic signature for a text to capture the context surrounding a metaphorical candidate. Unlike other approaches that try to discriminate metaphors from their context, their approach uses binary classifiers to compare the semantic signature for a text with that of known metaphors.

The above approaches attempted to capture governing context in various ways and were effective when applied to the problem of metaphor detection. However, these methods tend to over-classify literal instances as metaphorical when semantic cohesion is violated within their governing contexts. Additionally, these methods could

fail to detect extended metaphors, which span over wider contexts. In this paper, we specifically focus on the problem of discriminating literal instances from metaphorical instances by expanding the scope of what is captured within a context. Like (Mohler et al., 2013), we share the intuition that there could be associations between specific metaphors and their contexts, but we relax the assumption that metaphors must be similar to known metaphors.

3 Our Approach

To better capture the distinctions between metaphorical and literal usages of the same word (target word), we approach the task in two directions. First, we model how topics in context change for both metaphorical and literal instances of a target word (Section 3.1). Second, we consider the situational context for why individuals choose to use metaphor (Section 3.2). We use multi-level modeling to combine these two types of features with the specific target word to model interactions between the features and a particular metaphor (Section 3.3).

3.1 Topic Transition

In writing, cohesion refers to the presence or absence of explicit cues in the text that allow the reader to make connections between ideas (Crossley and McNamara, 2010). For example, overlapping words and concepts between sentences indicate that the same ideas are being referred to across these sentences. Metaphorically used words tend to be semantically incohesive with the governing context. Therefore, determining semantic or topical cohesion is important for metaphor detection.

However, even if a text is literal and cohesive, not all words within the text are semantically related. In example (1), a human could easily determine that "pillows", "music", "flickering candles", and "a foot massage" share the theme of relaxation. But it is difficult to define their relatedness computationally – these terms are not synonyms, hypernyms, antonyms, or in any other well-defined lexical relation. Additionally, even if the whole sentence is correctly interpreted as ways of indulging oneself, it is still semantically contrasted with the surrounding sentences about medicine. In this example, the target word "candle" is used literally, but the contrast between the

sentence containing the target word and its context makes it computationally difficult to determine that it is not metaphorical:

(1) ... yet encouraged to hear you have a diagnosis and it's being treated. Since you have to give up your scented stuff you'll just have to figure out some very creative ways to indulge yourself. *Soft pillows, relaxing music, flickering candles, maybe a foot massage.* Let's hope your new pain relief strategy works and the Neulasta shot is not so bad . I never had Taxotere, but have read it can be much easier than AC for many people. ...

Example (2) also shows semantic inconsistency between the candidate metaphor "boat" and the surrounding sentences about medicine. However, in this example, "boat" is metaphorically used. Thus, it is difficult to determine whether a word is metaphorical or literal when there is semantic contrast because both example (1) and example (2) show semantic contrast.

(2) When my brain mets were discovered last year, I had to see a neurosurgeon. He asked if I understood that my treatment was pallative care. *Boy, did it rock my boat to hear that phrase!* I agree with Fitz, pallative treatment is to help with pain and alleviate symptoms.....but definitely different than hospice care.

The primary difference between these two examples is in the nature of the semantic contrast. In example (1), the topic of the sentence containing "candle" is relaxation, while the topic of the previous and following sentences is medicine. The transition between medicine and relaxation tends to be more literal, whereas the transition between the topic in the sentence containing "boat" and the surrounding medical topic sentences tends to be more metaphorical.

We use these differences in the topic transition for metaphor detection. We consider topic transitions at the sentence level, rather than the word level, because people often represent an idea at or above the sentence level. Thus, topic is better-represented at the sentence level.

To model context at the sentence level, we first assign topics to each sentence using Sentence Latent Dirichlet Allocation (LDA) (Jo and Oh, 2011). Sentence LDA has two main advantages over standard LDA for our work. First, while standard LDA assumes that each word is assigned a topic derived from the topic distribution of a document, Sentence LDA makes the constraint that all words in the same sentence must be assigned the same topic. Due to this property, the generated topics are better aligned with the role or purpose of a sentence, compared to topics generated from LDA. Additionally, having each sentence assigned to one topic helps us avoid using heuristics for representing the topic of each sentence. [1]

Using Sentence LDA, we modeled four features to capture how the topic changes around the sentence where a target word resides. We refer to this sentence as the *target sentence*.

Target Sentence Topic (TargetTopic): We hypothesize that sentences containing a metaphor may prefer topics that are different from those of sentences where the same word is used literally. Hence, TargetTopic is a T-dimensional binary feature, where T is the number of topics, that indicates the topic assigned to the sentence containing the target word.

Topic Difference (TopicDiff): We hypothesize that a metaphorical sentence is more likely to be different from its neighboring sentences, in terms of topic, than a literal sentence. Therefore, TopicDiff is a two-dimensional binary feature that indicates whether the topic assigned to the target sentence is different from that of the previous and next sentences.

Topic Similarity (TopicSim): Under the same hypothesis as TopicDiff, TopicSim is a two-dimensional feature that represents the similarity between the topic of the target sentence and its previous and next sentences. Unlike TopicDiff, which is binary, TopicSim has continuous values between 0 and 1, as we use the cosine similarity between each topic's word distributions as topic similarity. Note that in Sentence LDA, all topics share the same vocabulary, but assign different probabilities to different words as in LDA although all tokens in a sentence are assigned to the

same topic in Sentence LDA.

Topic Transition (TopicTrans): The topic of a metaphorical sentence may extend over multiple sentences, so a topic transition may occur a few sentences ahead or behind the target sentence. TopicTrans looks for the nearest sentences with a different topic before and after the current target sentence and encodes the topics of the different-topic sentences. Hence, TopicTrans is a $2T$-dimensional feature, where T is the number of topics, that indicates the topics of the nearest sentences that have a different topic from the target sentence.

Topic Transition Similarity (TopicTransSim): The topics before and after a transition, even in the extended case for TopicTrans, are still expected to be more different in metaphorical cases than in literal cases, as we assume for TopicSim. Therefore, TopicTransSim is a two-dimensional continuous feature that encodes the cosine similarity between the topic of the target sentence and the topics of the nearest sentences that have a different topic before and after the target sentence.

3.2 Emotion and Cognition

Metaphors are often used to explain or describe abstract ideas, such as difficult concepts or emotions (Meier and Robinson, 2005). (Fainsilber and Ortony, 1987) showed that descriptions of feelings contain more metaphorical language than descriptions of behavior.

In our domain, writers are searching for support through the emotionally tumultuous experience of breast cancer and often turn to metaphor to express this emotion. For example, the word "road" can be used as a metaphor to express the emotional experiences of waiting for or passing through steps in treatment. A similar phenomenon is that markers of cognition, such as "I think", can occur to introduce the abstract source of the metaphor. In example (3), one breast cancer patient in our data describes her speculation about her condition metaphorically, writing,

> (3) i have such a long *road* i just wonder what to do with myself.

To encode these emotional and cognitive elements as features, we use Linguistic Inquiry Word Count (LIWC) (Tausczik and Pennebaker, 2010). LIWC is a tool that counts word use in certain

---

[1] We also tried standard LDA for assigning topics to sentences, by representing each sentence as a topic distribution over its words. However, this representation was not as informative as Sentence LDA in our task, so we leave out the LDA topics in further discussion.

psychologically relevant categories. Focusing on emotional and cognitive processes, we use the LIWC term lists for categories seen in Table 1.

LIWC category	Example Terms
affect	ache, like, sweet
positive emotion	passion, agree, giving
negative emotion	agony, annoy, miss
anxiety	embarrass, avoid
anger	assault, offend
sadness	despair, grim
cognitive mechanisms	if, could
insight	believe, aware
cause	make, pick
discrep	would, hope
tentativeness	anyone, suppose
certainty	never, true

Table 1: Selected LIWC categories.

We count the number of words that fall into each category within either an immediate or global context. For these LIWC features, we take the target sentence and its neighboring sentences as the immediate context and the entire post as the global context for a candidate metaphor instance. The counts for each category in either the immediate or global context are used as features encoded by what degree the immediate or global context expresses the emotional or cognitive category.

We expect words indicative of emotion and cognition to appear more frequently in metaphorical cases. Our preliminary statistical analysis on the development set revealed that this holds true within the target sentence and shows a tendency in the surrounding sentences.

3.3 Multi-Level Modeling

Our topical and emotion and cognition context features are general across target words. However, the specific features that are informative for metaphor identification may depend on the target word. To account for the specificity of target words, we use multi-level modeling (Daume III, 2007). The idea of multi-level modeling is to pair each of our features with every target word while keeping one set of features independent of the target words. There are then multiple copies of each topic transition and emotion/cognition feature, all paired with a different target word. Thus, if there are N target words, our feature space becomes $N + 1$ times larger.

4 Experiments

Our main experimental task is metaphor detection or disambiguation – given a post containing a candidate metaphor word, we aim to determine whether the word is used literally or metaphorically in context.

4.1 Data

We conducted experiments on a dataset of posts from a public breast cancer support group discussion forum, annotated by Jang et al. (2015). We chose to work on this dataset because it features metaphors occurring in naturalistic language.

In this dataset, posts are restricted to those containing one of seven candidate metaphors that appear either metaphorically or literally: "boat", "candle", "light", "ride", "road", "spice", and "train". We split the data randomly into a development set of 800 posts for preliminary analysis and a cross-validation set of 1,870 posts for classification as in (Jang et al., 2015).

4.2 Metrics

We report five evaluation metrics for every model: kappa, F1 score, precision, recall, and accuracy. Kappa, which corrects for agreement by chance, was calculated between predicted results and actual results. Because the dataset is skewed towards metaphorical instances, we rely on the first four measures over accuracy for our evaluation.

4.3 Baselines

We use the following two baselines: the feature set of (Jang et al., 2015) and a context unigram model.

Jang et al. (2015): We use the best configuration of features from Jang et al. (2015), the state-of-the-art model on our dataset, as a baseline. This feature set consists of all of their local context features (word category, semantic relatedness, concreteness), all of their global context features except lexical chaining (word category, global topic distribution), and context unigrams.

Context Unigram Model: All the words in a post, including the target word, are used as context features.

4.4 Settings

We ran Sentence LDA, setting the number of topics to 10, 20, 30, 50, and 100. α and β determine the sparsity of the topic distribution of each document and the word distribution of each topic,

Model	κ	F1	P-L	R-L	P-M	R-M	A
Unigram	.435	.714	.701	.434	.845	.943	.824
Unigram + AllTopic + AllLIWC***	.533	.765	.728	.550	.872	.937	.847
Unigram + MM AllTopic + MM AllLIWC***	.543	.770	.754	.546	.872	.946	.852
J	.575	.786	.758	.587	.882	.943	.859
J + AllTopic + AllLIWC*	.609	.804	.772	.626	.892	.943	.869
J + MM AllTopic**	.619	.809	.784	.630	.893	.947	.873
J + MM AllLIWC	.575	.787	.757	.589	.882	.942	.859
J + MM AllTopic + MM AllLIWC***	**.631**	**.815**	**.792**	**.642**	**.896**	**.948**	**.876**

Table 2: Performance on metaphor identification task. (**Models**) J: Jang et al. (2015), MM - Multilevel Modeling (**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, *: marginally statistically significant ($p < 0.1$), **: statistically significant ($p < 0.05$), ***: highly statistically significant ($p < 0.01$) improvement over corresponding baseline by Student's t-test.

respectively; the lower the sparser. Following convention, we set these parameters to 0.1 and 0.001, respectively, to enforce sparsity. We also removed the 37 most frequent words in the corpus, drawing the threshold at the point where content words and pronouns started to appear in the ranked list. The models with 10 topics performed the best on the development set, with performance degrading as the number of topics increased. We suspect that poorer performance on the models with more topics is due to feature sparsity.

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. For each experiment, we performed 10-fold cross-validation. We also trained the baselines with the same SVM settings.

4.5 Results

The results of our classification experiment are shown in Table 2. We tested our topical and emotion and cognition features in combination with lexical features from our baselines: unigram and Jang et al. (2015).

Adding our topical and emotion/cognition features to the baselines improved performance in predicting metaphor detection. We see that our features combined with the unigram features improved over the Unigram baseline although they do not beat the Jang et al. (2015) baseline. However, when our features are combined with the features from Jang et al. (2015), we see large gains in performance. Additionally, our multi-level modeling significantly improved performance by tak-
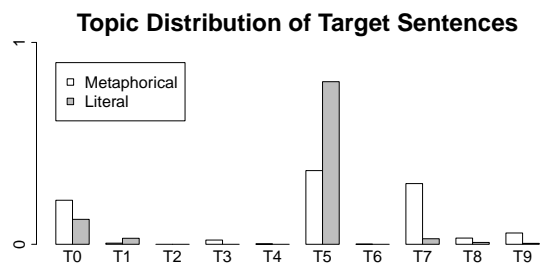

Figure 1: Proportions of topics assigned to target sentences, when target words were used metaphorically vs. literally. The proportions of metaphorical and literal cases are different with statistical significance of $p < 0.01$ by Pearson's chi-square test.

ing into account the effects of specific metaphors. The topical features added to the baseline led to a significant improvement in accuracy, while emotion and cognition features only slightly improved the accuracy without statistical significance. However, the combination of these emotion and cognition features with topical features (in the last row of Table 2) leads to improvement. We performed a Student's t-test for calculating statistical significance.

5 Discussion

Metaphorical instances tend to have personal topics. An author was more likely to use target words metaphorically when the target sentence relates more closely to their own experience of disease and treatment. Specifically, metaphors were relatively frequent when people shared their own disease experience (Topic 0, Topic 9) or sympa-

	Topic	Top Words	Example Sentences
0	Disease/ Treatment	get, chemo, if, they, as, out, can, like, now, she, feel, did, up, know, think, been, good, time, or, when	I'm scared of chemo and ct scans because it makes cancer come back and you become more resistance to treatment with drugs like these later.
1	Food	good, they, gt, can, like, eat, fat, or, if, some, one, as, them, get, up, fiber, think, more, what	*Martha's Way* Stuff a miniature marshmallow in the bottom of a sugar cone to prevent ice cream drips.
2	Emotions	love, great, laura, good, hope, like, debbie, amy, up, happy, too, everyone, day, glad, look, fun, mary, what, kelly, how	Too funny. / You're so cute! / ene23...the photo in the locket idea sounds great!
3	Time	chemo, week, go, last, then, next, weeks, taxol, good, done, treatment, first, start, one, more, rads, after, today, 'll, now	I am now 45, and just had my ONE year anniversary from finishing chemo last week!!
4	Greetings/ Thanks	thanks, hugs, hi, here, carrie, thank, welcome, love, us, glad, know, greg, good, everyone, thread, ladies, there, how, sorry, mags	Thank you so much for the story!! / Big Hugs!
5	People	she, he, they, out, get, up, her, when, like, one, as, from, there, our, time, did, if, can, go, what	She has three children and her twin sister has taken her and her 3 children in.
6	Support	good, hope, well, happy, everyone, doing, glad, luck, hear, better, take, jen, care, great, liz, birthday, hugs, lol, watson, feeling	YAY! / lol. / I wish you all good luck and peace.
7	Relation	what, know, she, as, can, her, cancer, if, there, has, think, been, how, like, our, who, when, they, would, us	She knows that she has BC but does not know that it has spread. / I just read your message and I wondered about you.
8	Religion	god, love, lord, us, prayers, our, bless, dear, her, lu, may, day, patti, thank, know, comfort, amen, xoxo, he, pray	Dear Lord, I come to you with a friend that is not doing well, Please bless her that her hands will reach for you threw the last part of her breast cancer fight.
9	Diagnosis	diagnosed, when, chemo, she, breast, years, stage, cancer, dx, now, found, nodes, no, after, lump, they, age, then, year, mastectomy	I was 64 when diagnosed wtth pure DCIS.....I had my ninght radiation treatment today. / I was diagnosed Nov 2007 at age 45.

Table 3: Topics learned by Sentence LDA.

Topic Distribution of the Sentences Nearest to the Target Sentence and with a Different Topic

Vs. Previous Sentence *Vs. Next Sentence*
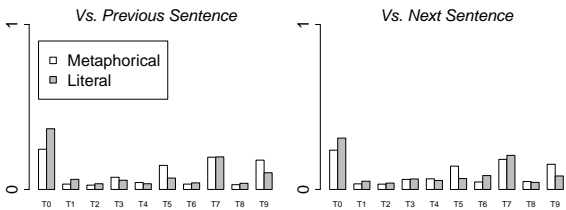
□ Metaphorical
□ Literal

Figure 2: Proportions of the topics of the sentences that are nearest to the target sentence and have a different topic from the target sentence. The proportions of metaphorical and literal cases are different with statistical significance of $p < 0.01$ by Pearson's chi-square test.

Proportions of Target Sentences With A Different Topic from Context

Vs. Previous Sentence *Vs. Next Sentence*

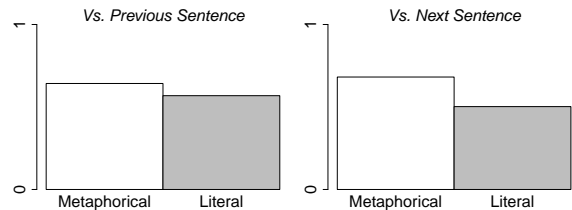Metaphorical Literal   Metaphorical Literal

Figure 3: Proportions of target sentences whose topic is different from that of the previous/next sentence, when target words were used metaphorically vs. literally. The proportions of metaphorical and literal cases are different with statistical significance of $p < 0.01$ by Pearson's chi-square test.

thized with other people's experiences (Topic 7), but were more infrequent when they simply talked about other people in Topic 5 (Figure 1). According to our closer examination of sample sentences, Topic 0 had many personal stories about disease and treatment, and Topic 7 was about learning and relating to other people's experiences. Example metaphorical expressions include "There is *light* during chemo." (Topic 0) and "Hi Dianne - I am glad I found your post as I am sort of in the same

boat." (Topic 7). Analysis of our LIWC features also supports the reflective nature of metaphors: "insight" and "discrepancy" words such as "wish", "seem", and "feel" occur more frequently around metaphorical uses of target terms.

The topics of the surrounding context (TopicTrans) were also informative for metaphor detection (Figure 2). However, the topics of the surrounding sentences followed an opposite pattern to the topics of the target sentence; talking

**Topic Similarity
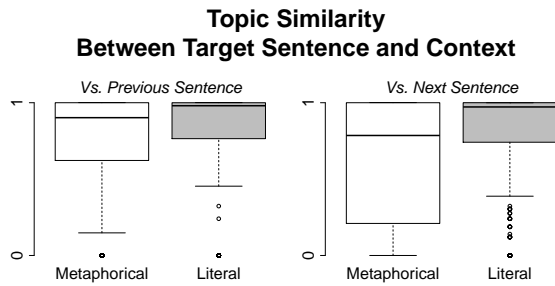Between Target Sentence and Context**

Figure 4: Cosine similarity between the topic of a target sentence and the topic of its previous/next sentence, when target words were used metaphorically vs. literally. The means of the metaphorical and literal cases are different with statistical significance of $p < 0.01$ by Welch's t-test.

**Topic Similarity Between Target Sentence
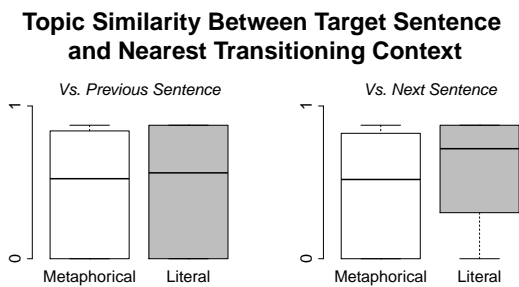and Nearest Transitioning Context**

Figure 5: Cosine similarity of the topic of a target sentence and the topic of the sentences that are nearest to the target sentence and have a different topic from the target sentence. The means of metaphorical and literal cases are different with statistical significance only for the next sentence, with $p < 0.01$ by Welch's t-test.

about other people (Topic 5) in the context of a target sentence led to more metaphorical usage of target words. Similarly, writers used target words more literally before or after they shared their personal stories (Topic 0). This pattern could be because the topic of the target sentence differs from the topics of the surrounding sentences in these instances, which would mean that the target sentence is a topic that is more likely to be literal. Topic 9, however, does not follow the same pattern. One possible reason is that Topic 9 and Topic 0 tend to frequently co-occur and be metaphorical. Thus, if a target word comes after or before Topic 9 and it is Topic 0, then this word may more likely be metaphorical.

Topic transitions are effective indicators of metaphor. Metaphorical instances accompanied more drastic topic transitions than literal instances.

This tendency, which matched our hypothesis, was shown in all our topic features. The immediately neighboring sentences of metaphorical instances were more likely to have a different topic from the target sentence than those of literal instances (Figure 3). Additionally, differences in topic between the target sentence and the neighboring sentences were greater for metaphorical instances (Figure 4). The nearest sentences with topics different from the target sentence (TopicTransSim) also showed this pattern (Figure 5). An interesting finding was that a topic transition after the target sentence was more indicative of metaphor than a transition before.

Emotion and cognitive words are discriminative depending on the metaphor. Emotion and cognition in the surrounding contexts, which were captured by the LIWC features, helped identify metaphors when combined with topical features. This result supports the claim in (Fainsilber and Ortony, 1987) that descriptions of feelings contain more metaphorical language than descriptions of behavior.

This effect, however, was limited to specific target words and emotions. For example, we saw a higher number of anxiety words in the immediate and global contexts of metaphors, but the trend was the opposite for anger words. This may be because our target words, "boat", "candle", "light", "ride", "road", "spice" and "train", relate more to anxiety in metaphors such as "bumpy road" and "rollercoaster ride", than to anger. On the other hand, cognitive words had more consistency, as words marking insight and discrepancy were seen significantly higher around metaphorical uses of the target words. These patterns, nevertheless, could be limited to our domain. It would be interesting to explore other patterns in different domains.

A multi-level model captures word-specific effects. Our features in context helped recognize metaphors in different ways for different target words, captured by the multi-level model. The paucity of general trends across metaphorical terms does not mean a limited applicability of our method, though, as our features do not suppose any specific trends. Rather, our method only assumes the existence of a correlation between metaphors and the theme of their context, and our multi-level model effectively identifies the interaction between metaphorical terms and their con-

texts as useful information.

For all the figures in this section, most target words have a similar pattern. See our supplemental material for graphs by target word.

6 Conclusion

We propose a new, effective method for metaphor detection using (1) sentence level topic transitions between target sentences and surrounding contexts and (2) emotion and cognition words. Both types of features showed significant improvement over the state-of-the-art. In particular, our system made significant gains in solving the problem of over-classification in metaphor detection.

We also find that personal topics are markers of metaphor, as well as certain patterns in topic transition. Additionally, language expressing emotion and cognition relates to metaphor, but in ways specific to particular candidate words. For our breast cancer forum dataset, we find more words related to anxiety around metaphors.

Our proposed features can be expanded to other domains. Though in other domains, the specific topic transition and emotion/cognition patterns would likely be different, these features would still be relevant to metaphor detection.

Acknowledgments

References

George Aaron Broadwell, Umit Boz, Ignacio Cases, Tomek Strzalkowski, Laurie Feldman, Sarah Taylor, Samira Shaikh, Ting Liu, Kit Cho, and Nick Webb. 2013. Using imageability and topic chaining to locate metaphors in linguistic corpora. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 102–110. Springer.

Scott A Crossley and Danielle S McNamara. 2010. Cohesion, coherence, and expert evaluations of writing proficiency. In *Proceedings of the 32nd annual conference of the Cognitive Science Society*, pages 984–989.

Hal Daume III. 2007. Frustratingly easy domain adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 256–263, Prague, Czech Republic, June. Association for Computational Linguistics.

Lynn Fainsilber and Andrew Ortony. 1987. Metaphorical uses of language in the expression of emotions. *Metaphor and Symbolic Activity*, 2(4):239–250.

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.

Yohan Jo and Alice Oh. 2011. Aspect and Sentiment Unification Model for Online Review Analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 815–824.

Beata Beigman Klebanov, Chee Wee Leong, and Michael Flor. 2015. Supervised word-level metaphor detection: Experiments with concreteness and reweighting of examples. *NAACL HLT 2015 3rd Metaphor Workshop*, page 11.

George Lakoff and M. Johnson. 1980. Metaphors we live by. *Chicago/London*.

Linlin Li and Caroline Sporleder. 2010. Using gaussian mixture models to detect figurative language in context. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 297–300, Stroudsburg, PA, USA. Association for Computational Linguistics.

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*, pages 25–28. Association for Computational Linguistics.

Brian P Meier and Michael D Robinson. 2005. The metaphorical representation of affect. *Metaphor and symbol*, 20(4):239–257.

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.

Marc Schulder and Eduard Hovy. 2014. Metaphor detection through term relevance. *ACL 2014*, page 18.

Ekaterina Shutova. 2015. Design and evaluation of metaphor processing systems. *Computational Linguistics*.

Caroline Sporleder and Linlin Li. 2009. Unsupervised recognition of literal and non-literal use of idiomatic expressions. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 754–762. Association for Computational Linguistics.

Tomek Strzalkowski, George Aaron Broadwell, Sarah Taylor, Laurie Feldman, Boris Yamrom, Samira Shaikh, Ting Liu, Kit Cho, Umit Boz, Ignacio Cases, et al. 2013. Robust extraction of metaphors from novel data. *Meta4NLP 2013*, page 67.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer. 2014. Metaphor detection with cross-lingual model transfer.

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.