

Effects of Situational Factors on Metaphor Detection in an Online Discussion Forum

Hyeju Jang, Miaomiao Wen, and Carolyn Penstein Rosé

Language Technologies Institute

Carnegie Mellon University

Pittsburgh, PA 15213, USA

{hyejuj, mwen, cprose}@cs.cmu.edu

Abstract

Accurate metaphor detection remains an open challenge. In this paper, we explore a new type of clue for disambiguating terms that may be used metaphorically or literally in an online medical support community. In particular, we investigate the influence of situational factors on propensity to employ the metaphorical sense of words when they can be used to illustrate the emotion behind the experience of the event. Specifically we consider the experience of stressful illness-related events in a poster's recent history as situational factors. We evaluate the positive impact of automatically extracted cancer events on a metaphor detection task using data from an online cancer forum. We also provide a discussion of specific associations between events and metaphors, such as *journey* with diagnosis or *warrior* with chemotherapy.

1 Introduction

In this paper we present a novel approach to metaphor detection that leverages situational factors in the life of a speaker that alter the propensity to employ the metaphorical sense of specific terms. In recent years, the field of language technologies has made advances in the area of metaphor detection by leveraging some linguistic regularities such as lexical selection, lexical co-occurrence, and abstractness versus concreteness. On the other hand, we know that metaphor is creative at its core, and these linguistic regularities, though essential, are bounded in their ability to enable accurate metaphor detection in a broad sense. In contrast to previous approaches focusing on these linguistically inspired features, we

begin to explore situational factors coming from a pragmatic perspective, related to the reasons why people choose to use metaphors. The situational factors may provide a complementary set of indicators to partner with tried and true linguistically inspired features in order to increase performance. Specifically, we explore expressions of metaphors used in a cancer support community in connection with discussion around stressful cancer events. In particular, we provide evidence that propensity to employ metaphorical language increases around the time of stressful cancer events.

Describing an experience metaphorically is an effective conversational strategy for achieving social goals that are relevant within an online medical support community. For example, a metaphor may be useful for drawing the listener closer by revealing not just what has been experienced, but how the speaker is personally engaged with the event, such as *journey* and *battle* (Jang et al., 2014). For example, the *journey* metaphor conveys the experience of cancer treatment as a process of progressing along a path in which the cancer patient is a traveler, whereas the *battle* metaphor conveys a more active attitude towards cancer treatment by comparing cancer treatment to conflict and war where the speaker is positioned as a warrior. In this way, metaphors may be used to build solidarity or a sense of camaraderie as they increase insight into the speaker's personal experience and thus facilitate empathetic understanding between the participants (Ritchie, 2013).

Beyond the social implications of using a metaphor, there are implications at the cognitive level as well. In particular, metaphor is a type of linguistic tool used to express an abstraction. As

such, usage of metaphor requires somewhat more cognitive effort than the equivalent literal description. Usage of a metaphor may thus reflect the effort the speaker has invested in making sense out of the associated experience.

Both cognitive and social factors may contribute towards an elevated level of usage of specific metaphors that are associated with the experience of a stressful cancer event in the recent past of a speaker. Specifically, speakers experience a need for more social support during and soon after a stressful event, and thus may engage in behaviors that are useful for building closeness and drawing others in. Additionally, as part of the coping process, experiencers of stressful cancer events are faced with the need to adjust to a new reality after the experience, and this adjustment process may be reflected in linguistic mechanisms that are associated with abstraction and reasoning. Leveraging this insight, we hypothesize that for ambiguous terms (those that can be used either in a literal or metaphorical sense), the concentration of metaphorical use will be elevated within a short window of time following the experience of the associated cancer events. We thus hypothesize that a context variable associated with these events will be a useful clue for increasing accuracy at disambiguating the interpretation of these terms.

In this paper, we present a corpus analysis of data extracted from an online medical support community, where technology has been deployed to extract mentions of specific cancer events (e.g. diagnosis, chemotherapy, etc.). First, we investigate how popular metaphors we find to be unambiguous in our data from the discussion forum are used in connection with major cancer events. This validates the proposed association between cancer events and metaphor usage. Second, we evaluate the extent to which event information can be helpful for a computational metaphor disambiguation task over more ambiguous candidate metaphor words. In this work, we quantitatively verify the effectiveness of considering situational features in metaphor detection.

The major contribution of this work from a computational perspective is to introduce novel types of features for automatic metaphor detection. Metaphor is not a purely linguistic phenomenon only, but it is language in use. It can depend on

a variety of factors including the mood, audience, identity of speaker, and the situational context of the speaker. Thus, we believe that combining insights both from linguistics and language in use will be able to benefit metaphor detection. Our hope is that this work opens a door to more diverse kinds of situational features to be used for metaphor detection, together with linguistically inspired features. In addition, our work reinforces and extends earlier insights into social and cognitive factors that influence usage of metaphor in discussion, and illustrates a new impact of accurate event extraction.

The remainder of the paper is organized as follows. Section 2 relates our work to prior work on computational metaphor detection. Section 3 describes the data used for our experiment. Section 4 explains the event extraction method we adopted. Section 5 illustrates popular metaphors related to cancer events in our data through a statistical analysis. Section 6 presents our successful metaphor disambiguation experiments. Section 7 concludes the paper with a discussion of limitations and next steps in the work.

2 Related Work

In this section, we introduce two main bodies of relevant prior work in language technologies: case studies in online medical support communities and computational metaphor detection.

2.1 Case Studies in Online Medical Support Communities

Analysis of language patterns in online cancer forums have shown effects of time and experience. For example, with respect to time, Nguyen and Rosé (2011) examine how language use patterns are linked with increased personal connection with the community over time. They show consistent growth in adoption of community language usage norms over time. Prior work on online cancer support discussion forums also shows that participants' behavior patterns are influenced by the experience of stress-inducing events. For example, Wen and Rosé (2012) show that frequency of participants' posting behavior is correlated with stress-inducing events. Wen et al. (2011) conducted a study to analyze patterns of discussion forum posts relating

to one specific woman's cancer treatment process. However, these studies have not performed computational analysis on the role of metaphor in these tasks. Metaphor use in this domain is highly prevalent, and plays an important role in analysis of language use, however its usage patterns in this type of context have not been systematically explored.

2.2 Computational Metaphor Detection

There has been much work on computational metaphor detection. Among these published works, the approaches used have typically fallen into one of three categories: selectional preferences, abstractness and concreteness, and lexical incoherence.

Selectional preferences relate to how semantically compatible predicates are with particular arguments. For example, the verb *eat* prefers *food* as an object over *chair*. The idea of using selectional preferences for metaphor detection is that metaphorically used words tend to break selectional preferences. In the example of *The clouds sailed across the sky*, *sailed* is determined to be a metaphor since *clouds* as a subject violates its selectional preferences. Selectional preferences have been considered in a variety of studies about metaphor detection (Martin, 1996; Shutova and Teufel, 2010; Shutova et al., 2010; Shutova et al., 2013; Huang, 2014)

The abstractness/concreteness approach associates metaphorical use with the degree of abstractness and concreteness within the components of a phrase. In an phrase of adjective and noun such as *green idea* and *green frog*, the former is considered metaphorical since an abstract word (*idea*) is modified by a concrete word (*green*), while the latter is considered literal since both words are concrete (Turney et al., 2011). Broadwell et al. (2013) use measures of imageability to detect metaphor, a similar concept to abstractness and concreteness.

The lexical coherence approach uses the fact that metaphorically used words are semantically not coherent with context words. Broadwell et al. (2013) use topic chaining to categorize words as non-metaphorical when they have a semantic relationship to the main topic. Sporleder and Li (2009) also use lexical chains and semantic cohesion graphs to detect metaphors.

To the best of our knowledge, there has been no computational work on the effect of situational fac-

tors, such as the experience of stressful events, on computational metaphor detection. Demonstrating how situational factors could be useful for computational metaphor detection is one of our contributions.

3 Data

We conduct experiments using data from discussion boards for an online breast cancer support group. Participants in the discussion forums are mainly patients, family members, and caregivers. People use the discussion for exchanging both informational support and emotional support with each other by sharing their stories, and through questioning and answering. Some people begin participating in this forum immediately after being diagnosed with cancer, while others do not make their first post until a later event in the cancer treatment process, such as chemotherapy (Wen and Rosé, 2012).

The data contains all the public posts, users, and profiles on the discussion boards from October 2001 to January 2011. The dataset consists of 1,562,459 messages and 90,242 registered members. 31,307 users have at least one post, and the average number of posts per user is 24.

We picked this dataset for our study of relationship between metaphor and situational factors for two reasons. First, people in this community have a common set of events (e.g cancer diagnosis, chemotherapy, etc.) that are frequently discussed in user posts. Second, people use metaphorical expressions quite frequently in this domain. Thus, the dataset is suitable for a study about metaphor use related with user events. Below is an example post containing metaphors. Some parts in the post have been changed for private information.

Meghan, I was diagnosed this pst 09/02/07. I was upset for a day when I realized after I had two mammograms and the ultrasound that I had cancer-I didn't have a diagnosis, but I knew. After the ultrasound came the biopsy and then the diagnosis, I was fine. I did research. I made up my mind about what treatment I thought I wanted. I was good...I really was fine up to my visit with the surgeon last week. That made it really real for me.

I am waiting for my breast MRI results, and I have to have an ultrasound needle guided auxillary node biopsy before I even get to schedule my surgery. My PET showed other issues in the breast, thus the MRI and the biopsy. Be kind to yourself. It will be a *roller coaster ride* of emotions. Some days really up and strong, other days needing lots of hugs and kleenex. Melody

4 Extracting Cancer Event Histories

The cancer events investigated in this paper include *Diagnosis*, *Chemotherapy*, *Radiation Therapy*, *Lumpectomy*, *Mastectomy*, *Breast Reconstruction*, *Cancer Recurrence* and *Metastasis*. All these eight events induce significant physical, practical and emotional challenges. The event dates are extracted from the users' posts as well as the "Diagnosis" and "Biography" sections in their user profiles. 33% of members filled in a personal profile providing additional information about themselves and their disease (e.g., age, occupation, cancer stage, diagnosis date).

We apply the approach of Wen et al. (2013) to extract dates of cancer events for each of the users from their posting histories. A temporal tagger retrieves and normalizes dates mentioned informally in social media to actual month and year referents. Building on this, an event date extraction system learns to integrate the likelihood of candidate dates extracted from time-rich sentences with temporal constraints extracted from event-related sentences.

Wen et al. (2013) evaluate their event extraction approach in comparison with the best competing state-of-the-art approach and show that their approach performs significantly better, achieving an 88% F1 (corresponding to 91% precision and 85% recall) at resolution of extracted temporal expressions to actual calendar dates, and correctly identifies 90% of the event dates that are possible given the performance of that temporal extraction step.

We adopt the same method to extract all users' cancer event dates in our corpus. Note that even were we to use a perfect event extraction system, we can only extract events that the users explicitly mention in their posts. Users may experience additional events during their cancer treatment process,

and simply choose not to mention them during their posts.

5 Investigation into the Connection between Metaphor and Events

As users continue to participate in the cancer community we are studying, over time they experience more and more significant cancer events. Earlier work (Wen and Rosé, 2012) shows elevated levels of participation frequency and posting frequency around the time of and immediately after experiencing one of these stress-causing events. This pattern suggests that one way users work to process their traumatic experience is by participating in the forum and obtaining support from other people who are going through similar experiences. Since using metaphorical language suggests elevated levels of cognitive effort related to the associated concept, it is reasonable to expect that users may also engage in a higher concentration of metaphorical language during this time as well as an additional reflection of that processing. In this section, we investigate how the use of metaphor changes with respect to specific traumatic cancer events. We examine a set of common metaphors to see whether situational factors, i.e. cancer events, affect their use. We use cancer event dates extracted in (Wen et al., 2013) as described in Section 4

5.1 Before and After Events

As our first analysis of the relationship between metaphor use and events, we pick eight unambiguous metaphor words in our data – *journey*, *boat*, *warrior*, *angel*, *battle*, *victor*, *one step at a time*, and *roller coaster ride* – and consider the distribution of these metaphors around each event. We categorized these metaphors as unambiguous based on their usage within a small sample of posts we analyzed by hand. Since these are unambiguous, we can be sure that each time we detect these words being used, the speaker is making a metaphor. For each metaphor-event pair, we construct a graph showcasing the frequency of the metaphor usage both before and after the event. We center each user's post dates around the month of the event, so times on the x-axis are relative dates rather than absolute dates (the center of the graph corresponds to the actual event month).

The graphs for *journey* and *warrior* paired with the diagnosis event are shown in Figure 1 and Figure 2, respectively.

Certain metaphor/event pairs show a peak around the event, or at 1 year after the event, for example on the anniversary of diagnosis, which is a significant event in the life of a cancer patient. However, the pattern does not hold across all such pairs, making it difficult to generalize. For example, in Figure 1, we see a peak of metaphor frequency occurring at the time of the event, but in Figure 2, we do not see such a peak at the time of the event, but see other peaks both before and after the event date. Another complicating factor is that different users experience different cancer treatment timelines. For instance, one user might experience these events over a long period of time, whereas another user may encounter these events in quick succession (Wen and Rosé, 2012). These factors motivated us to consider other methods, including hierarchical mixed models, for more in-depth analysis.

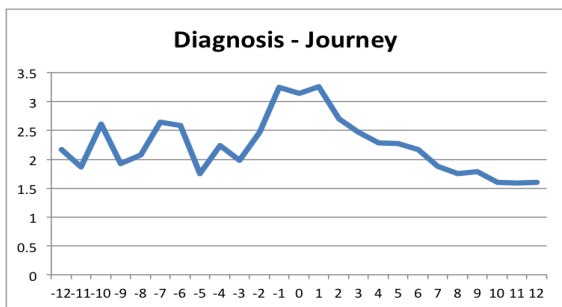


Figure 1: Distribution of *journey* metaphor centered around diagnosis event (x-axis: months from event, y-axis: average frequency of metaphor usage)

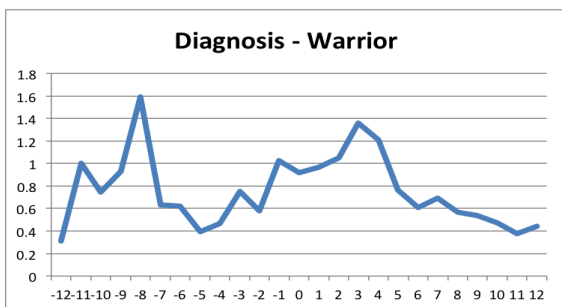


Figure 2: Distribution of *warrior* metaphor centered around diagnosis event (x-axis: months from event, y-axis: average frequency of metaphor usage)

5.2 Associated Events Analysis

Hierarchical mixed models enable us to model the effect of the experience of a cancer event in the history of a user while controlling for other important factors, such as time and personal tendency. We prepared data for analysis by sampling users. We identified the list of users who used any of our target metaphors at least once, and extracted all the posts of those users. In our models, we treat the message as the unit of analysis, and the dependent measure is always either the presence or absence of a specific metaphor, or the presence or absence of metaphorical language more generally, in all cases indicated by a dichotomous variable. Independent variables including dichotomous indicators of the experience of a specific cancer event in the recent past. We treat each user post as being in the critical period of a cancer event if the post date falls within a time window of two months prior to the event month to two months after the event month, which we selected based on informal observation. Data statistics are shown in Table 1.

We tested the association between each dependent variable and the set of independent variables. These hierarchical mixed models were built using the Generalized Linear Latent and Mixed Models (GLLAMM) add-on package in STATA (Rabe-Hesketh and Skrondal, 2008; Rabe-Hesketh et al., 2004), using maximum likelihood estimation to estimate the models. A random intercept is included for each poster, which is necessary for avoiding obtaining biased estimates of the parameters since there were multiple data points for each user, and users varied in their tendency to use metaphorical language or not. We also experimented with time as an independent variable to control for potential consistent increases in usage of metaphorical language over time, but we did not find any such strong effect, and so we dropped this variable from our models.

We did not find significant effects with a dependent measure that indicated that any of the set of metaphors were used, however, we did find significant associations between metaphors and events when we used dependent variables associated with specific metaphors. Our finding was that the subset of events associated with a metaphor varied by metaphor in a way that made sense given the conno-

metaphor	#		%	
	M	L	M	L
journey	5,787	1,329,560	0.43	99.57
boat	21,398	1,313,849	1.60	98.40
warrior	3,462	1,331,785	0.26	99.74
angel	16,025	1,319,222	1.20	98.80
battle	6,347	1,328,900	0.48	99.52
victor	3,540	1,331,707	0.27	99.73
one step at a time	1,554	1,333,693	0.12	99.88
roller coaster ride	536	1,334,711	0.04	99.96
all	64,755	1,270,492	4.85	95.15

Table 1: Corpus-wide unambiguous popular metaphor use statistics (among posts where the user used the metaphor at least once) (**M**: posts that contain each metaphor, **L**: posts that do not contain each metaphor).

candidate	associated events
journey	diagnosis, recurrence, mastectomy
boat	diagnosis, reconstruction
warrior	chemo
angel	chemo, rads, mets
battle	diagnosis, rads, lumpectomy
victor	chemo, rads, reconstruction
one step at a time	diagnosis
roller coaster ride	diagnosis, reconstruction

Table 2: Metaphor candidates and their associated events

tation of the metaphor. For instance, *warrior* is associated with chemo, and *journey* is associated with diagnosis, recurrence, and mastectomy. Associations for all metaphors used for analysis are listed in Table 2.

6 Metaphor Disambiguation

Knowing that there is a significant association between the experience of a cancer event and the usage of a metaphor opens up the possibility for using knowledge of a user’s experience of cancer events in the interpretation of their language choices. In particular, if they use a word that may or may not be metaphorical, and the metaphorical usage is associated with a cancer event that occurred in their recent past, then the model should be more likely to pre-

dict the metaphorical interpretation. Conversely, if the user is not within the critical period of the event associated with the potential metaphorical interpretation, the metaphorical interpretation should be correspondingly less preferred. We hypothesize that usage of this contextual information might improve the accuracy of disambiguation of potentially metaphorical language. In this section, we test that hypothesis in a corpus based experiment conducted this time on a set of ambiguous, potentially metaphorical words.

6.1 Task

Our task is metaphor disambiguation: given a candidate word, decide whether the word is used metaphorically or literally in a post. For example, *road* in (1) is used metaphorically, and *road* in (2) is used literally. The task is to classify *road* into metaphor and literal use.

- (1) Great hobbies! ... My hobby that I love is *road* bike riding. My husband and I both have bikes and we love to ride. ... That’s the beauty of living in the south is that you can ride all year long.
- (2) Another thing to consider is cosmetic outcome. ... If you have a recurrence of cancer and have to do a mast down the *road*, reconstruction is more difficult after having radiation. ...

6.2 Data Annotation

We picked six metaphor candidates that appear either metaphorically or literally in the breastcancer corpus: *candle*, *light*, *ride*, *road*, *spice*, and *train*.

We employed MTurk workers to annotate metaphor use for candidate words. A candidate word was given highlighted in the full post it came from. MTurkers were instructed to copy and paste the sentence where a given highlighted word is contained to a given text box to make sure that MTurkers do not give a random answer. They were given a simple definition of metaphor from Wikipedia along with a few examples to guide them. Then, they were questioned whether the highlighted word is used metaphorically or literally. Each candidate word was labeled by five different MTurk workers, and we paid \$0.03 for annotating each word. To control annotation quality, we required that all workers have a United States location and have 98% or more of their previous submissions accepted. We filtered out annotations whose the first task of copy and paste failed, and 18 out of 11,675 annotations were excluded.

To evaluate the reliability of the annotations by MTurkers, we calculated Fleiss’s kappa (Fleiss, 1971). Fleiss’s kappa is appropriate for assessing inter-reliability when different items are rated by different judges. The annotation was 1 if the MTurker coded a word as a metaphorical use, otherwise the annotation was 0. The kappa value is 0.80.

We split the data randomly into two subsets, one for analysis of related events, and the other for classification. The former set contains 803 instances, and the latter contains 1,532 instances. The unusual number of instances within each subset arises from the fact that some posts contain multiple metaphors, and we specifically chose to set aside 1,500 posts for classification.

6.3 Analysis on Associated Events

We performed a statistical analysis on the six metaphor candidate words as in Section 5.2. We combined the users from all the six metaphor candidates, and extracted posts of these users. Independent variables for the model were binary values for each event, where the value is 1 if a post was written in the critical period (defined previously in Sec-

candidate	#		%	
	N	L	N	L
candle*	4	18	18.18	81.81
light	503	179	73.75	26.25
ride	234	185	55.85	44.15
road	924	129	87.75	12.25
spice*	3	21	12.50	87.50
train	94	41	69.63	30.37
all	1762	573	75.46	24.54

Table 3: Metaphor use statistics of data used for MTurk (* indicates metaphor candidates for which the literal usage is more common than the non-literal one, N: nonliteral use L: literal use).

candidate	associated events
candle	none
light	diagnosis, rads, mast
ride	diagnosis
road	diagnosis, rads
spice	none
train	mast

Table 4: Metaphor candidates and their associated events

tion 5.2), and 0 otherwise. The dependent variable is a binary value regarding the usage of a metaphor candidate within a post. If a particular post does not include a metaphor candidate or if a post includes a literally used metaphor candidate, the binary dependent value is set to 0. Otherwise, it is set to 1.

The results of conducting the hierarchical mixed model analysis on the data similar to the one conducted above on non-ambiguous metaphors suggest that some candidate words show an association with different cancer events as shown in Table 4.

6.4 Classification

We used the LightSIDE (Mayfield and Penstein-Rosé, 2010) toolkit for extracting features and classification. For the machine learning algorithm, we used the support vector machine (SVM) classifier provided in LightSIDE with the default options. We used basic unigram features extracted by LightSIDE.

To see the effect of event information for classification, we defined two sets of event features. One is a feature vector over all the events, consisting of

model	Accuracy	Kappa	Precision	Recall	F1 score
(1) word	0.8133	0.3493	0.8105	0.9827	0.8884
(2) context unigram	0.8094	0.4701	0.8651	0.8860	0.8754
(3) context unigram + event	0.8127	0.4777	0.8657	0.8903	0.8778
(4) context unigram + associated event	0.8146	0.4729	0.8612	0.8998	0.8801
(5) context unigram + fs	0.8277	0.5155	0.8731	0.9033	0.8879
(6) context unigram + event + fs	0.8336	0.5325	0.8772	0.9067	0.8917
(7) context unigram + associated event + fs	0.8244	0.504	0.8695	0.9033	0.8861

Table 5: Performance on metaphor disambiguation evaluation. (6) is significantly better than (5) [p=0.013] (fs.: used feature selection)

both binary variables to indicate whether or not a post belongs to the critical period of each event, and numerical variables to indicate how many months the post is written from a known event. We will refer to these features as *event* in Table 5. The other is a binary variable to indicate whether or not a post belongs to the critical period of *any* of the associated events for the given metaphor (defined in Section 6.3). We will refer to this feature as *associated event* in Table 5.

We used multilevel modeling for the features when including *associated event*. We also used the FeatureSelection feature in LightSIDE, where a subset of features is picked on each fold before passing it to the learning plugin. We performed 10-fold cross validation for these experiments.

Because we want to see the effect of event information, we compare our model with a unigram model that uses only the word itself as in (Klebanov et al., 2014), and the context unigram model which uses all the context words in a post as features as baselines.

6.5 Results

Table 5 displays the results for our experiments. First, we observe the strong performance of the unigram baseline. As in (Klebanov et al., 2014), our evaluation also shows that just using the word currently being classified gives relatively high performance. This result suggests that our candidate words are popular metaphors repeatedly used metaphorically in this domain, as precision is above 80%.

Second, surprisingly, we do not see improvement on accuracy from adding the context words as features. However, we do observe that this addition

results in a higher kappa value than just using the candidate words themselves.

Finally, we can see both *event* and *associated event* features show promising results. Both additions give higher result when added to the context unigram model, and the *event* features continue to show improvement when considering models with feature selection. The best model, using *event* features with feature selection, shows significant improvement ($p < 0.05$) over the next best model of context unigram with feature selection.

7 Conclusion

In this paper, we discussed how situational factors affect people’s metaphor use. We presented a study in an online medical support community, which contains a variety of related events (e.g. diagnosis, chemotherapy, etc.). First, we investigated how popular unambiguous metaphors in the discussion forum are used in relation to major cancer events. Second, we demonstrated that event information can be helpful for a computational metaphor disambiguation task over ambiguous candidate metaphor words. In this work we quantitatively verified the effect of situational features.

Our analysis showed that some popular unambiguous metaphors in the discussion forum are used in connection with stressful cancer events. Usage of different metaphors is associated with different cancer events. We also observed that the personal tendency factor is about 10 times as strong as the situational factor. For our future work, it will be an interesting problem to design a model considering the personal tendency factor. It will require a latent variable model to properly tease these factors apart.

In addition, our metaphor disambiguation experiments validated the proposed association between cancer events and metaphor usage. Using event information as features showed significant improvement. Although our classification results using associated event information show weak improvement to no improvement depending on whether feature selection is used, it is important to note that our analysis consistently identified a strong relationship between metaphors and their associated events (Table 4). Therefore, we believe that it is crucial to develop a classification model that can better leverage the metaphor-event association, which remains as our future work. We also want to try different sized context windows for the critical period of a cancer event in order to see the effect of time with respect to situational factors.

One limitation of this research is that our analysis relies on the event extraction results. Although the event extraction approach we adopted is currently the best performing state-of-the-art technique, it still makes mistakes that could make our analysis inaccurate. Another limitation is that it is hard to obtain data big enough to split the data into subparts for both the hierarchical mixed model analysis and classification.

Acknowledgments

This research was supported in part by NSF Grant IIS-1302522.

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.

Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378–382.

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

Hyeju Jang, Mario Piergallini, Miaomiao Wen, and Carolyn Penstein Rosé. 2014. Conversational metaphors

in use: Exploring the contrast between technical and everyday notions of metaphor. *ACL 2014*, page 1.

Beata Beigman Klebanov, Chee Wee Leong, Michael Heilman, and Michael Flor. 2014. Different texts, same metaphors: Unigrams and beyond. *ACL 2014*, page 11.

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

Elijah Mayfield and Carolyn Penstein-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.

Dong Nguyen and Carolyn P Rosé. 2011. Language use as a reflection of socialization in online communities. In *Proceedings of the Workshop on Languages in Social Media*, pages 76–85. Association for Computational Linguistics.

Sophia Rabe-Hesketh and Anders Skrondal. 2008. *Multilevel and longitudinal modeling using Stata*. STATA press.

Sophia Rabe-Hesketh, Anders Skrondal, and Andrew Pickles. 2004. *Gllamm manual*.

SL. David Ritchie. 2013. *Metaphor (Key Topics in Semantics and Pragmatics)*. Cambridge university press.

Ekaterina Shutova and Simone Teufel. 2010. Metaphor corpus annotated for source-target domain mappings. In *LREC*.

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*, pages 1002–1010. Association for Computational Linguistics.

Ekaterina Shutova, Simone Teufel, and Anna Korhonen. 2013. Statistical metaphor processing. *Computational Linguistics*, 39(2):301–353.

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.

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.

Miaomiao Wen and Carolyn Penstein Rosé. 2012. Understanding participant behavior trajectories in online health support groups using automatic extraction

- methods. In *Proceedings of the 17th ACM international conference on Supporting group work*, pages 179–188. ACM.
- Kuang-Yi Wen, Fiona McTavish, Gary Kreps, Meg Wise, and David Gustafson. 2011. From diagnosis to death: A case study of coping with breast cancer as seen through online discussion group messages. *Journal of Computer-Mediated Communication*, 16(2):331–361.
- Miaomiao Wen, Zeyu Zheng, Hyeju Jang, Guang Xiang, and Carolyn Penstein Rosé. 2013. Extracting events with informal temporal references in personal histories in online communities. In *ACL (2)*, pages 836–842.