Identifying Simple Narrative Structure in Personal Narratives

Reid Swanson, Elahe Rahimtoroghi, Thomas Corcoran and Marilyn A. Walker
Natural Language and Dialog Systems Lab
University of California Santa Cruz
Santa Cruz, CA 95064, USA
{reid,elahe,maw}@soe.ucsc.edu, tcorcora@ucsc.edu

1 Introduction

Unlike any prior time in human history, personal narratives about many different life experiences are being told online, and are widely available in social media sources such as weblogs. See Figure 1. These narratives provide a valuable resource for learning knowledge about the world that would be otherwise difficult to obtain. They are particularly well suited to learning about causal and temporal relationships because their interpretation explicitly depends on the coherence of these relationships (Trabasso and van den Broek, 1985). They also provide us with a wealth of commonsense knowledge about people, the types of activities they engage in and the attitudes they hold.

Recently, there has been a growing interest in computational methods for modeling narrative based on the causal and temporal relationships from text (Schank and Ableson, 1977; Manshadi et al., 2008; Beamer and Girju, 2009; Elson and Mckeown, 2009; Chambers and Jurafsky, 2009; Zhichao Hu et al., 2013). However, stories are not just about the events and actions that occur. Instead, they are generally situated within a particular time, setting and social group. They also include emotional reactions to those events and their outcomes, either stated or implied (Goyal et al., 2010) To our knowledge, none of the current research in this area accounts for these aspects of narrative. We show below that 2/3 of the clauses in personal narratives are not action clauses. This implies that any automated system for causal reasoning about events that does not take these distinctions into account will preform poorly.

In this paper we will present a simple labeling scheme, derived from Labov & Waletzkey's (L&W) theory of oral narrative (Labov, 1997), which are useful for separating clauses that indicate causal relationships from those that provide setting information or emotional reactions. We then evaluate the performance of automatically classifying these labels using supervised machine learning techniques.

2 L&W's Theory of Narrative

The elements of personal narratives described above are the building blocks for structured narrative discourse, and are partially identified by L&W categories as illustrated in Figure 1.

L&W define a story as a sequence of ACTION clauses (events), of which at least two must be temporally joined, e.g. clauses 1-5 and 7,8 in Figure 1. Stories also contain ORIENTATIONS (setting the scene, describing the characters), e.g. utterances 1-2. An orientation clause introduces the time and place of the events of the story, and identifies the participants of the story and their initial behavior. To properly understand narrative structure, orientations need to be identified as a separate type of utterance. EVALUATION clauses provide information on the consequences of the narrated events as they pertain to the plans, goals and desires of the participants and give the reason for telling the story. Without EVALUATION there is no story, merely a boring recitation of events. L&W state that the EVALUATION clauses may also describe the events that did not occur, may have occurred, or would occur in the future in the story.

3 Data Collection & Annotation

For our dataset, we sampled 50 personal stories across a variety of topics from a corpus of over 1 million stories automatically identified by Swanson & Gordon (Gordon and Swanson, 2009).

#	Category	Story Clause
1	Orientation	This hasn't been helped by the intense pressure in town due to the political transition coming to an end.
2	Orientation	This week things started alright and on schedule.
3	Action	But I managed to get myself arrested by the traffic police (rouleage) early last Wednesday.
4	Action	After yelling excessively at their outright corrupted methods
5	Action	and asking incessently for what law I actually broke,
6	Action	they managed to bring me in at the police HQ.
7	Action	I was drawing too much of a curious crowd for the authorities.
8	Action	In about half an hour at police HQ I had charmed every one around.
9	Action	I had prepared my "gift" as they wished.
10	Evaluation	Decision witheld, they decided that I neednt to bother,
11	Evaluation	they liked me too much.
12	Evaluation	I should go free.
13	Action	I even managed to meet famous Raus, the big chief.
14	Evaluation	He was too happy to let me go when he realized I was no one.
15	Action	But then, a Major at his side noticed my Visa was expired.
16	Evaluation	Damn!
17	Orientation	My current Visa is being renewed in my other passport at Immigration's.
18	Evaluation	Fuck.

Table 1: An excerpt from an example story from our corpus annotated with the L&W categories.

These stories are spontaneous recountings of everyday events in the life of the author published on their weblog. In this work we treated the basic unit of discourse as an independent clause. This resulted in a collection of 1,652 manually segmented clauses. We split the corpus of 50 narratives into 4 groups and annotated them in a round robin fashion among three annotators to refine our guidelines and process. We also developed a hierarchical annotation scheme that extends the 3 base L&W categories where we found ambiguities and gaps in the theory when applied to our dataset. We were able to obtain the highest level of agreement (Fleiss' κ 0.63) by annotating each clause using the most specific set of labels and then mapping these back to the original L&W categories. Once completed, there were 421 orientation, 436 action, 719 evaluation and 26 not story clauses. Note, over 2/3 of the clauses would be incorrectly treated as ACTION in previous work on causal reasoning.

4 Machine Learning Experiments

Rahimtoroghi et al. (Rahimtoroghi et al., 2013) first demonstrated the viability of automatically labeling narrative text using L&W categories on a corpus of 20 Aesop's fables. Precision and recall on these simple narratives is generally above 0.9 with the exception of ORIENTATION, which had a recall of 0.45. The high accuracy is in part due to the formulaic nature of these fables. In contrast, the personal narratives are spontaneous texts generated by ordinary people with no guarantee of any formal writing experience. Here, we evaluate the performance on noisy user generated narratives.

We conducted a series of machine learning experiments with our annotated data. We randomly split the corpus of 50 narratives into a training and test set consisting of 40 and 10 narratives. Using the L&W definitions, and an analysis of our training data, we derived a number of features that we automatically extracted from the text. These included the relative position of the clause, the part-of-speech of the main verb, if the clause contained a negation, lexical semantic categories from LIWC, dependency relations, lexical unigrams, and several other linguistic features. We used information gain to find the highest value features.

We tried several standard classification algorithms (Naive Bayes (Witten and Frank, 2005), Confidence Weighted Linear Classifier (Dredze et al., 2008), Maximum Entropy (Witten and Frank, 2005)) and a sequential classifier (Lafferty et al., 2001). To evaluate the performance of each algorithm we performed a 10-fold cross-validation on the training data. We also selected the optimal number of features using a five segment line search in the range of 2^2 to 2^{12} for each algorithm. In general, Naive Bayes achieves the highest F-Score among the algorithms, although the results are not significantly better than the other models except for the CRF (p-value < 0.05). We are able to achieve an average F-score of 0.683 on the extended label set using a 10-fold validation of the training data. We also see a large increase in performance when mapping the labels to the reduced label set, which increases the F-Score to 0.762.

References

Brandon Beamer and Roxana Girju. 2009. Using a bigram event model to predict causal potential. In *Proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing*, CICLing '09, page 430441, Berlin, Heidelberg. Springer-Verlag.

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*, ACL '09, page 602610, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mark Dredze, Koby Crammer, and Fernando Pereira. 2008. Confidence-weighted linear classification. In *Proceedings of the 25th international conference on Machine learning*, pages 264–271, Helsinki, Finland. ACM.

David K. Elson and Kathleen R. Mckeown. 2009. A tool for deep semantic encoding of narrative texts. In *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*, Suntec, Singapore, August.

Andrew S. Gordon and Reid Swanson. 2009. Identifying personal stories in millions of weblog entries. In *Third International Conference on Weblogs and Social Media, Data Challenge Workshop*, San Jose, CA, May.

Amit Goyal, Ellen Riloff, Hal Daume III, and Nathan Gilbert. 2010. Toward plot units: Automatic affect state analysis. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, page 1725, Los Angeles, CA, June. Association for Computational Linguistics.

W Labov. 1997. Some further steps in narrative analysis. *Journal of Narrative and Life History*, 7(1-4):395–415.

John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning*, pages 289, 282. Morgan Kaufmann Publishers Inc.

Mehdi Manshadi, Reid Swanson, and Andrew S. Gordon. 2008. Learning a probabilistic model of event sequences from internet weblog stories. In *Twenty-first International Conference of the Florida AI Society, Applied Natural Language Processing track*, Florida.

Elahe Rahimtoroghi, Reid Swanson, and Marilyn Walker. 2013. Evaluation, orientation, and action in interactive StoryTelling. In *Workshop on Intelligent Narrative Technologies 6*, Boston, MA, October.

Roger C Schank and Robert P Ableson. 1977. *Scripts, plans, goals, and understanding: an inquiry into human knowledge structures*. L. Erlbaum Associates ; distributed by the Halsted Press Division of John Wiley and Sons, Hillsdale, N.J.; New York.

Tom Trabasso and Paul van den Broek. 1985. Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24(5):612–630, October.

Ian H. Witten and Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*. Morgan Kaufmann, 2 edition, June.

Zhichao Hu, Elahe Rahimtoroghi, Larissa Munishkina, Reid Swanson, and Marilyn Walker. 2013. Unsupervised induction of contingent event pairs from film scenes. In *Conference on Empirical Methods in Natural Language Processing*, Seattle, WA, October.