Extracting and Aggregating Aspect-Level Sentiment from Product Reviews

Matthew Long, Desmond C. Ong, Shane Soh
Stanford University
{mlong14, dco, shanesoh}@stanford.edu

Abstract

Previous work in aspect-level sentiment analysis—identifying the sentiment associated with products and their individual attributes, like battery life—have mostly been formulated as supervised learning problems, requiring known labels of both the relevant aspects and their sentiment. Here we propose a hybrid method where we first generate aspects from natural language text via unsupervised clustering of word vector representations, and secondly extract aspect-level sentiment. We further propose a deep learning architecture for aggregating and summarizing these aspect-level sentiment within and across reviews.

1 Introduction

In today's e-commercialized society, the consumer not only has access to online stores through which she might purchase anything she desires, but she also has unprecedented access to a deluge of information—most notably, product reviews written by other consumers—with which she can make her decision. Unfortunately, sifting through hundreds of reviews across tens of different websites to acquire specific information about the product and its important attributes (e.g. *battery life* for electronics) is a time-consuming chore. Because of this, there has been much recent research tackling the two separate but connected components of this problem: (1) entity-level or aspect-specific sentiment analysis, and (2) summarization and aggregation of reviews.

Within the context of a product review, the first component, *aspect-specific sentiment analysis*, involves identifying individual "aspects" (which could be the product itself, or attributes of the product), and subsequently identifying the sentiment—positive, neutral, or negative judgments—associated with those aspects [1-5]. Previous methods have relied on graphical models (e.g. Latent Dirchlet Analysis [1-3], Conditional Random Fields [4]), or directly modeling sentiment compositionality [5]. In this paper, we propose extending recent successful advances in deep learning [6-7] to address this problem. In particular, we seek to improve and extend the work in [7], who propose several deep learning architectures like the Recursive Neural Tensor Network (RNTN) to extract aspect and sentiment in a single step. Most of the previous work [1-2,4-7], except [3], are supervised methods that require labeled aspects/sentiment for training. We propose using unsupervised methods to generate candidate aspects via word vector representations, followed by simultaneous extraction of aspect and sentiment.

The second problem involves aggregation of reviews, or constructing a *meta-review*. Previous work [8-10] has shown that simple averaging of the "stars" of reviews for an individual product is suboptimal. Here we propose a deep learning architecture to aggregate the previously identified aspect-specific sentiment across multiple sentences within a review, and furthermore, to aggregate these sentiment across multiple reviews.

2 Problem Statement

Describe your problem precisely specifying the dataset to be used, expected results and evaluation

Amazon Reviews of Electronics [11]

3 Technical Approach and Models

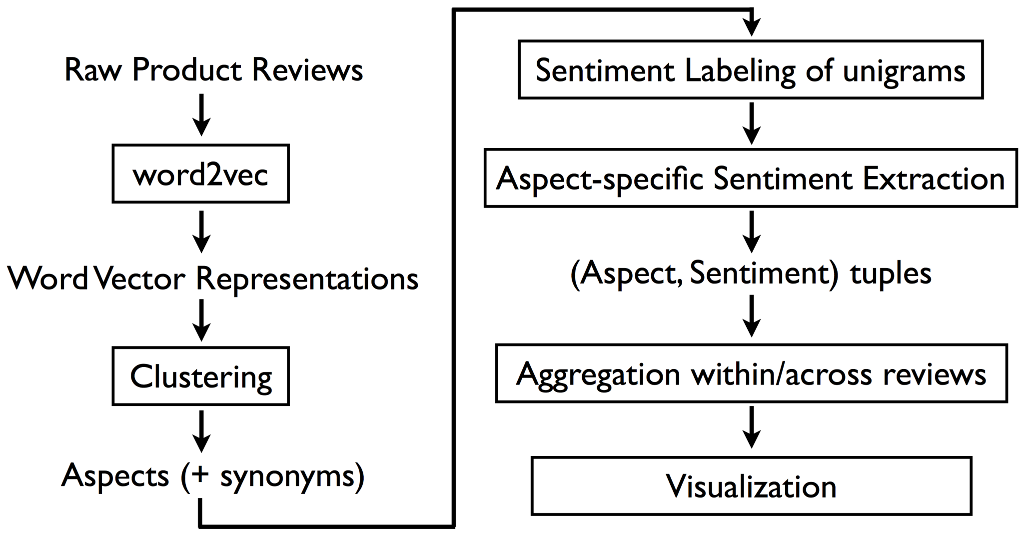

Figure 1: Proposed Workflow. Boxed items are processes, while non-boxed items are outputs. The left half of the diagram details unsupervised aspect identification, while the right half of the diagram involves aspect-specific sentiment extraction, summarization and visualization. Except for the Sentiment Labeling and Visualization processes, every other process involves deep learning.

Our proposed workflow can be divided into three main parts: **Aspect Identification**, **Aspect-Specific Sentiment Extraction**, and **Sentiment Aggregation**.

3.1 Aspect Identification

The first part involves an unsupervised approach to generating aspects. For example, for a given electronic item such as a mobile phone, relevant aspects could be the battery life, screen size, weight, cost, and more. Not all of these aspects are relevant across all electronic items: for a computer, weight might be less of an issue, while screen size might be moot for a screen-less music player. First, we would run word2vec to obtain a word vector representation of the data. Then, we would perform k-means clustering (or some other clustering algorithm) on the word2vec representations. This would generate clusters of similar words, such as: {build construction, build quality, durability, etc}. We can then define the weighted centroid of these clusters as **Aspects**, and the other words in the cluster as relevant synonyms. This allows us to then label all the aspect-words in the dataset.

3.2 Aspect-Specific Sentiment Extraction

An additional pre-processing step would be to label all unigrams with a trained sentiment analysis model. Once this is done, and with the aspect-labels from the previous part, we now have a dataset labeled with aspect and (unigram-) sentiment.

Next, we build off and improve on the model in [7] to use a RNTN or improved architecture to extract (aspect, aspect-specific sentiment) tuples from the data.

3.3 Aspect-Specific Sentiment Aggregation

The final part of our project involves sentiment aggregation. This involves aggregating aspect-specific sentiment across sentences within a review, and at a further step, aggregating across reviews. We are still currently reading relevant literature (e.g. [8-10]) to come up with ideas on how to optimally construct this. A non-deep learning approach might be to hand-specify features (review length; timing – more recent reviews might be more important; etc). A proposal would be to have a deep learning network (such as a recursive neural network with individual reviews as tokens) to try to come up with an optimal way of aggregating this information.

An application-related goal would be to have visualization of our output, in terms of a word cloud of important aspects (e.g., sized by word frequency and colored by sentiment), or a "bar chart" that shows the relevant aspects and their associated sentiment.

4 Intermediate/Preliminary Experiments & Results

State and evaluate your results up to the milestone

Acknowledgments

We would like to acknowledge funding for computing resources provided by the Deep Social Learning Lab at Stanford.

References

[1] Titov, I., & McDonald, R. T. (2008). A Joint Model of Text and Aspect Ratings for Sentiment Summarization. In *ACL* (Vol. 8, pp. 308-316).

[2] Jo, Y., & Oh, A. H. (2011). Aspect and sentiment unification model for online review analysis. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 815-824). ACM.

[3] Brody, S., & Elhadad, N. (2010). An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 804-812). Association for Computational Linguistics.

[4] Engonopoulos, N., Lazaridou, A., Paliouras, G., & Chandrinos, K. (2011). ELS: a word-level method for entity-level sentiment analysis. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*

[5] Moilanen, K., & Pulman, S. (2009). Multi-entity Sentiment Scoring. In *Recent Advances in NLP* (pp. 258-263).

[6] Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on Empirical Methods in Natural Language Processing (EMNLP)* (Vol. 1631, p. 1642).

[7] Lakkaraju, H., Socher, R, & Manning, C. (2014). Aspect Specific Sentiment Analysis using Hierarchical Deep Learning. *NIPS Workshop on Deep Learning and Representation Learning*

[8] Ghose, A., & Ipeirotis, P. G. (2007). Designing novel review ranking systems: predicting the usefulness and impact of reviews. In *Proceedings of the Ninth international conference on Electronic commerce* (pp. 303-310). ACM.

[9] Chen, P. Y., Dhanasobhon, S., & Smith, M. D. (2008). All reviews are not created equal: The disaggregate impact of reviews and reviewers at Amazon.com. Available at SSRN: http://ssrn.com/abstract=918083

[10] Dai, W., Jin, G. Z., Lee, J., & Luca, M. (2012). Optimal aggregation of consumer ratings: an application to Yelp.com (No. w18567). National Bureau of Economic Research.

[11] McAuley, J., Targett, C., Shi, J., & van den Hengel, A. (2015). Image-based recommendations on styles and substitutes. *ACM Special Interest Group on Information Retrieval (SIGIR)*