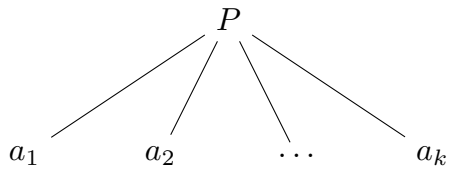
Hierarchical Aspect-Sentiment Analysis for Online Reviews (using Deep Learning Techniques?)

Arzoo Katiyar AK979@CORNELL.EDU
Cornell University

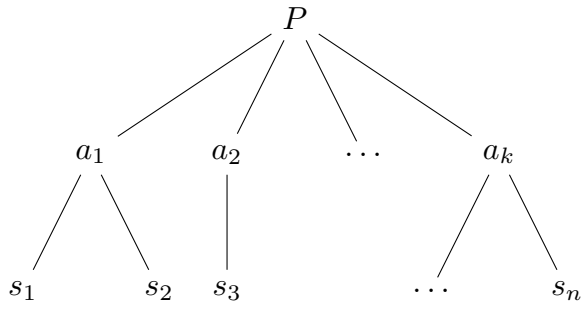
1. Progress made

1.1. Proposed Model

Let R denote a review for a product P and let $\{s_1, s_2, \ldots s_n\}$ are the sentences in a review such that every sentence has an associated aspect. For simplicity, we assume that there is a single aspect associated with every sentence. Let $\{a_1, a_2 \ldots a_k\}$ be the aspects that can be arranged in a hierarchy. Again for the sake of simplicity, we consider a very simple hierarchy as shown below.



Since every sentence has an aspect, we can modify the above diagram to include the sentences as below.



Trivially we know that a review need not contain sentences corresponding to every aspect for that product. Also *we assume that we know a gold standard hierarchy for the product* we consider.
In order to use the model described in (Socher et al., 2013), we want to find vector representation of the sentences at the leaf level in the hierarchy.

1.2. Sentence Model

We want to compute a vector for a sentence s given the sequence of words in s and a vector for each of the words. We propose to use hierarchical convolutional neural network similar to described in (**?**). At the initial stage of composition, the value of a feature in the sentence vector is a function of the values of the same feature in the word vectors. In the next stage, we take compositionality to yield local effects and then yield increasingly more global effects across all the words in the sentence by adopting convolutional kernels of gradually increasing sizes that ultimately span the whole sentence. Figure 1 shows the hierarchical convolutional neural network for sentential compositionality where the bottom layer represents a single feature across all the word vectors in the sentence and the top layer is the value of corresponding feature in the resulting sentence vector.

User generated online reviews contain rich information related to the various aspects of a product. A recent study (Chen & Xie) shows that the customers seek opinions from other users by reading their reviews for the product. Due to the recent popularity of these sites, large number of reviews for different products are available thus leading to the need to extract and evaluate opinions and present them to the user.

Most of the previous work deal with finding the overall sentiment of a review or recently on multi-aspect sentence labeling where each sentence in a review is labeled according to the aspect it discusses along-with finding aspect-specific star ratings for each review. An example review from OpenTable.com as provided in the paper (Lu et al.) to illustrate the aspects of a review for a restaurant is :

" The food was very good, but it took over half an hour to be seated, ... and the service was terrible. The room was very noisy and cold wind blew in from a curtain next to our table. Desserts were very good, but because of [the] poor service, I'm not sure we'll ever go back! "

cnn.jpg

Figure 1. Figure used from (Kim et al., 2013) to show the part of tree structure for the LAPTOP review corpus.

The above review expresses positive opinion toward the restaurant's food but negative opinion towards the ambiance and the service. The goal was to label each sentence in the review with the aspect it expresses (food, ambiance, service) and also find the sentiment for every aspect. A more recent work (Kim et al., 2013) introduces a hierarchical structure in sentiment analysis for reviews. If we consider the example in Figure 1 from the paper, it describes a tree structure over the different aspects of the reviews. Each node in the tree is itself a two-level tree, where the root represent the aspect and the children represent the associated sentiment polarities. Different users are interested in different levels of granularity of aspects and sentiments. Some users may be interested in general aspects such as screen or CPUs in a laptop, while other users might be interested in more specific aspects such as CPU frequency or cache size. Thus it is desirable to have a hierarchy of the different aspects(at each granular level). The paper infer the aspect-sentiment tree from the review texts.

Another independent work on sentiment analysis (Socher et al., 2013) introduces a Recursive Neural Tensor Network(RNTN) for sentiment composition. Semantic word spaces for single words have lately been used by many researchers. Since these vector spaces do not capture the meaning of the longer phrases properly, compositionality in semantic vector spaces has received a lot of attention by the researchers. This paper uses fine-grained sentiment labels for phrases in the parse tree of the sentences and use RNTN for sentiment composition for the positive/negative sentiment classification in a sentence.

The general idea of this project is to find out if the deep learning techniques introduced above can be used to find the overall sentiment of a document using some similar compositionality techniques over the sentences (or phrases?). However, we restrict ourselves to the online user-generated product reviews where there is a possibility of finding a structure in the document and our goal is to be able to use this hierarchy similar to the parse tree in the case of recursive neural networks.

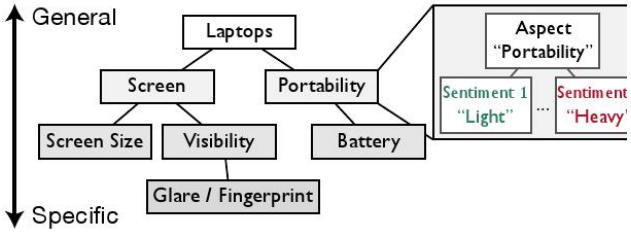


Figure 2. Figure used from (Kim et al., 2013) to show the part of tree structure for the LAPTOP review corpus.

2. Statement of the Problem

Given an online product review $R = \{sen_1, sen_2, ..\}$ we want to find out the sentiment polarities $\{p_{a_1}, p_{a_2}, ..\}$ for the different aspects $\{a_1, a_2, ..\}$ at various granular level $\{g_1, g_2, ..\}$ and thus find out the sentiment(S) of the overall product review.

As a starting point, we want to assume that the hierarchy among the different aspects of the products is already provided and hence the goal is to be able to come up with an equivalent vector representation of the sentences (phrases or words ?) under the aspects at the finest granularity level. The immediate key challenges in this project involve

- Finding out the semantic vector representation for sentences under an aspect in the hierarchy.
  We may need to consider using vector representations for the words or phrases instead of the sentences.
- Compositionality of the sentiments in the sentences under an aspect in the hierarchy.
  However, this does not seem intuitive for the case where a single aspect is discussed at different

places in the review. If all the sentences about a particular aspect are next to each other, then we expect that the sentences are related to each other in some way through discourse analysis.

- Compositionality of the sentiments over the finer aspects to find the sentiment of the parent aspect. Since the hierarchical structure need not be a binary tree, we need to figure out how (if) the compositionality as described in the paper (Socher et al., 2013) need to be modified.

3. General Approach

As discussed above, we would like to assume that the hierarchy among the different aspects of the product is already known. Then, we would like to formulate the problem similar to the semantic compositionality over a sentence. We start with the parsing of the product review based on the hierarchical structure. Then assign initial uniformly distributed vector representations to the sentences and use the same compositionality function as described in the paper cited above. Similarly, we want to find out if using the words or phrases from the sentences is better able to capture the sentiments for an aspect.

The first and foremost task is to be able to find out if our problem is captured well by the model we propose. Since there has not been much work on finding the sentiment of the review using the sentiments of the different aspects in the review, we are restricted to the hierarchy model in the paper cited above for the initial experiments.

Another interesting idea from the paper that builds aspects-hierarchy for online reviews is to find out how it performs on any document (any text). It will be interesting to find out the different clusters within a document that it might generate.

4. Resources

We hope to conduct experiments on two publicly available datasets, Amazon.com reviews of laptops (LAPTOPS) and digital SLRs available at http://uilab.kaist.ac.kr/research/WSDM11 The readings are provided in the references section.

5. Schedule

- **16 March** : Start with some initial experiments based on the model that we finalise before this period.
- **20 April** : Provide the results from the experiments on the datasets and compare the results with the existing techniques.
- **12 May** : Improve the model with the suggestions, if any, during the presentation and also finish the report writing by then.

References

Chen, Yubo and Xie, Jinhong. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Manage. Sci.*

Kim, Suin, Zhang, Jianwen, Chen, Zheng, Oh, Alice H., and Liu, Shixia. A hierarchical aspect-sentiment model for online reviews. In *AAAI*, 2013.

Lu, Bin, Ott, Myle, Cardie, Claire, and Tsou, Benjamin K. Multi-aspect sentiment analysis with topic models. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops*, ICDMW '11.

Socher, Richard, Perelygin, Alex, Wu, Jean, Chuang, Jason, Manning, Christopher D., Ng, Andrew Y., and Potts, Christopher. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642, October 2013.