BANDHAKAVI, A., WIRATUNGA, N., MASSIE, S. and LUHAR, R. 2018. Context extraction for aspect-based sentiment analytics: combining syntactic, lexical and sentiment knowledge. In Bramer, M. and Petridis, M. (eds.) Artificial intelligence xxxv: proceedings of the 38th British Computer Society's Specialist Group on Artificial Intelligence (SGAI) International conference on innovative techniques and applications of artificial intelligence (AI-2018), 11-13 December 2018, Cambridge, UK. Lecture notes in artificial intelligence, 11311. Cham: Springer [online], pages 357-371. Available from: https://doi.org/10.1007/978-3-030-04191-5_30

Context extraction for aspect-based sentiment analytics: combining syntactic, lexical and sentiment knowledge.

BANDHAKAVI, A., WIRATUNGA, N., MASSIE, S., LUHAR, R.

2018

Context extraction for aspect-based sentiment analytics: combining syntactic, lexical and sentiment knowledge

Anil Bandhakavi[1], Nirmalie Wiratunga[1], Stewart Massie[1], and Rushi Luhar[2]

[1] School of Computing, Robert Gordon University, Aberdeen, UK
a.s.bandhakavi, n.wiratunga, s.massie @rgu.ac.uk
[2] SentiSum, London, UK
rushi@sentisum.com

Abstract. Aspect-level sentiment analysis of customer feedback data when done accurately can be leveraged to understand strong and weak performance points of businesses and services and also formulate critical action steps to improve their performance. In this work we focus on aspect-level sentiment classification studying the role of opinion context extraction for a given aspect and the extent to which traditional and neural sentiment classifiers benefit when trained using the opinion context text. We introduce a novel method that combines lexical, syntactical and sentiment knowledge effectively to extract opinion context for aspects. Thereafter we validate the quality of the opinion contexts extracted with human judgments using the BLEU score. Further we evaluate the usefulness of the opinion contexts for aspect-sentiment analysis. Our experiments on benchmark data sets from SemEval and a real-world dataset from the insurance domain suggests that extracting the right opinion context combining syntactical with sentiment co-occurrence knowledge leads to the best aspect-sentiment classification performance. From a commercial point of view, accurate aspect extraction, provides an elegant means to identify ”pain-points” in a business. Integrating our work into a commercial CX platform is enabling the companys[3] clients to better understand their customer opinions.

Keywords: Aspect Extraction · Sentiment Analysis · Natural Language Processing · Machine Learning

1 Introduction

Sentiment analysis (SA) is the computational study of opinionated text with increasing relevance to on-line commercial applications. Sentence level analysis of opinionated content is common but these ignore sentence structure and semantic constructs [14, 10]. Basically they attempt to detect the overall polarity of a sentence, paragraph, or text span, irrespective of the entities mentioned (e.g. restaurant) and their aspects (e.g. price). Increasingly, more granular analysis is needed to better understand the target of the opinion, referred to as the aspect, as well as the context within which that sentiment is being expressed [13]. Indeed the ability to analyze opinionated content beyond just the surface level is crucial to discover meaningful business insights for companies. For

[3] https://www.sentisum.com/

instance given *the food was amazing but the service could have been better*, we can observe that although the overall sentence polarity can be viewed as being positive, there is to some degree a level of negative polarity also being expressed towards aspect, *service*, when sentence context, *the service could have been better*, is inspected more closely.

Context-aware analysis calls for methods that not only extract aspects, but also extract relevant context about each aspect from within the sentence in order to infer the polarity of sentiment (positive, negative or neutral) and its strength expressed numerically on a positive to negative scale [12]. For example, in the sentence *food is good at the restaurant but the price is too high*, there are two aspects (food, price) discussed with differing sentiment. In this example, traditional SA would identify the sentiment to be either positive or negative which is less useful for understanding the specific opinion of the user about food and price at the restaurant.

Typically feedback content has multiple aspects with differing sentiment towards them. Context can allow us to map the relevant sentiment to its associated aspect. Therefore context extraction is important to support the disambiguation of this mapping and therefore improve aspect level sentiment analysis. Accordingly our contributions are:

- a comparative study of context extraction approaches on benchmark and real-world data sets;
- a novel hybrid approach for aspect context extraction which combines syntactic analysis and sentiment co-occurrence knowledge; and
- integration of the approach as a scalable extension into a commercial system that analyses high volume real-world customer feedback data for insight discovery.

In the rest of the paper we review related literature in Section 2. In Sections 3 and 4 we formulate the different methods of opinion context extraction and also describe the sentiment classifiers used. Section 5 describes our evaluation with insights on the experimental datasets and analysis of the results. In Section 6 we consider the role of context extraction in a real world analytics system before presenting our conclusions in Section 7.

2 Related Work

Aspect-Based Sentiment Analysis identifies both the sentiment present in the text as well as the specific target on which the sentiment is expressed. Context plays a key role in mapping sentiment to its target aspect. Aspect-Based Sentiment Analysis can be considered as three staged pipeline: aspect extraction, opinion context extraction and aspect-level sentiment analysis.

Various methods have been used for identifying and extracting aspect terms, for example Conditional Random Fields (CRF), Support Vector Machines (SVM), Random Trees and Random Forest. One approach is to extract all the different nouns and noun phrases from the text and consider them as candidate aspect terms [9]. Schouten develop a co-occurrence based method for category discovery using a dictionary-based sentiment classification algorithm through which aspects can be identified by an annotation process [20]. Alternatively, aspect extraction can be modeled as a sequential labeling task with features extracted for CRF training [22]. In addition to the common features used in Named Entity Recognition (NER) systems, it also uses available external resources for building different name lists and word clusters.

Supervised machine learning can also be used to extract the aspect term [22]. An aspect can be expressed by a noun, adjective, verb or adverb. In [17] the aspect term is extracted by casting it as a sequence tagging task, in which each token in a candidate sentence is denoted as either *beginning*, *inside* or *outside* (BIO). CRFs are used for extracting aspect terms along with the BIO model for representation [3]. The CRFs together with a linear chain CRF are used for determining conditional probability. The authors employ a graph co-ranking approach, to model aspect terms and opinion words as graph nodes, and then they generate three different sub-graphs defining their bond between the nodes [6]. To obtain a list of dependable aspect terms, the candidate nodes are ranked using a combined random walk on the three sub graphs. In this work we extract aspects manually and focus on evaluating different approaches for extracting the context associated with each aspect and also the impact of such contexts on different sentiment classifiers to predict aspect level sentiment.

In order to analyze opinion with reference to a specific aspect (feature) requires, firstly the extraction of phrases or context, followed by the specific aspects related to sentiment or opinion analysis. Context extraction methods tend to utilize frequency related metrics such as relevance and interestingness metrics [24]; as well as the use of dependency parser based extraction patterns [4, 23]. Common to both is the use of noun and verb phrases (NPs, VPs) as indicators of product features and the surrounding dictionary opinion words as opinions. Features are constructed using the phrase dependency tree to extract relations among all product features and opinions that were later used in aspect and opinion expression extraction. Although, these approaches fail to discover aspect specific opinion phrases, the use of NPs in extracting candidate opinion phrases has that effect, and is similar to [5]. In other relevant work on extracting opinion contexts related to aspects include the use of specific rules to refine the dependency tree parse by only accepting it when it fits the specific patterns [18]. Analyzing information regarding predecessors, successors and siblings in a given predecessor tree are common strategies used in these syntax-based methods [8]. Our work also takes advantage of dependency parsers, but additionally we combine analysis of the tree with sentiment co-occurrence statistics to extract candidate opinion phrases for aspect-sentiment analysis. Specifically this allows us to effectively prune the parse tree and home-in on the relevant context content.

The state-of-the-art in sentiment analysis shows a diverse landscape in terms of approaches - from rule-bases and sentiment lexicons [15] to the more supervised classification models generated by shallow and deep learning methods [19]. Rule-based systems and general-purpose lexicons are normally manually created, whilst domain-specific lexicons tends to be generative by leveraging labeled or weakly-labeled text with sentiment classes (e.g. positive, negative) [7]. On the other hand machine/deep learning systems for sentiment analysis apply supervised learning to learn sentiment classifiers to predict the polarity of a given text. A common approach is to learn features from text related to vocabulary (e.g. n-grams), part-of-speech (POS) information, polarity, negation [14, 16, 1]. Recent success in neural architectures include both shallow (e.g. fastText) [10] and deep networks (e.g. convolutional neural networks, recursive/recurrent neural networks) [11, 21]. Typically these have been found to be effective at learning features when there is large amounts of training data. In this paper we do not propose a new sentiment classifier, however we evaluate the effectiveness of different state-of-the-art sentiment analyzers trained using the text generated from the proposed opinion context extraction methods for predicting aspect-level sentiment.

3 Opinion Context Extraction Approaches

The main question we address in this paper is how best to choose the words that constitute the context of a given aspect. Possible approaches are to select words from the full sentence, from a lexical window, or from a syntactic window. The context of the aspect can then be taken as the bag-of-words contained in the associated sentence or window. The sentiment associated with the aspect can then be determined by passing the extracted context to one of several supervised state-of-the-art sentiment classifiers.

3.1 Sentence level context

The baseline strategy for context extraction is to simply use the entire sentence as containing the relevant context for any given aspect, a, in that sentence. Sentiment classification is applied to the entire sentence bearing a and the corresponding prediction assigned to a:

$$sentiment\_classifier(a,\text{sentence}(a)) \tag{1}$$

Where $sentiment\_classifier()$ is a function that predicts the sentiment expressed, in relation to the aspect as positive, negative or neutral. This context extraction approach is reasonable if the sentence contains only a single aspect and the sentiment words in the sentence are used to express opinion towards that aspect. However in real life data, e.g. customer feedback data, sentences often contain multiple aspects and the sentiment towards each can be either positive, negative or neutral. Therefore using the entire sentence as a context is not ideal to accurately determine the opinion towards aspects. In the following sections we propose three alternative approaches that extract part of a sentence with an aim to identify the opinion targeted towards specific aspects.

3.2 Lexical window of context

In this approach we identify a window of k words around an aspect as the context window from which to extract text for sentiment analysis. The size of the window is chosen empirically to be 3. More formally, let a sentence be denoted as $S = \{w_1, \ldots, w_n\}$. Assuming $w_x \in S$ as the aspect a, the lexical window of context for w_x is extracted as follows:

$$Context_{lex}(w_x) = \begin{cases} [w_{x-k} : w_{x+k}] & if x<n-k \text{ and } x>k \\ [w_1 : w_{x+k}] & if x <k \\ [w_{x-k} : w_n] & if x+k >n \end{cases} \tag{2}$$

This approach assumes that the opinion words targeting an aspect occur close by, in the window of k words from the aspect, and that extracting the words within that window gives a useful bag-of-words for analyzing the sentiment of the aspect. For instance with $k = 3$ we would extract just, *The*, to the left-side and, *was amazing but*, to the right-side as our context given the underlined aspect in the following sentence:

The food was amazing but the service could have been better.

Sentiment classification is applied using the lexical context associated with the aspect a and the corresponding prediction is assigned to a as follows:

$$sentiment\_classifier(a, Context_{lex}(a)) \tag{3}$$

3.3 Syntactical window of context

With complex sentences involving multiple aspects one cannot rely solely on adjacency of text as a cue to context identification. For instance in the following examples, *We use that restaurant for Italian food on week days and have always found they serve promptly and the sauces are great*, sentiment about the food itself appears towards the end of the sentence, whilst the mention of the aspect (*food*) is at the beginning of the sentence. Accordingly to link opinion to aspects we need to study syntactical relationships between these components.

procedure SYNTACTIC(w_x,k)
    $L \leftarrow \phi$
    $L \leftarrow$ L.append(w_x)
    while $k > 0$ **do**
        for $l_i \in L$ **do**
            if parent(l_i)$\notin L$ **then**
                L = L.append(parent(l_i)
            else if children(l_i) $\notin L$ **then**
                L = L.append(children(l
            else
                continue
        k = $k - 1$
    return get_context_text(L)

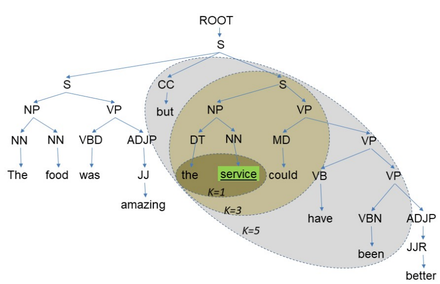

Fig. 1: Algorithm1 - Syntactical window of context algorithm (left) and parse tree analysis for k up to 3 (right).

In the syntactically-informed windowing approach, we study the dependency relationships within a sentence to extract the window of k words to form the context for aspect, w_x. Unlike the lexical window which ignores the syntactic relationships between words, this approach starts from the aspect node and incrementally traverses the dependency parse tree in either direction to arrive at the context text for sentiment analysis. The standard tool used in natural language processing for learning the syntactic structure of sentences is a dependency parser. In this work we use trees constructed through Spacy[4] to extract the relevant text window.

More formally given a sentence, $S = \{w_1, \ldots, w_n\}$, we make use of its dependency tree, $T = \{t_1, \ldots, t_n\}$, where each $t_i \in T$ is a triplet $(w_i, parent(w_i), children(w_i))$ where $parent(w_i) \in S$ and $children(w_i) \in S$. Assuming $w_x \in S$ as the aspect a, the syntactical window of context for w_x is extracted using the algorithm detailed in figures 1. Essentially the approach as described in algorithm 1 traverses the dependency tree to increasing levels guided by the parameter k to collect tree nodes and their corresponding words in order to generate the context for a given aspect.

The example tree (see Figure 1) illustrates how for different values of k we are able to traverse the tree from a given aspect node (such as *service*) to form the neighborhood and therein extract the relevant text appearing within that neighborhood. Clearly the higher the value of k the greater the neighborhood reach. In our experiments we explore the impact of neighborhood size on the different datasets for context extraction.

Once the context text is extracted sentiment classification is applied using the discovered syntactic context associated with the aspect a and the corresponding prediction is assigned to a as follows:

$$sentiment\_classifier(a, Context_{synt}(a)) \tag{4}$$

3.4 Syntactical sentiment weighted co-occurrence window of context

A sentiment-rich corpus of text can be used to learn how often a list of sentiment words and aspects co-occur. Furthermore, this knowledge can be used to guide the traversal of the dependency tree to collect the words that influence the aspect unlike the previous approach which uses distance between words within the tree. Essentially the co-occurrence statistics provides a gauge of how relevant a specific node is likely to be for a given aspect. Aggregating these scores for any candidate sub-tree associated with the aspect provides a heuristic with which we can select the most relevant sub-tree for our context extraction. Unlike 3.3, here we are able to commit to the most promising sub-tree thereby disregarding neighboring sub-trees that are less promising in terms of aspect sentiment relatedness.

In our example tree in figure 2 we can see two candidate sub-trees associated with the aspect *service*. However the correct context relates to the sub-tree containing the opinionated words *but*, *could* and *better*, whilst that containing *tasty* should actually be relevant only to aspect *food*. To disambiguate this context we aggregate co-occurrence statistics within each candidate sub-tree, and associate the sub-tree to the aspect having the higher score. Accordingly the higher co-occurrence score with the sentiment-rich word, *better*, suggests that *service* is more likely to be assigned to the right sub-tree over the left, because *food* is likely to have a stronger co-occurrence with *tasty* (this score is not shown in figure). In this way we combine the syntactic dependencies between words with co-occurrence statistics between aspects and sentiment bearing words to extract context for aspect-sentiment analysis from a sentence.

[4] https://spacy.io/

procedure SYNT_SENTI_COOC(w_x,T,A)
  $w_{x_{context}} \leftarrow$ None
  $candidate\_trees \leftarrow \phi$
  for subtree in T **do**
    $w_{x_{sc}} \leftarrow$ sent_cooc_score(w_x, subtree)
    for $w_i \in A$ **do**
      if $w_i \notin$ subtree or w_i == w_x **then**
        continue
      $w_{i_{sc}} \leftarrow$ sent_cooc_score(w_i, subtree)
      if $w_{i_{sc}} > w_{x_{sc}}$ **then**
        discard_tree(subtree)
        continue
      else
        candidates.append(subtree)
  candidate = get_longest_tree(candidates)
  $w_{x_{context}} \leftarrow$ get_context_text(candidate)
  return $w_{x_{context}}$

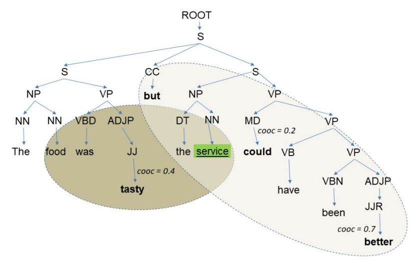

Fig. 2: Algorithm2 - Syntactical sentiment wighted co-occurrence window of context (left) and parse tree illustration (right) for arbitrary subtree.

More formally, let a sentence be denoted as $S = \{w_1, \ldots, w_n\}$. Let T be the dependency tree corresponding to S and T^* be a subtree of T. Let A be the set of aspects identified for a corpus of reviews. Assuming $w_x \in S$ as the aspect a, the algorithm for co-occurrence informed window of context for w_x is extracted as shown in the algorithm in figure 2. The aggregated sentiment weighted co-occurrence statistics for any given tree is obtained using $get\_senti\_cooc\_score$.

Note that although the illustration (in Figure 2) shows two arbitrary subtrees, essentially the approach described in algorithm 2 for a given aspect a, extracts all the subtrees that contains it and selects the sub-tree from them such that there is no other aspect in the subtree with a sentiment weighted co-occurrence score higher than that of a. Sentiment classification is applied using the discovered context associated with the aspect a and the corresponding prediction is assigned to a as follows:

$$sentiment\_classifier(a, Context_{synt\_senti\_coocc}(a)) \tag{5}$$

4 Sentiment Classifier

A diverse set of sentiment classifiers ranging from feature engineering-based (e.g. NRC sentiment) [12] to shallow neural networks (e.g. fastText) [10] to deep neural networks (e.g.convolutional neural network (CNN)) [11] is used to evaluate context extraction quality.

NRC sentiment applies feature engineering to extract different features based on n-grams, part-of-speech (POS) information and sentiment information. n-grams are extracted from text at the sentence level and also within a sentence span whose scope is decided using a dependency tree. The sentence level text is defined as the *surface context* and *parse context* is the sentence span identified using the dependency tree. The parse context is used to extract integer valued features concerning POS information (e.g. no of adjectives, no of verbs etc). Sentiment features extracted are namely: total positive score, total negative score, max sentiment score etc. In our case NRC sentiment classifier uses the sentence level text for n-gram feature extraction and the opinion context text from the proposed Opinion Context Extraction (OCE) methods to extract POS and sentiment related features. Finally we used Support Vector Machine as the classification algorithm to learn the co-relation between the features extracted and the sentiment classes.

fastText is a one-layer shallow neural network. The supervised version of fastText learns the association between words and classes (word vectors) and uses that in turn (average) to learn document representations. The document vectors and the sentiment class labels are modeled using a softmax function to learn a sentiment classifier. We used the context text from the different OCE methods along with the sentiment class labels to train fastText classifier for aspect sentiment prediction.

The CNN used in this work applies one layer of convolution and one layer of pooling on top of word embeddings. The word embeddings are generated for each of the domains (restaurants and insurance) using the unsupervised version of fastText. We used fastText embeddings as it is known to enrich the word embeddings with sub-word information thereby better capturing syntactic variations in the vocabulary [2]. Similar to supervised fastText we feed the context text from the different OCE methods as input for CNN to learn a neural sentiment classifier for aspect sentiment prediction.

5 Evaluation

The aim of the evaluation is to measure the quality of the opinion contexts extracted against human judgments and also to validate the usefulness of the proposed opinion context extraction methods for effective aspect-sentiment analysis. Our evaluation is a comparative study of the performance of the different opinion context extraction methods using evaluation tasks such as text overlap and sentiment analysis.

5.1 Datasets

We used three different data sets (customer reviews) from the domains of restaurants and insurance for our evaluation. The restaurant data sets are official benchmark data sets from the SemEval competition for 2015[5] and 2016[6]. The data set for the insurance domain is a commercial data set. The SemEval data sets are provided with marked aspects and aspect level sentiment labels. We have manually identified a list of about 700 aspects from the insurance domain reviews and have created a sample data set containing these aspects and passed them to Amazon Mechanical Turk to get the aspect level sentiment annotations. Further we have used a sample of sentences from the SemEval

[5] http://alt.qcri.org/semeval2015/task12/
[6] http://alt.qcri.org/semeval2016/task5/

and the insurance data sets to manually annotate the opinion contexts for the aspects. Table 1 captures the volume of data (sentences) for SemEval and insurance data sets.

Table 1: Sentiment Datasets

	SemEval-2015		SemEval-2016		Insurance	
class	Train	Test	Train	Test	Train	Test
Positive	941	193	1008	568	3615	1811
Negative	274	123	430	196	2077	1042
Neutral	36	22	52	36	757	381

5.2 Methods

The following different methods are part of our comparative study:

1. Opinion context extraction method using sentence as a context (refer Section 3.1)
2. Opinion context extraction method using lexical window of words as a context (refer Section 3.2)
3. Opinion context extraction method using syntactic window of k words as a context (refer Section 3.3)
4. Opinion context extraction method using syntactic features and sentiment co-occurrence statistics (refer Section 3.4)
5. Each of the above methods are used to generate a context text for aspects which in turn is used to train the 3 sentiment classifiers from Section 4.

5.3 Results and Analysis

In this section we present the results obtained for different opinion context extraction approaches in text overlap and aspect-sentiment classification tasks.

Text Overlap Analysis Figures 3 and 4 show the text overlap between the dataset's human annotated opinion phrases and the extracted opinion phrases from sentences using the proposed methods. We used BLEU score as a metric to quantify the quality of the extracted opinion phrases. The figures capture sentences sorted by length (ascending order) on the x-axis and BLEU score on the y-axis. It was observed that for shorter sentences the best BLEU score was from sentence based opinion context extraction approach. On the other hand for longer sentences approaches that extract a span of text within the sentence were found to have better BLEU score. This suggests that it is useful to have a context that accurately captures the opinion about an aspect instead of using the entire sentence.

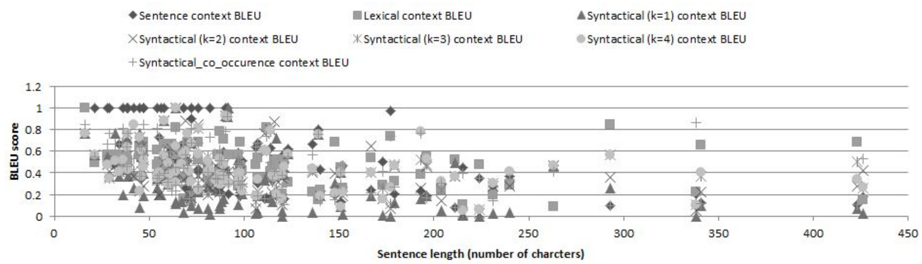

Fig. 3: BLEU score for opinion context extraction methods on insurance data

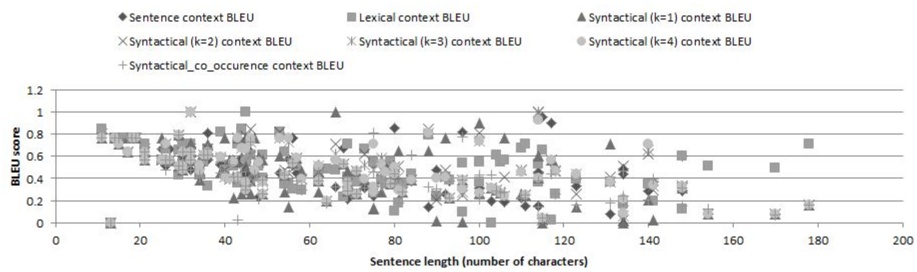

Fig. 4: BLEU score for opinion context extraction methods on SemEval data

Further we also investigated the aspect composition in the sample sentences selected for text overlap analysis. We have segregated the data into two categories namely: sentences with *single aspect* and sentences with *multiple aspects* and observed the BLEU scores within each category. Figure 5 shows the BLEU scores for insurance and SemEval datasets organized by the aspect composition in the sentences.

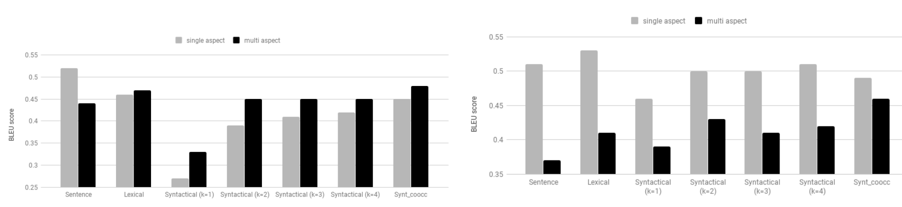

Fig. 5: BLEU score by aspect composition for Insurance (left) and SemEval (right) data

As before we found that for single aspect sentences, considering the complete sentence as a context has higher BLEU score compared to using part of a sentence as a context. On the other hand, for multi-aspect sentences (specially with the insurance data) context extraction was found to have higher BLEU scores than using the entire

Table 2: Aspect Sentiment Classification Results

Classifier	Sentence	Lexical	Syntactic (k=1)	Syntactic (k=2)	Syntactic (k=3)	Syntactic (k=4)	Syntactic Senti_co_occur
				SemEval-2015 data			
FastText	65.28	58.45	60.83	66.76	66.17	68.24	69.45
NRC	69.76	72.81	70.13	70.67	69.12	69.37	**75.24***
CNN	67.14	60.45	62.23	67.35	68.12	70.42	72.26
				SemEval-2016 data			
FastText	73.29	62.72	67.75	71.66	76.19	76.44	77.15
NRC	74.68	77.83	77.14	78.33	78.46	**79.72***	78.67
CNN	72.86	63.24	68.12	70.83	74.12	75.89	76.68
				Insurance data			
FastText	75.72	62.85	64.78	68.37	70.93	70.09	77.45
NRC	78.74	80.05	81.69	82.65	82.77	82.73	**82.87***
CNN	76.12	64.12	66.23	70.23	72.89	74.57	78.67

sentence. This suggests that it is important to identify the right spans of text within a sentence and associate each with the aspects contained in the sentence.

Aspect sentiment classification Here we use sentiment classification as a means to find out how effective each extraction method is for aspect-level sentiment prediction. Table 2 shows the aspect level sentiment prediction results (best overall accuracy highlighted in bold) for SemEval and the insurance data sets. It was observed that using sentence text as the context for aspect sentiment analysis is a strong competition for the other methods which use only a span of text within the sentence as context text for aspect sentiment prediction. We believe this is due to the presence of single aspect bearing sentences, where the entire sentence is a description about one aspect and its sentiment. However context-aware methods outperform the sentence based method suggesting that extracting the right window of text around the aspect is useful for aspect sentiment classification. Further amongst the opinion context methods in general the lexical context based approach has the weakest performance. This could be due to the ineffectiveness of the lexical window in capturing the relevant sentiment words that target the aspects.

For the syntactic context approach the performance of sentiment analysis improves consistently as the value of k increases from 1 to 4. This suggests that having longer context which include immediate as well as distant syntactic dependents for the aspect is more effective to capture the relevant opinion words that target the aspects thereby boosting the sentiment classifier performance. Further the approach which combines the syntactic dependency information with the sentiment co-occurrence information to extract the opinion context surrounding an aspect either records the best performance (SemEval-2015, insurance) or is comparable with the performance of the approach which uses only the syntactic context (SemEval-2016). Overall these results suggests that the sentiment co-occurrence guided sub-tree selection heuristic for disambiguating aspect contexts, is specifically beneficial for multi-aspect sentence analysis.

Amongst the different sentiment learners used, NRC sentiment classifier performs best consistently outperforming its neural counterparts. Amongst the neural methods CNN performs better than fastText. This suggests that with more depth in the neural network there is scope for learning better predictive models. We believe that NRC sentiment classifier learns features that are complementary and are collectively effective in predicting the sentiment at the aspect level. Further in the case of NRC since it uses both the sentence level text and opinion context within the sentence for feature extraction we found it to be having an advantage over the neural classifiers which consider only the opinion context text as input.

Finally we observed that overall performance scores tend to be higher on the real-world insurance data set compared to the SemEval data sets. This could be due to the fact that the insurance data set has more training examples than the SemEval data sets. Nevertheless it is extremely promising to find that the context-aware methods outperform the sentence based method in aspect sentiment predictions confirming our assumption that identifying sentiment at the level of entities and aspects present in the sentence is more important than overall sentiment analysis for extracting value from customer feedback data.

6 Real world analytics system

The technology described in this paper is used by SentiSum to help their clients better understand feedback in the form of on-line reviews and customer satisfaction surveys. The aspect extraction method discussed here helps identify key aspects (referred to as topics) in the feedback text. Topics are grouped into higher level categories which correspond to stages in a customer's interaction with the company. This is known as a customer journey (see Fig 6. Here, we have highlighted the topic ”claim” which is part of the ”claims” customer journey stage for a UK insurer. The system further extracts sentence fragments that describe the sentiment towards the identified topics. This fragment is then fed into domain specific sentiment classifiers generating a positive/negative/neutral label. The distribution of classes over topics and over time forms the info graphic which is used to generate trend insights on customer feedback.

We also compute a customer satisfaction score which is a weighted average of sentiment over selected topics (83% Fig 6). Notice how aspect extraction directly contributes to the discovery of action steps - by unearthing topics that have attracted negative opinion which if fixed can further boost the overall score. These are the customer ”pain points” and drawing attention to these are a valuable feature of the SentiSum offering.

7 Conclusion

In this paper we investigated the role of opinion context extraction for aspect-level sentiment classification with the aim of evaluating the extent to which traditional and neural sentiment classifiers benefit when trained using the opinion context text. We proposed four methods to extract opinion contexts surrounding aspects using lexical, syntactic and sentiment co-occurrence knowledge. Further we validated the quality of the opinion contexts extracted with human judgments using the BLEU score and also

Fig. 6: SentiSum Dashboard Analytics

through standard aspect-sentiment classification tasks. Our experiments on benchmark data sets from SemEval and a real-world dataset from the insurance domain suggests that extracting the right opinion context is effective at improving classification performance. Specifically our heuristic which combines syntactical features with sentiment co-occurrence knowledge leads to significantly better aspect-sentiment classification performance and is currently deployed in the commercial product of SentiSum. Our work in the commercial domain has clearly demonstrated that successful opinion context extraction methods is of key importance for aspect discovery, as they have the power to positively shape the delivery of customer experience analytics.

References

1. Arora, S., Mayfield, E., Penstein-Rosé, C., Nyberg, E.: Sentiment classification using automatically extracted subgraph features. In: NAACL-HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text pp 131–139 (2010)
2. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. In: arXiv preprint arXiv:1607.04606 (2016)
3. Brychcn, T., Konkol, M., Steinberger, J.: Uwb: Machine learning approach to aspect-based sentiment analysis. In: SemEval (2014): 817 (2014)
4. Chen, Y.Y., Wiratunga, N., Lothian, R.: Effective dependency rule-based aspect extraction for social recommender systems. In: 21st Pacific Asia Conf on information systems (2017)
5. Fei, G., Chen, Z., Liu, B.: Review topic discovery with phrases using the polya urn model. In: COLING. pp. 667–676 (2014)
6. Garca-Pablos, A., Cuadros, M., Rigau, G.: V3: Unsupervised aspect based sentiment analysis for semeval-2015 task 12. In: SemEval-2015 (2015)
7. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. Processing (2009)

8. Hazem, A., Morin, E.: Improving bilingual lexicon extraction from comparable corpora using window-based and syntax-based models. In: Computational Linguistics and Intelligent Text Processing. Springer (2014)
9. Hu, M., Liu, B.: Mining opinion features in customer reviews. In: Proc 19th National Conf on AI 2004 (2004)
10. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759 (2016)
11. Kim, Y.: Convolutional neural networks for sentence classification. In: Proc Conf on Empirical Methods in NLP (EMNLP). pp. 1746–1751 (2014)
12. Kiritchenko, S., Zhu, X.D., Cherry, C., Mohammad, S.: Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In: SemEval@COLING (2014)
13. Laddha, A., Mukherjee, A.: Extracting aspect specific opinion expressions. In: Proc Conf on Empirical Methods in NLP. pp. 6270–637 (2016)
14. Mohammad, S.M., Kiritchenko, S., Zhu, X.: Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In: Proc 7th Int Workshop on Semantic Evaluation. pp. 321–327 (2013)
15. Muhammad, A., Wiratunga, N., Lothian, R.: Contextual sentiment analysis for social media genres. Knowledge-Based Systems. **108**, 92–101 (2016)
16. Nakagawa, T., Inui, K., Kurohashi, S.: Dependency tree-based sentiment classification using crfs with hidden variables. In: Human Language Technologies: Conf of the North American Chapter of the ACL. pp. 786–794 (2010)
17. Nandan, N., Dahlmeier, D., Vij, A., Malhotra, N.: Sap-ri: A constrained and supervised approach for aspect-based sentiment analysis. In: SemEval 2014 : 517 (2014)
18. Otero, P.G.: Comparing window and syntax based strategies for semantic extraction. In: Computational Processing of the Portuguese Language. PROPOR 2008. LNCS, vol 5190. Springer (2008)
19. Ribeiro, F.N., Araujo, M., Goncalves, P., Goncalves, M.A., Benevenuto, F.: Sentibench- a benchmark comparision of state-of-the-paractice sentiment analysis methods. Springer EPJ Data Science (2016)
20. Schouten, K., Frasincar, F., de Jong, F.: Commit-p1wp3: A co-occurrence based approach to aspect-level sentiment analysis. In: Proc 8th Int Workshop on Semantic Evaluation (SemEval 2014) (2014)
21. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C., Ng, A., Potts, C.: Recursive deep models for semantic compositionality over a seniment treebank. In: Proc of EMNLP (2013)
22. Toh, Z., Wang, W.: Dlirec: Aspect term extraction and term polarity classification system. In: Proc 8th Int Workshop on Semantic Evaluation (SemEval 2014). pp. 235–240 (2014)
23. Wu, Y., Zhang, Q., Huang, X., Wu, L.: Phrase dependency parsing for opinion mining. In: Proc Conf on Empirical Methods in NLP. vol. 3, pp. 1533–1541 (2009)
24. Zhao, W.X., Jiang, J., He, J., Song, Y., Achananuparp, P., Lim, E.P., Li., X.: Topical keyphrase extraction from twitter. In: Proc 49th Annual Meeting of the ACL: Human Language Technologies-Volume 1, pp 379-388 (2011)