

Aspect Extraction Using Coreference Resolution and Unsupervised Filtering

Deon Mai

School of Computer Science
The University of Adelaide
deon.mai@student.adelaide.edu.au

Wei Emma Zhang

School of Computer Science
The University of Adelaide
wei.e.zhang@adelaide.edu.au

Abstract

Aspect extraction is a widely researched field of natural language processing in which aspects are identified from the text as a means for information. For example, in aspect-based sentiment analysis (ABSA), aspects need to be first identified. Previous studies have introduced various approaches to increasing accuracy, although leaving room for further improvement. In a practical situation where the examined dataset is lacking labels, to fine-tune the process a novel unsupervised approach is proposed, combining a lexical rule-based approach with coreference resolution. The model increases accuracy through the recognition and removal of coreferring aspects. Experimental evaluations are performed on two benchmark datasets, demonstrating the greater performance of our approach to extracting coherent aspects through outperforming the baseline approaches.

1 Introduction

Aspect-based sentiment analysis (ABSA) is a task involving the identification of key terms (words and phrases) that refer to important parts, features, attributes, or properties of a targeted product, object or service, along with associated sentimental emotions, opinions or evaluations. What started out as a simple document-level classification task (Hu and Liu, 2004), i.e., using reviews to differentiate positive from negative, has evolved into a heavily researched field of natural language processing and information retrieval (Godbole et al., 2007). As social presence becomes more standard, the need for detecting opinions in comments or reviews becomes more present. Due to the multi-perspective opinion-oriented nature of the comments, this task will require sentence or phrase-level aspect extraction. The system must be able to locate the expressions of aspects on a sentence-level, for example in

the following examples, the aspects and their associated sentiment are clear; *seaweed* and *chewy*, and *coronavirus* and *hate, terrible* respectively: “*The seaweed was too chewy*”, and “*Hate it, the coronavirus is terrible*”.

The existing approaches for extracting aspect are in two branches: supervised and unsupervised. Supervised approaches often formulate ATE as a token-level sequence labeling problem, achieving better accuracy than unsupervised methods in general (Li and Lam, 2017; Li et al., 2018; Zhou et al., 2019; Ma et al., 2019). However, these approaches generally require annotated data and can run into domain adaptation issues. Moreover, in reality human labelling is a time-consuming and laborious work, motivating the unsupervised approach. Topic model based approaches were proposed for this purpose (Mukherjee and Liu, 2012). These approaches model the text corpus as a mixture of opinion topics, treating the task as a problem in topic coreference resolution. This process labels aspects relating to the extracted opinion topic while dealing with coreferring aspects (Stoyanov and Cardie, 2008; Brody and Elhadad, 2010; Poria et al., 2016). Although the aspects interpreted by these models express a corpus well, they aren’t coherent; individual aspects are of low quality, consisting of irrelevant or distantly-related concepts. The work in (Hu and Liu, 2004) first proposed a manually-defined rule-based approach to extract product features through observing frequent nouns and noun chunks. This approach sparked the development of numerous approaches based on frequent term mining and dependency parsing (Zhuang et al., 2006). Later, the work in (Qiu et al., 2011) proposed a unique approach to learn syntactic relations using dependency trees. Although innovative, the rule-based models heavily relied on predefined rules which only worked well when the aspect terms are confined to a small group of nouns.

In our project, we target the issue of conducting aspect-based sentiment analysis when there is the lack of labelled data, which presents a practical challenge. To this end, as a starting point, we propose an unsupervised approach for aspect extraction on the data corpus, forming the foundation of our following works. We particularly seek an advanced rule-based approach due to its efficiency and independence from manual efforts. We first extract candidate aspects using dependency parsing and coreference resolution. A careful selection process is then applied using unsupervised techniques; inspecting the candidates for duplicate and incorrect aspects. Specifically, syntactic rules are applied on the part of speech (POS) and dependency information of a document to convert it into a candidate aspect list. This candidate list is then reduced to a final list by first applying coreference resolution, removing candidates that refer to an already existing aspect to avoid duplicity. Finally, an unsupervised filtering technique is applied on the candidates, calculating the cosine similarity of an aspect’s word embedding to its neighbours and removing those that don’t meet an optimal threshold. Overall, our proposed approach consists of several stages where in each the candidate list is reduced. This allows our model to overcome the small noun group restraint by first taking in a broad list of noun phrases. A clustering process is applied to complete the categorisation task to an extent.

2 Methodology

The workflow of our proposed unsupervised aspect extraction method can be broken down into four sub-processes: i) Pre-processing and text handling; ii) Noun chunk extraction via dependency parsing; iii) Candidate extraction using rules and coreference resolution; and iv) Aspect term refinement.

2.1 Pre-processing

The pre-processing performed in our approach took the form of two main tasks, applied to each entry in the data set. The first task aims to remove all *stop words* from the text. A list of extremely common English words are pre-defined, representing the stop words to be removed. The next stage of pre-processing involved converting each word to its *lemmatized* (base) form. This is done in order to ensure words correctly match their dictionary entry in the future, specifically within the *Selecting Candidate Aspects* stage, where sentiment words

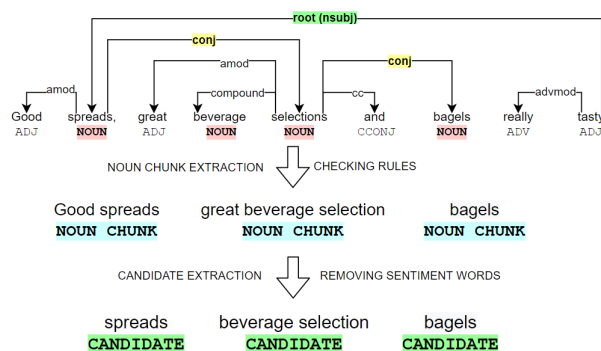


Figure 1: Example of Candidate Aspect Extraction.

need to be identified from a lexicon.

2.2 Noun Chunk Extraction

For each document, the POS and dependency of each word is analysed. We particularly utilize Stanford POS tagger¹ and Stanford dependency parser² to get the POS tags and the dependencies in the sentences. The most important POS tags for this purpose are **NOUN** (noun), **PROPN** (proper noun) and **PRON** (pronoun). The most important relationships used are **nsubj** (nominal subject), **nsubjpass** (nominal subject, passive), **doobj** (direct object), **pobj** (preposition object), **pcomp** (preposition complement), **dative** (dative), **appos** (appositional modifier), **attr** (attribute) and **conj** (conjunct). Using these tags, all *noun chunks* are extracted from the text using a set of lexical checks. These checks are iterated through each token in the document. The first step in Figure 1 illustrates this process. Highlighted are the important POS and dependency tags, which are used to extract the noun chunks. Tokens that have an important POS tag with any of the important dependency tags *except for conj* are extracted as a noun chunk along with all of their syntactic descendants. The tokens with conjunct dependency are held as potential noun chunks - if the first syntactic parent of non-conjunct is found with another important dependency tag, then the original conjunct token is part of a noun chunk. Hence, it is extracted along with its syntactic descendants.

2.3 Candidate Extraction

Using the previously extracted noun chunks, a list of candidate aspects is selected using coreference

¹<https://nlp.stanford.edu/software/tagger.shtml>

²<https://nlp.stanford.edu/software/lex-parser.shtml>

resolution and additional rules. Iterating through the noun chunks for each document, a step-by-step approach is taken following the processes:

2.3.1 Coreference Resolution (CoRef)

This step involves checking each noun chunk to determine if it was already mentioned previously in a different form. For example, in “*The pasta was so tasty. It also had perfect texture.*”, “pasta” and “it” both corefer to the same aspect. Only the first mention of the aspect, “pasta”, should be a candidate which is both “so tasty” and “perfect texture”, such as:

- *Pre-CoRef*: The **pasta**_{ASPECT} was so **tasty**_{pasta}. **It**_{ASPECT} also had perfect **texture**_{It}
- *Post-CoRef*: The **pasta**_{ASPECT} was so **tasty**_{pasta}. It also had perfect **texture**_{pasta}

We adopt (Clark and Manning, 2016) to implement the coreference resolution with minimal fine-tuning. A blacklist for resolving coreferences is created (i.e., the system will judge these words), including the following pronouns: *he, she, it, they, them, him, her, his, hers, its, we, and us*. After performing coreference resolution, each noun chunk is tested to see if it corefers to a previously existing aspect. If so, it is removed from the candidate list.

2.3.2 Individual Token Checking

Once noun chunks satisfy the existing conditions, each token is checked individually. Observing that occasionally opinion words are included in the noun chunks, for example “*very large portions*” produces a noun chunk “*large portions*”, the token is checked to see if it is an opinion word. Using a sentiment lexicon³, tokens are first searched for matching the surface forms, and then matching the lemma forms using their POS tags. If a corresponding match is found within the lexicon, the token is removed from the noun chunk. An example of this is illustrated in the last step of Figure 1.

2.4 Aspect Terms Refinement

Taking a similar approach to Lee et al.’s neighbourhood-based filtering technique (Lee et al., 2017), the extracted candidate aspect list is purged. The aim of this is to remove all false or incoherent aspects. Candidates that stood out had a tendency to be incorrect— if they were not closely related to

other candidates, they were most likely the false aspects. Converting each a candidate into its neural word embedding form (we adopted Word2Vec as the word embedding), they are purged based on their semantic similarity to other candidates. A similarity score between two candidates is expressed as the cosine of the angle over their word vectors. The overall similarity score of a candidate is separated into two sub-scores:

- *AvgSim*: the average similarity of a candidate to all the other candidates. This is calculated by finding the sum of all the similarities and dividing the sum by the number of candidate aspects,

$$AvgSim(a) = \frac{\sum_{n \in C} similarity(a, n)}{|C|}$$

where a is the subject candidate, $similarity(a, n)$ is the similarity score between a and one candidate aspect n , and C is the list of candidate aspects.

- *MaxSim*: the maximum similarity a candidate has to another candidate.

Combining the two scores, an empirically-determined threshold is developed – candidates with similarity scores under this threshold are purged from the aspect list. If a candidate has many other similar candidates, then it coheres that the two scores will be great enough such that it will be considered a valid aspect.

3 Experiments and Results

In this section, We report our evaluations of the proposed approach on two benchmark datasets.

3.1 Experimental Setup

We introduce the datasets used to evaluate our approach, the comparative works as well as the configurations of our approach.

Datasets. We evaluated on two *benchmark datasets*: *i) SemEval 2014 Task 4 - Restaurant* (Pontiki et al., 2014) includes 402 labelled reviews of various restaurants and cafés, used for evaluating our aspect term extraction approach against previous approaches. The dataset has been re-used in SemEval’s later competitions. *ii) SentiHood* (Saeidi et al., 2016) is a labelled corpus of various urban neighbourhood discussions, in which aspects are generalised to two entities. The SentiHood dataset

³Bing Liu’s subjectivity clues (Liu, 2010) lexicon.

Approach	Precision	Recall	F ₁
UFAL	0.50	0.72	0.59
Blinov	0.70	0.72	0.71
iTac	0.37	0.40	0.38
Pre-CoRef (ours)	0.60	0.82	0.70
CoRef (ours)	0.79	0.77	0.78

Table 1: Aspect Term Extraction (SemEval 2014)

focuses on categorical aspect extraction, i.e., aspect term categorization.

Comparative Approaches. In order to validate our model’s performance of aspect term extraction, we compared it against three previous best-performed unsupervised approaches from the SemEval 2014 task 4 competition. We did not observe very recent works for the exact the same aspect extraction purpose. Specifically, the comparing works are *UFAL* (Veselovská and Tamchyna, 2014), *Blinov* (Blinov and Kotelnikov, 2014) and *iTac* (Bornebusch et al., 2014).

Configurations for Aspect Term Refinement. In purging stage, we experimented with different thresholds: the Average Similarity score threshold is tested from 10% to 40%, while the Max. Similarity threshold is tested from 50% to 75%. The best possible combination is discovered through empirical studies, presenting the most accurate purge of incorrect candidates. The *AvgSim* threshold is set to 0.2, and the *MaxSim* is set to 0.55 – candidates that are on average less than 20% similar to other words, or share less than an apex of 55% similarity to another word are purged from the candidate list.

Using our final aspect list, we plot the trained word embeddings of each aspect and perform K-Means clustering to evaluate the coherence and accuracy of aspect categories. The purpose for the clustering is because the benchmark datasets we used for evaluation include aspect category information which would help us evaluate our approach from the category perspective.

3.2 Results and Discussions

We report the evaluation results and the discussions in this section.

3.2.1 Comparisons on SemEval-14

We evaluated the performance of our model on the annotated reviews in the restaurant corpus. The criteria for assessment calculates how accurately the predictions match the true aspects. This is measured by precision, recall, and F_1 scores. The results and comparisons are summarised in Table 1.

The middle ranking of the Pre-CoRef model can be attributed to the purging process, removing incorrect aspects. Without the removal of these candidates (e.g. person relations *boyfriend*, *girlfriend*, and locations *New York*), the precision is considerably lower. On the other hand, the top ranking recall score, while surprising, reflects the accuracy of our rule-based system. Aspects such as person relations (e.g., *boyfriend*, *wife*) are removed, along with those not caught out by the named entity recogniser or pre-processing stage. For example, locations (e.g. *New York*) and the restaurant name itself. The change in results following the implementation of CoRef is as expected. The increased precision proves validity, as the candidate aspect list is severely reduced due to the purge of corefering pronouns. Conversely, the reduction in recall implies meaningful aspects were also purged.

Analysing the errors, we found that if the incorrect aspect is mentioned prior to the correct aspect, the correct aspect is removed. Take the sentence “*Although it’s expensive, the steak was great!*” as an example. The CoRef model identifies *it* as the original aspect, and *steak* as the coreferring aspect, hence removing *steak* from the candidate list. Our final results place our model ahead of all unsupervised approaches in all three scores.

Due to the rule-based noun chunk extraction model and similarity filtering, certain aspects were incorrectly missed. We realised that our model is incapable of identifying unconver sentiment word meanings. From “*The sweet lassi was excellent*”, the correct aspect is *sweet lassi*. An error occurred here in two stage. First, our model extracts only *lassi*, as *sweet* is seen as a sentiment word and removed. During the filtering process, *lassi* is then removed as it was deemed too dissimilar to other aspects. This is most likely due to the fact that it is a foreign dish and hence a possibly uncovered word in the vocabulary.

3.2.2 Findings on SentiHood

The SentiHood dataset focuses on categorical aspect extraction. Included in the corpus are 11 categories for aspects: *live*, *safety*, *price*, *quiet*, *dining*, *nightlife*, *transit-location*, *touristy*, *shopping*, *green-nature*, and *multicultural*. To match these true categories, the number of clusters for the adopted K-Means algorithm is set to 11. We compared these true categories to our aspect categories obtained through grouping together our extracted aspects. In doing so, this clustering process

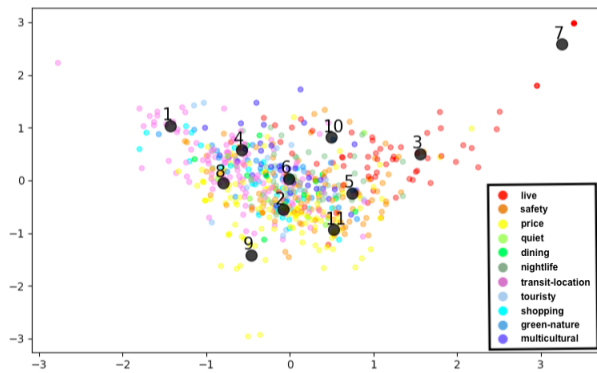


Figure 2: Clustering on Document Embeddings (SentiHood).

reveals the similarity shared between our aspect categories and the true categories; hence the coherence of our aspects. The similarity of our clusters to the true categories indicates the effectiveness of our approach from another angle. To do this, the sentences are converted into their document vector form, trained through doc2vec, and plotted. The results are seen in Figure 2. The colour of each individual point reflects the corresponding true category of that sentence. Our clusters (i.e., the numbers in the figure) could match the ground truth clusters. For example, cluster 3 matches the “safety” cluster. This indicates our aspect term extraction is effective. Moreover, the *live*, *transit-location*, and *price* categories appear to be closely related in the graph in their respective clusters. This is an extension and possible future work to our project, delving into sentiment identification and analysis.

4 Conclusions and Future Work

We have proposed and implemented an approach to aspect extraction utilising an unsupervised rule-based coreference resolution model. The basis of this approach is to apply a rule-based checking system on noun chunks extracted from the text. What started as a simple model has proven itself to be a valid approach, outperforming previous similarly unsupervised approaches. Additionally, the clusters produced on each aspect’s word vector are coherent to a satisfactory level, reflecting the eligibility of our baseline model.

To improve the purging process, word vectors can be learned for a much larger vocabulary. If this can be implemented, foreign dish words such as *rasamalai* won’t be incorrectly ruled out as aspects due to them not being in the vocabulary. Slang in-

terpretations such as *rule* in “*the food options rule!*” can be investigated by using a similar technique to the stop word list. We will also involve machine learning techniques to improve the rule-based approach. Through training our model with the output of rules as an indicator feature for a discriminative learning model, we can expect that our rules are fine-tuned and adaptable to different corpora. Furthermore, to avoid mistakes in clustering where similar words included in different categories are graphed in similar locations, additional learning can be acquired by our model. Extra checks can be performed once a certain black-listed word is found in an aspect, and word embeddings can be trained further. In addition, we will perform sentiment analysis on the extracted aspects and investigate whether public sentiment can reflect the real-estate prices.

References

- Pavel Blinov and Eugeny V. Kotelnikov. 2014. Blinov: Distributed Representations of Words for Aspect-Based Sentiment Analysis at SemEval 2014. In *Proc. of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 140–144, Dublin, Ireland.
- Fritjof Bornebusch, Glauca Cancino, Melanie Diepenbeck, Rolf Drechsler, Smith Djomkam, Alvine Nzeungang Fansu, Maryam Jalali, Marc Michael, Jamal Mohsen, Max Nitze, Christina Plump, Mathias Soeken, Fred Tchambo, Toni, and Henning Ziegler. 2014. iTac: Aspect based sentiment analysis using sentiment trees and dictionaries. In *Proc. of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 351–355, Dublin, Ireland.
- Samuel Brody and Noemie Elhadad. 2010. An Unsupervised Aspect-Sentiment Model for Online Reviews. In *Proc. of 2010 Conference of the North American Chapter of the Association of Computational Linguistics: Human Language Technologies (NAACL-HLT 2010)*, pages 804–812, Los Angeles, California, USA.
- Kevin Clark and Christopher D. Manning. 2016. Deep reinforcement learning for mention-ranking coreference models. In *Proc. of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)*, pages 2256–2262, Austin, Texas, USA.
- Namrata Godbole, Manjunath Srinivasaiah, and Steven Skiena. 2007. Large-scale sentiment analysis for news and blogs. In *Proc. of the First International Conference on Weblogs and Social Media (ICWSM 2007)*, Boulder, Colorado, USA.

- Minqing Hu and Bing Liu. 2004. Mining and Summarizing Customer Reviews. In *Proc. of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004)*, pages 168–177, Seattle, Washington, USA.
- Annie Lee, Wenjie Zi, Afsaneh Fazly, Brandon Seibel, and Anderson De Andrade. 2017. Unsupervised Aspect Extraction from Free-form Conversations. In *Proc. of the 2017 Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM 2017)*.
- Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018. Aspect term extraction with history attention and selective transformation. In *Proc. of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI 2018)*, pages 4194–4200, Stockholm, Sweden.
- Xin Li and Wai Lam. 2017. Deep Multi-Task Learning for Aspect Term Extraction with Memory Interaction. In *Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, pages 2886–2892, Copenhagen, Denmark.
- Bing Liu. 2010. Sentiment analysis and subjectivity. *Handbook of Natural Language Processing, Second Edition*, pages 627–666.
- Dehong Ma, Sujian Li, Fangzhao Wu, Xing Xie, and Houfeng Wang. 2019. Exploring Sequence-to-Sequence Learning in Aspect Term Extraction. In *Proc. of the 57th Conference of the Association for Computational Linguistics (ACL 2019)*, pages 3538–3547, Florence, Italy.
- Arjun Mukherjee and Bing Liu. 2012. Aspect Extraction through Semi-Supervised Modeling. In *Proc. of the 50th conference of the Association for Computational Linguistics (ACL 2012)*, pages 339–348, Jeju Island, Korea.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proc. of the SemEval 2014*, pages 27–35.
- Soujanya Poria, Iti Chaturvedi, Erik Cambria, and Federica Bisio. 2016. Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. In *Proc. of 2016 International Joint Conference on Neural Networks (IJCNN 2016)*, pages 4465–4473, Vancouver, BC, Canada.
- Baojun Qiu, Kang Zhao, Prasenjit Mitra, Dinghao Wu, Cornelia Caragea, John Yen, Greta Greer, and Kenneth Portier. 2011. Get Online Support, Feel Better – Sentiment Analysis and Dynamics in an Online Cancer Survivor Community. In *Proc. of the Third International Conference on Social Computing (SocialCom 2011)*, pages 274–281, Boston, MA, USA.
- Marzieh Saeidi, Guillaume Bouchard, Maria Liakata, and Sebastian Riedel. 2016. SentiHood: Targeted Aspect Based Sentiment Analysis Dataset for Urban Neighbourhoods. In *Proc. of the 26th International Conference on Computational Linguistics (COLING 2016)*, pages 1546–1556, Osaka, Japan.
- Veselin Stoyanov and Claire Cardie. 2008. Topic Identification for Fine-Grained Opinion Analysis. In *Proc. of the 22nd International Conference on Computational Linguistics (COLING 2008)*, pages 817–824, Manchester, UK.
- Kateřina Veselovská and Aleš Tamchyna. 2014. ÚFAL: Using Hand-crafted Rules in Aspect Based Sentiment Analysis on Parsed Data. In *Proc. of the 8th International Workshop on Semantic Evaluation (SemEval@COLING 2014)*, pages 694–698, Dublin, Ireland.
- Yan Zhou, Longtao Huang, Tao Guo, Jizhong Han, and Songlin Hu. 2019. A Span-based Joint Model for Opinion Target Extraction and Target Sentiment Classification. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI 2019)*, pages 5485–5491, Macao, China.
- Li Zhuang, Feng Jing, and Xiao-Yan Zhu. 2006. Movie Review Mining and Summarization. In *Proc. of the 2006 ACM CIKM International Conference on Information and Knowledge Management (CIKM 2006)*, pages 43–50, Arlington, Virginia, USA.