Aspect Extraction Using Coreference Resolution and Unsupervised Filtering for Forum Discussions

ABSTRACT

Aspect extraction is a widely researched field of natural language processing in which aspects are identified from text as a means for information. Previous studies have introduced various approaches of 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 precision through the recognition and removal of coreferring aspects. Experimental results are performed on three data-sets; two benchmark and one real-world, demonstrating the greater performance of our approach on extracting coherent aspects through outperforming baseline approaches.

CCS CONCEPTS

• **Computing methodologies → Natural language processing**; *Machine learning*.

KEYWORDS

aspect extraction, coreference resolution, unsupervised learning

ACM Reference Format:
. 2018. Aspect Extraction Using Coreference Resolution and Unsupervised Filtering for Forum Discussions. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/1122445.1122456

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 [5], i.e. using reviews to differentiate positive from negative, has evolved into a heavily researched field of natural language processing and information retrieval [4]. 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. In supervised approaches, labelled aspect terms are utilised, training a model to distinguish aspects in text [7]. 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 [9]. 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 [13]. 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 [5] 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 [15]. Qiu et al. 2011 [11] 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 aim to conduct aspect-based sentiment analysis on real-world forum discussions in order to examine public sentiment. However, the lack of labelled data presents a practical challenge. To this end, as a starting point, we propose an unsupervised approach for aspect extraction on forum discussions, 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 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.

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

2 METHODOLOGY

The workflow 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; iv) Aspect term refinement. In step i) we perform standard pre-processes to remove stop words and lemmatize the document. In this section we discuss the other steps in detail.

2.1 Noun Chunk Extraction

For each document, the part of speech and dependency of each word is analysed. The most important part of speech 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), **dobj** (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 part of speech and dependency tags, which are used to extract the noun chunks. Tokens that have an important part of speech 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 it's syntactic descendants.

2.2 Candidate Extraction

Using the previously extracted noun chunks, a list of candidate aspects is selected using coreference resolution and additional rules. Iterating through the noun chunks for each document, a step-by-step approach is taken:

2.2.1 Coreference Resolution (CoRef). This 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 [3] 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.

2.2.2 Individual Token Checking. Once noun chunks satisfy the existing conditions, each token is checked individually. Observing that occasionally opinion words are included in noun chunks, for

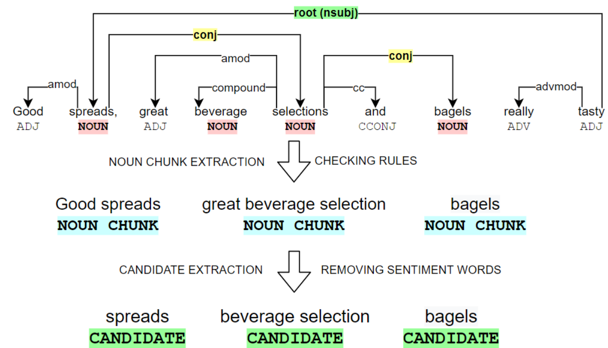

Figure 1: Example of candidate aspect extraction process.

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 lexicon[1], tokens are first searched for matching surface forms, and then matching lemma forms using their part of speech tag. 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.

2.3 Aspect Terms Refinement

Taking a similar approach to Lee et al.'s neighbourhood-based filtering technique [6], 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 false aspects. Converting each candidate into its neural word embedding form, 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 other candidates. This is calculated by finding the sum of all similarities and dividing 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 candidate n, and C is the list of candidate aspects.
- *MaxSim*: the maximum similarity a candidate has to another candidate.

Combining the two results, an empirically-determined baseline is developed; candidates with similarity scores under this baseline 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

We report our evaluations and findings in this section.

[1] Bing Liu's subjectivity clues [8] 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: Results on the SemEval 2014 restaurant corpus

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 evaluate on two ***benchmark datasets***: *i) Sem-Eval 2014 Restaurant Corpus* [10]: Included in this dataset are 402 labelled reviews of various restaurants and cafés, used for evaluating our aspect term extraction approach against previous approaches. *ii) SentiHood* [12]: a labelled corpus of various urban neighbourhood discussions, in which aspects are generalised to two entities. The SentiHood dataset focuses on categorical aspect extraction, i.e. aspect term categorization. We apply our model on ***forum discussions***, from which we aim to perform sentiment analysis later on. The dataset is constructed from multiple threads in the "Real Estate" section of the popular local forum, **Whirlpool**[2], in Australia. Each thread is found by looking for the suburb names in the state of New South Wales (NSW). The discussions reflect the public sentiment on the housing market in NSW.

Comparative Approaches. In order to validate our model's performance of aspect term extraction, we compare it against three previous best-performed unsupervised approaches from the SemEval 2014 task 4 competition. Specifically, they are *ÚFAL* [14], *Blinov* [1] and *iTac* [2]. All works are rule-based approaches.

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%. Through trial and error, the best possible combination is discovered, presenting the most accurate purge of incorrect candidates. The AvgSim threshold is set to 0.2; the MaxSim 0.55 - i.e. 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.

3.2 Results and Discussions

3.2.1 *Comparisons on SemEval-14.*

We evaluate the performance of our model on 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. The change in results following the implementation of CoRef is as expected. The increased precision proves validity, as the candidate aspect list is

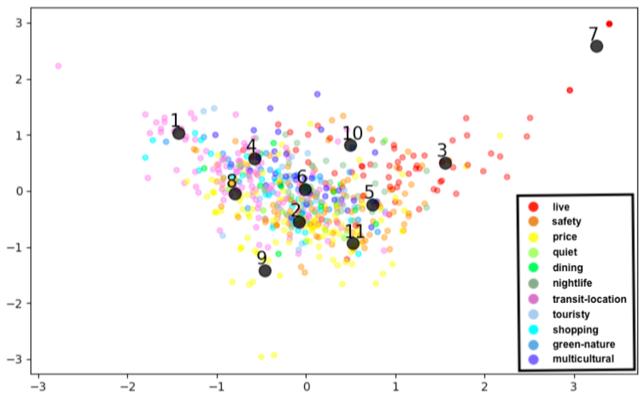

Figure 2: Clustering result on document embeddings, SentiHood corpus.

severely reduced due to the purge of coreferring 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!*". 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 slang 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 foreign word - our model operates solely on an English-based vocabulary.

3.2.2 *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 is set to 11. We compare these true categories to our aspect categories obtained through grouping together our extracted aspects. In doing so, this clustering process 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 indicate 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 can be seen in Figure 2. The colour of each individual point reflects the corresponding true category of that sentence. There is a varying level of coherence between the clusters. For example, the *live, transit-location*, and *price* categories appear to be closely related in the graph in their respective clusters. The categories which have similar properties are dispersed around each other: *quiet* and *safety*, and *shopping, dining* and *nightlife* are similar categories and hence close to each other when graphed. Each of the vectors on the graph appear close-knit; this is expected as the

[2]https://forums.whirlpool.net.au/forum/154

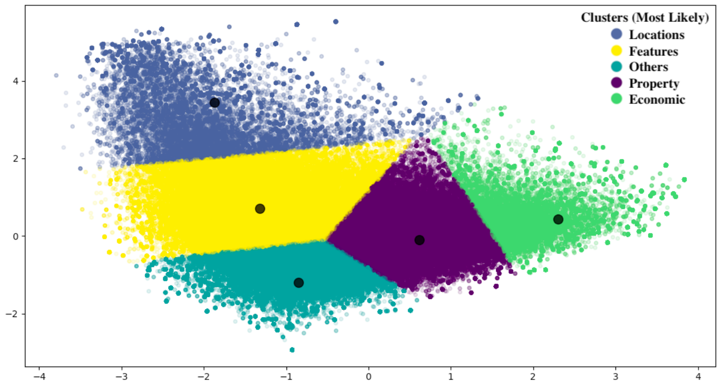

Figure 3: Clustering extracted aspect terms of Whirlpool dataset.

overall topic of discussion is similar. The overall coherence of our clusters reflect the eligibility of our category extraction process.

3.2.3 Applying on Whirlpool Discussions. We extract a complete set of aspects before plotting their word embeddings. Figure 3 displays a scatter plot of each word vector for every aspect along with the resulting clusters; over 900,000 aspects. We find the difference in number of aspects extracted to be 448,980 aspects before and after implementing CoRef. The reduction in aspects of over 30% indicate the effectiveness of our CoRef approach in removing coreferring pronoun clutter. We detect five distinguishable aspect categories from manually observing the dataset: *locations, economy, property, features*, and *others*. Scanning the results for the most common and closest aspects, we found the clusters pleasingly reflect the categories observed in the corpus. Obviously this is not 100% accurate; due to the sparse nature of the unlabelled corpus, several aspects are spread across the different clusters. For example, *bank loan* and *bank property* are two different categorical aspects; *Economy* and *Property* respectively. Due to *bank* being involved in each aspect, their word embeddings translate to similar points on the scatter graph, ending up in the *Economic* category. On the other hand, aspects that have a high level of clarity are most often located in their correct category; reflecting the coherence displayed across the aspects extracted by our model.

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 interpretations 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. Further, we will perform sentiment analysis on the extracted aspects and investigate whether public sentiment can reflect the real-estate prices.

REFERENCES

[1] Pavel Blinov and Eugeny Kotelnikov. 2014. Blinov: Distributed Representations of Words for Aspect-Based Sentiment Analysis at SemEval 2014. In *Proc. of the SemEval 2014*. 140–144.
[2] Fritjof Bornebusch, Glaucia Cancino, Melanie Diepenbeck, Rolf Drechsler, Smith Djomkam, Alvine Nzeungang Fanseu, 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 SemEval 2014*. 351–355.
[3] Kevin Clark and Christopher D. Manning. 2016. Deep Reinforcement Learning for Mention-Ranking Coreference Models. In *Proc. of the EMNLP 2016*. 2256–2262.
[4] Namrata Godbole, Manjunath Srinivasaiah, and Steven Skiena. 2007. Large-Scale Sentiment Analysis for News and Blogs. In *International Conference on Weblogs and Social Media*.
[5] Minqing Hu and Bing Liu. 2004. Mining and Summarizing Customer Reviews. In *Proc. of the KDD 2004*. 168–177.
[6] Annie Lee, Wenjie Zi, Afsaneh Fazly, Brandon Seibel, and Anderson De Andrade. 2017. Unsupervised Aspect Extraction from Free-form Conversations. In *Proc. of the KDD 2017*.
[7] Fangtao Li, Chao Han, Minlie Huang, Xiaoyan Zhu, Ying-Ju Xia, Shu Zhang, and Hao Yu. 2010. Structure-Aware Review Mining and Summarization. In *Proc. of COLING 2010*. 653–661.
[8] Bing Liu. 2010. Sentiment analysis and subjectivity. *Handbook of Natural Language Processing, Second Edition* (2010), 627–666.
[9] Arjun Mukherjee and Bing Liu. 2012. Aspect Extraction through Semi-Supervised Modeling. In *Proc. of the ACL 2012*. 339–348.
[10] 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*. 27–35.
[11] 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 SocialCom 2011*. 274–281.
[12] Marzieh Saeidi, Guillaume Bouchard, Maria Liakata, and Sebastian Riedel. 2016. SentiHood: Targeted Aspect Based Sentiment Analysis Dataset for Urban Neighbourhoods. In *Proc. of the COLING 2016*. 1546–1556.
[13] Veselin Stoyanov and Claire Cardie. 2008. Topic Identification for Fine-Grained Opinion Analysis. In *Proc. of COLING 2008*. 817–824.
[14] Kateřina Veselovská and Aleš Tamchyna. 2014. ÚFAL: Using Hand-crafted Rules in Aspect Based Sentiment Analysis on Parsed Data. In *Proc. of the SemEval 2014*. 694–698.
[15] Li Zhuang, Feng Jing, and Xiao-Yan Zhu. 2006. Movie Review Mining and Summarization. In *Proc. of the CIKM 2006*. 43–50.