V3: Unsupervised Aspect Based Sentiment Analysis for SemEval-2015 Task 12

Aitor García-Pablos, Montse Cuadros
Vicomtech-IK4 research center
Mikeletegi 57
San Sebastian, Spain
{agarciap,mcuadros}@vicomtech.org

German Rigau
IXA Group
Euskal Herriko Unibertsitatea,
San Sebastian, Spain
german.rigau@ehu.es

Abstract

This paper presents our participation in SemEval-2015 task 12 (Aspect Based Sentiment Analysis). We participated employing only unsupervised or weakly-supervised approaches. Our attempt is based on requiring the minimum annotated or hand-crafted content, and avoids training a model using the provided training set. We use a continuous word representations (Word2Vec) to leverage in-domain semantic similarities of words for many of the involved subtasks.

1 Introduction

The continuous growing of textual content on the Internet has motivated an important research on finding automatic ways of processing and exploiting this valuable source of information. That is one of the reasons why sentiment analysis has become a very active research field during the last decade (Pang and Lee, 2008; Liu et al., 2012; Zhang and Liu, 2014). Sentiment analysis aims to detect and classify the polarity of sentiments expressed in a text. The granularity of this classification goes from the overall polarity of full documents to paragraphs, sentences or, as in Aspect Based Sentiment Analysis (ABSA), the sentiment about precise aspects being opinionated (Hu and Liu, 2004) (Popescu and Etzioni, 2005) (Wu et al., 2009) (Zhang et al. , 2010).

In this paper we describe our participation in SemEval-2015 task 12[1] (Pontiki et al., 2015), which is about ABSA. We have participated in all subtasks

---

[1] http://alt.qcri.org/semeval2015/task12/

employing unsupervised or weakly supervised approaches.

The rest of the paper is structured as follows. Section 2 introduces the SemEval-2015 task 12 competition and provided datasets, and a brief introduction about how we have approached the different slots. Sections 3, 4 and 5 describe more in detail the employed techniques. Section 6 shows the results of the evaluation, and finally section 7 summarizes the conclusions.

2 Our approach

SemEval2015 task 12 was about ABSA. The task was divided into 3 slots. Slot 1 was about classifying review sentences into *entity-attribute* pairs, being the entity a main aspect of the reviewed item (e.g. food, drinks, location) and the attribute a particular facet of that aspect (e.g. food-quality, food-price, etc.). Slot 1 runs on two domains, restaurants and laptops. Slot 2 was about detecting explicit mentions to aspect-terms that are being reviewed (e.g. service in "The service was attentive"). Slot 2 runs only on restaurants domain. Slot 3 was about detecting the polarity/sentiment for the given gold entity-attribute pairs in sentences (see slot 1). Slot 3 was meant for restaurants and laptops domain, plus an additional hidden domain (i.e. revealed in the last moment and with no training data available) which resulted to be about hotels.

Two training datasets were provided. The first dataset contains 254 annotated reviews about restaurants (a total of 1315 sentences). The second dataset contains 277 annotated reviews about laptops (a total of 1,739 sentences). The annotation consists of

quintuples of aspect-term, entity-attribute, polarity, and starting and ending position of the aspect-term. When there is no explicit aspect-term mentioned *"null"* is employed to fill the gap. Only the restaurants dataset contains aspect-term annotation.

Our aim is to apply only unsupervised or minimally supervised techniques. We have applied different unsupervised approaches to the different slot tasks avoiding the use of the provided datasets to train a supervised system. We have used them only to evaluate and tune the performance of the employed techniques. For some of the employed techniques we have also used big unlabeled datasets. In particular, for the domain of restaurants we have employed a subset of 100k restaurant reviews from Yelp dataset[2]. We name this corpus as Yelp-restaurants. For laptops domain we have used a subset about 100k reviews from a big dataset of Amazon electronic device reviews [3] (retaining only the ones that contain the word laptop). We name this corpus as Amazon-laptops.

3   Aspect term extraction

SemEval2015 Task 12 slot 2 was about detecting mentions to explicit aspect terms, but only for restaurant domain (i.e. other slots run for restaurants and laptop domains).

For aspect term extraction our aim is to bootstrap a list of candidate domain aspect terms and use it to annotate the reviews of the same domain. We have implemented a system inspired in the method described at (Liu et al., 2014). In this work the authors employ what they call a graph co-ranking approach. They model aspect-terms (AT) and opinion-words (OW) as graph nodes, and then they generate three different sub-graphs defining different types of relations (what they call semantic-relations and opinion-relations) between the nodes. Finally they rank the nodes using a combined random walk on the three sub-graphs to obtain a list of reliable aspect-term candidates. Due to space limitations we cannot explain all the details here. Please, refer to the original article for more in detail explanation.

Based on some of these ideas we have imple-

mented a system that aims to rank aspect-terms modeling them as a graph. From our datasets (i.e. Yelp-restaurants and Amazon-laptops) we have taken nouns as aspect term candidates, and adjectives as opinion word candidates, filtering out those words that appear less than 5 times. These are the nodes to build our graph. Then we have computed our own definition of semantic relations and opinion relations to build sub-graphs as follows:

- Opinion relations (AT-OW edges): we have computed how many times each AT has some syntactical dependency relation with each OW, from a certain set of dependency relations (i.e. direct object, adjectival modifier, attribute of a copulative verb). The result of this count is used as the weight of the edges between AT and OW nodes.

- Semantic relations (AT-AT and OW-OW edges): we have computed a continuous word representation of the datasets employing Word2Vec[4] (Mikolov et al., 2013) (with the following parameters: skip-grams, vector size of 200, context window of 5, hierarchical softmax). Then we have used the cosine similarity between word vectors as the weight of the semantic relation edges.

Once we have built the graph with the different type of nodes and different type of weighted edges, we execute a PageRank (Brin and Page, 1998) (alpha parameter set to 0.15) to score and rank the nodes. With the obtained score we generate an ordered list of aspect terms. We have done this only for restaurants since it was the only domain requested in the task 12 slot 2. Example of some of the higher scored words for restaurant domain are: *food, service, place, restaurant, portion, atmosphere, experience, dish, meal, burger*.

The obtained aspect term list is then cropped to retain only the top N ranked words, and this cropped word list is used to annotate the given sentences performing a simple lemma matching.

3.1 Multiword handling

Handling multiword terms is important in an ABSA system (e.g. it is not the same to detect just memory than flash memory and/or RAM memory, etc.). Multiword terms affect also to some opinion expressions like top notch. Finally, multiword terms arise from usual collocations of single terms, so they vary between domains.

In order to bootstrap a list of candidate multiword terms for each given domain, we have employed again our own Yelp-restaurants and Amazon-laptops datasets. We have computed Log-Likelihood Ratio (LLR) of n-grams (with n¡=3) to detect the more salient word collocations keeping the top K candidates (i.e. the ones with higher confidence of being a true multiword).

Examples of obtained multiwords for restaurants: *happy hour, onion ring, ice cream, spring roll, live music*.

Examples of obtained multiwords for laptops: *tech support, power supply, customer service, operating system, battery life*.

We have used this list in a pre-processing step to merge individual words into a single token when they match a multiword in the list.

4 Entity-attribute detection

The definition of entity-attribute detection in slot 1 states the difference between entities (coarse grained aspects that are being reviewed, e.g. food, drinks) and attributes (particular facet that is being actually reviewed, e.g. price, quality). This subtask runs both for restaurant and laptop domain. Due to the big amount of possible combinations and the consequent overlapping of some of them, this subtask becomes very difficult for an unsupervised system. In order to employ a weakly-supervised approach we have faced this subtask defining some representative seed words for each possible entity (e.g. food: chicken, salad, rice) and attribute (e.g. price: expensive, cheap). Then we reused the Word2Vec model for each domain to compute the similarity between sentence words and the seed words. When the accumulated similarity with some entity and attribute seed words is salient enough, we annotate the sentence with that entity-attribute pair. If the similarity is low, or is equally distributed among a every can-

Word	Polarity Score	Polarity label
delicious	0.424	positive
tasty	0.439	positive
inexpensive	0.341	positive
slow	-0.182	negative
arrogant	-0.254	negative
mediocre	-0.051	negative

Table 1: Examples of polarity values obtained from the restaurants polarity lexicon.

didate entity, we leave the sentence unlabeled.

5 Polarity detection

For polarity detection we have developed a polarity lexicon reusing the generated Word2Vec model for each domain. The intuition we have followed is that a polarity word in a domain should be more "similar" to a set of "very positive" words than to a set of "very negative" words, and vice versa.

We have employed the in-domain generated Word2Vec models because the polarity of words may vary between domains and we want to capture the polarity for each particular domain.

Let POS be a domain-independent positive word (e.g. excellent) and NEG a domain-independent negative word (e.g. horrible). Let W be the set of words we want to know the polarity. Let sim be the similarity between words (computed using the Word2Vec model for the domain). Then for each $w \in W$ we calculate its polarity using (1).

$$\text{polarity}(w) = \text{sim}(w, POS) - \text{sim}(w, NEG) \tag{1}$$

We obtain $polarity(w) > 0$ if the word is more similar to POS than to NEG and vice versa. This gives us a continuous value from very positive to very negative, but we have simplified it to a binary labeling: "positive" for any word w with $polarity(w) >= 0$ and "negative" if $polarity(w) < 0$.

In the table 1 we can see some examples of words, their punctuation in the positive-negative axis, and the assigned polarity label.

With these sentiment lexicons for each of the domains we have performed the annotation of the sentences. We have faced the annotation as a simple polarity count process. For each sentence we counted the polarity of the words regarding our in-domain

Slot 2 systems	Restaurants F-score
Baseline	0.48
V3 (ours)	0.45
Best	0.70
Average	0.52

Table 2: Results on the restaurant reviews for slot 2.

Slot 1 systems	Restaur. F-score	Laptops F-score
Baseline	0.51	0.46
V3 (ours)	0.41	0.25
Best	0.62	0.50
Average	0.53	0.45

Table 3: Results on the restaurant and laptops reviews for entity-attribute detection (SemEval-2015 task 12 slot 1).

Slot 3 accuracy	Restaurants	Laptops	Hotels
Baseline	0.635	0.699	0.716
V3 (ours)	0.694	0.683	0.710
Best	0.786	0.793	0.805
Average	0.713	0.713	0.712

Table 4: Results on the restaurant, laptops and hotels for slot 3.

lexicons and labeled the provided gold quintuples with the most frequent polarity. We have taken into account the negation words (e.g. not) present in the sentence in order to reverse the polarity of the words within a certain window (one token before and two tokens after the current word).

6 Experiments and results

We have participated in SemEval-2015 Task 12 slot 1 (entity-attribute detection), slot 2 (aspect-term detection) and slot 3 (polarity detection). In general the task definition is more challenging than in SemEval-2014 ABSA competition[5] as the average results of all participants indicate. The participation number is also lower and varies between of subtasks and domains (15 participants for restaurants slot 1, 9 for laptops slot 1, 21 for restaurants slot 2, and an average of 14 for slot 3 in the three available domains). As far as we know, we are the only team that has faced the competition using unsupervised approaches. As expected, the supervised systems obtain better results in general than our unsupervised one.

Slot 2 (detecting explicit aspect terms) was only available for restaurants. After performing the steps described in section 3, we employed the top 500 bootstrapped terms to annotate the provided set of reviews using a simple lemma matching. The results are shown in table 2, together with the official results of the supervised baseline, the best performing system, and the average of all participants.

Slot 1 (detecting entity-attribute pairs in sentences) was available both for restaurants and laptops. We employed the described manual bag of words plus Word2Vec approach. The results are quite modest as it can be appreciated in table 3.

Slot 3 (polarity annotation) was available both for restaurants and laptops, plus and additional hidden domain. This hidden domain, which was about hotels, was revealed in the last moment and no training data was provided. For this hidden domain we had no time to develop its own sentiment lexicon so we employed the one from restaurants domain. The results for all domains are shown in table 4.

7 Conclusions

In this paper we have described our participation in SemEval-2015 task 12 (ABSA). We have approached all subtasks from an unsupervised or weakly-supervised point of view. To our opinion this year the tasks were more challenging than in the previous SemEval ABSA edition. We have explored different ways of approaching the challenges without requiring a manually labeled train set. We have made an intensive use of continuous word representations (e.g. Word2Vec) to exploit semantic similarities between words and despite the low results we have found some promising ideas. In the future we will explore how to improve the developed systems and how to combine with other unsupervised or semi-supervised techniques to achieve competitive results.

Acknowledgments

References

Brin, Sergey and Page, Lawrence 1998. The anatomy of a large-scale hypertextual Web search engine *Computer networks and ISDN systems*

Hu, Minqing and Liu, Bing 2004. Mining opinion features in customer reviews *AAAI*

Bo Pang and Lillian Lee 2008. Opinion mining and sentiment analysis *Foundations and trends in information retrieval*,

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean 2013. Efficient Estimation of Word Representations in Vector Space *Proceedings of Workshop at ICLR*

Liu, Kang and Xu, Liheng and Zhao, Jun 2014. Extracting Opinion Targets and Opinion Words from Online Reviews with Graph Co-ranking *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*

Bing Liu 2012. Sentiment analysis and opinion mining *Synthesis Lectures on Human Language Technologies*

Maria Pontiki, Dimitrios Galanis, Haris Papageogiou, Suresh Manandhar, and Ion Androutsopoulos 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado*

Popescu, AM and Etzioni, Oren 2005. Extracting product features and opinions from reviews *Natural language processing and text mining*

Wu, Yuanbin and Zhang, Qi and Huang, Xuanjing and Wu, Lide 2009. Phrase dependency parsing for opinion mining *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3*

Zhang, L and Liu, Bing and Lim, SH and O'Brien-Strain, E 2010. Extracting and ranking product features in opinion documents *Proceedings of the 23rd International Conference on Computational Linguistics*

Zhang, Lei and Liu, Bing 2014. Aspect and Entity Extraction for Opinion Mining *Data Mining and Knowledge Discovery for Big Data*