Integrating Terminology Extraction and Word Embedding for Unsupervised Aspect Based Sentiment Analysis

Luca Dini
Innoradiant
Grenoble, France
luca.dini@innoradiant.com

Paolo Curtoni
Innoradiant
Grenoble, France
paolo.curtoni@innoradiant.com

Elena Melnikova
Innoradiant
Grenoble, France
elena.melnikova@innoradiant.com

Abstract

English. In this paper we explore the advantages that unsupervised terminology extraction can bring to unsupervised Aspect Based Sentiment Analysis methods based on word embedding expansion techniques. We prove that the gain in terms of F-measure is in the order of 3%.

Italiano. *Nel presente articolo analizziamo l'interazione tra syistemi di estrazione "classica" terminologica e systemi basati su techniche di "word embedding" nel contesto dell'analisi delle opinioni. Domostreremo che l'integrazione di terminogie porta un guadagno in F-measure pari al 3% sul dataset francese di Semeval 2016.*

1 Introduction

The goal of this paper is to bring a contribution on the advantage of exploiting terminology extraction systems coupled with word embedding techniques. The experimentation is based on the corpus of Semeval 2016. In a previous work, summarized in section 4, we reported the results of a system for Aspect Based Sentiment Analysis (ABSA) based on the assumption that in real applications a domain dependent gold standard is systematically absent. We showed that by adopting domain dependent word embedding techniques a reasonable level of quality (i.e. acceptable for a proof of concept) in terms of entity detection could be achieved by providing two seed words for each targeted entity. In this paper we explore the hypothesis that unsupervised terminology extraction approaches could further improve the quality of the results in entity extraction.

The paper is organized as follows: In section 2 we enumerate the goal of the research and the industrial background justifying it. In section 3 we provide a state of the art of ABSA particularly focused towards unsupervised ABSA and its relationship to terminology extraction. In section 4 we summarize our previous approach in order to provide a context for our experimentation. In section 5 we prove the benefit of the integration of unsupervised terminology extraction with ABSA, whereas in 6 we provide hints for further investigation.

2 Background

ABSA is a task which is central to a number of industrial applications, ranging from e-reputation, crisis management, customer satisfaction assessment etc. Here we focus on a specific and novel application, i.e. capturing the voice of the customer in new product development (NPD). It is a well-known fact that the high rate of failure (76%, according to Nielsen France, 2014) in launching new products on the market is due to a low consideration of perspective users' needs and desires. In order to account for this deficiency a number of methods have been proposed ranging from traditional methods such as KANO (Wittel et al., 2013) to recent lean based NPD strategies (Olsen, 2015). All are invariantly based on the idea of collecting user needs with tools such as questionnaire, interviews and focus groups. However with the development of social networks, reviews sites, forums, blogs etc. there is another important source for capturing user insights for NPD: users of products (in a wide sense) are indeed talking about them, about the way they use them, about the emotions they raise. Here it is where ABSA becomes central: whereas for applications such as e-reputation or brand monitoring capturing just the sentiment is largely enough for the specific purpose, for NPD it is crucial to capture the entity an opinion is referring to and the specific feature under judgment.

ABSA for NPD is a novel technique and as such it might trigger doubts on its adoption: given the investments on NPD (198 000 M€ only in the cosmetics sector) it is normal to find a certain reluctance in abandoning traditional methodologies for voice of the customer

collection in favor of social network based AB-ABSA. In order to contrast this reluctance, two conditions need to be satisfied. On the one hand, one must prove that ABSA is feasible and effective in a specific domain (Proof of Concept, POC); on the other hand the costs of a high quality in-production system must be affordable and comparable with traditional methodologies (according to Eurostat the spending of European PME in the manufacturing sector for NPD will be about 350,005.00 M€ in 2020, and PME usually have limited budget in terms of "voice of the customer" spending).

If we consider the fact that the range of product/services which are possible objects of ABSA studies is immense[1], it is clear that we must rely on almost completely unsupervised technologies for ABSA, which translates in the capability of performing the task *without* a learning corpus.

3 State of the Art

3.1 Semeval2016's overview

SemEval is " an ongoing series of evaluations of computational semantic analysis systems"[2], organized since 1998. Its purpose is to evaluate semantic analysis systems. ABSA (Aspect Based Sentiment Analysis) was one of the tasks of this event introduced in 2014. This type of analysis provides information about consumer opinions on products and services which can help companies to evaluate the satisfaction and improve their business strategies. A generic ABSA task consists to analyze a corpus of unstructured texts and to extract fine-grained information from the user reviews. The goal of the ABSA task within SemEval is to directly compare different datasets, approaches and methods to extract such information (Pontiki et al., 2016).

In 2016, ABSA provided 39 training and testing datasets for 8 languages and 7 domains. Most datasets come from customer reviews (especially for the domains of restaurants, laptops, mobile phones, digital camera, hotels and museums), only one dataset (telecommunication domain) comes from tweets. The subtasks of the sentence-level ABSA, were intended to identify all the opinion tuples encoding three types of information: Aspect category, Opinion Target Expression (OTE) and Sentiment polarity. Aspect is in turn a pair (E#A) composed of an Entity and

an Attribute. Entity and attributes, chosen from a special inventory of entity types (e.g. "restaurant", "food", etc.) and attribute labels (e.g. "general", "prices", etc.) are the pairs towards which an opinion is expressed in a given sentence. Each E#A can be referred to a linguistic expression (OTE) and be assigned one polarity label.

The evaluation assesses whether a system correctly identifies the aspect categories towards which an opinion is expressed. The categories returned by a system are compared to the corresponding gold annotations and evaluated according to different measures (precision (P), recall (R) and F-1 scores). System performance for all slots is compared to baseline score. Baseline System selects categories and polarity values using Support Vector Machine (SVM) based on bag-of-words features (Apidianaki et al., 2016).

3.2 Related works on unsupervised ABSA

Unsupervised ABSA. Traditionally, in ABSA context, one problematic aspect is represented by the fact that, given the non-negligible effort of annotation, learning corpora are not as large as needed, especially for languages other than English. This fact, as well as extension to "unseen" domains, pushed some researchers to explore unsupervised methods. Giannakopoulos et al. (2017) explore new architectures that can be used as feature extractors and classifiers for Aspect terms unsupervised detection.

Such unsupervised systems can be based on syntactic rules for automatic aspect terms detection (Hercig et al., 2106), or graph representations (García-Pablos et al., 2017) of interactions between aspect terms and opinions, but the vast majority exploits resources derived from distributional semantic principles (concretely, word embedding).

The benefits of word embedding used for ABSA were successfully shown in (Xenos et al., 2016). This approach, which is nevertheless supervised, characterizes an unconstrained system (in the Semeval jargon a system accessing information not included in the training set) for detecting Aspect Category, Opinion Target expression and Polarity. The used vectors were produced using the skip-gram model with 200 dimensions and were based on multiple ensembles, one for each E#A combination. Each ensemble returns the combinations of the scores of constrained and unconstrained systems. For Opinion Target

expression, word embedding based features extend the constrained system. The resulting scores reveal, in general, rather high rating position of the unconstrained system based on word embedding. Concerning the advantages derived from the use of pre-trained in domain vectors, they are also described in (Kim, 2014), who makes use of convolutional neural networks trained on top of pre-trained word vectors and shows good performances for sentence-level tasks, and especially for sentiment analysis

Some other systems represent a compromise between supervised and unsupervised ABSA, i.e. semi-supervised ABSA systems, such an almost unsupervised system based on topic modelling and W2V (Hercig et al., 2016), and W2VLDA (García-Pablos et al., 2017). The former uses human annotated datasets for training, but enrich the feature space by exploiting large unlabeled corpora. The latter combines different unsupervised approaches, like word embedding and Latent Dirichlet Allocation (LDA, Blei et al., 2003) to classify the aspect terms into three Semeval categories. The only supervision required by the user is a single seed word per desired aspect and polarity. Because of that, the system can be applied to datasets of different languages and domains with almost no adaptation.

Relationship with Term Extraction. Automatic Terminology Extraction (ATE) is an important task in NLP, because it provides a clear footprint of domain-related information. All ATE methods can be classified into linguistic, statistical and hybrid (Cabré-Castellvi et al., 2001).

The relationship between word embedding and ATE method is successfully explored for tasks of term disambiguation in technical specification documents (Merdy et al., 2016). The distributional neighbors of the 16 seed words were evaluated on the basis of the three corpora of different size: small (200,000 words), medium (2 M words) and large (more than 200 M words). The results of this study show that the identification of generic terms is more relevant in the large sized corpora, since the phenomenon is very widespread over the contexts. For specified terms, medium and large sized corpora are complementary. The specialized medium corpora brings a gain value by guaranteeing the most relevant terms. As for the small corpora, it does not seem to give usable results, whatever the term. Thus, the authors conclude that word2vec is an ideal technique to constitute semi-automatically term lexicon from very large corpora, without being limited to a domain.

Word2vec's methods (such as skip-gram and CBOW) are also used to improve the extraction of terms and their identification. This is done by the composed filtering of Local-global vectors (Amjadian et al., 2016). The global vectors were trained on the general corpus with GloVe (Pennington et al., 2014), and the local vectors on the specific corpus with CBOW and Skip-gram. This filter has been made to preserve both specific-domain and general-domain information that the words may contain. This filter greatly improves the output of ATE tools for a unigram term extraction.

The W2V method seems useful for the task of categorizing terms using the concepts of an ontology (Ferré, 2017). The terms (from medical texts) were first annotated. For each term an initial vector was generated. These term vectors, embedded into the ontology vector space, were compared with the ontology concept vectors. The calculated closest distance determines the ontological labeling of the terms.

Word2vec method is used also to emulate a simple ontology learning system to execute term and taxonomy extraction from text (Wohlgenannt and Minic, 2016). The researchers apply the built-in word2vec similarity function to get terms related to the seed terms. But the minus-side of the results shows that the candidates suggested by word2vec are too similar terms, as plural forms or near synonyms. On the other hand, the evaluation of word2vec for taxonomy building gave the accuracy of around 50% on taxonomic relation suggestion. Being not very impressive, the system will be improved by parameter settings and bigger corpora.

In the experiments described in this paper we exploit only the Skip-gram approach based on the word2vec implementation. It is important to notice that this choice is not due to a principled decision but to not functional constraints related the fact that that algorithm has a java implementation, is reasonably fast and it is already integrated with Innoradiant NLP pipeline.

4   Previous Investigations

The experiments described in Dini et al. (under review), have been performed by using Innoradiant's Architecture for Language Analytics (henceforth IALA). The platform implements a standard pipelined architecture composed of classical NLP modules: Sentence Splitting → Tokenization → POS tagging → lexicon access → Dependency Parsing → Feature identification

→ Attitude analysis. Inspired by Dini et al. (2017) and Dini and Bittar (2016), sentiment/attitude analysis in IALA is mainly symbolic. The basic idea is that dependency representations are an optimal input for rules computing sentiments. The rule language formalism is inspired by Valenzuela-Escárcega et al. (2015) and thanks to its template filling capability, in several cases, the grammar is able to identify the **perceiver** of a sentiment and, most importantly the **cause**, of the sentiment, represented by a word in an appropriate syntactic dependency with the sentiment-bearing lexical item. For instance the representation of the opinion in *I hate black coffee.* would be something such as:

```
<Opinion trigger="1" perceiver="0"
cause="3">.
```

(where integers represent position of words in a CONLL like structure).

By default entities (which are normally products and services under analysis) are identified since early processing phases by means of regular expressions. This choice is rooted in the fact that by acting at this level multiword entities (such as *hydrating cream*) are captured as single words since early stages.

The goal of the Dini et al. (2018) work was to minimize the domain configuration overhead by i) expanding automatically the polarity lexicon to increase polarity recall and ii) to perform entity recognition by providing only two words (seeds) for each target entity.

Both goals were achieved by exploiting a much larger corpus than Semeval, obtained by automatically scraping restaurant review from TripAdvisor. The final corpus was composed of 3,834,240 sentence and 65,088,072 lemmas. From this corpus we obtain a word2vec resource by using the DL4j library (skip-gram). The resource (W2VR, henceforth) was obtained by using lemma rather than surface forms. Relevant training parameters for reproducing the model are described in that paper.

We skip here the description of i) (polarity expansion) as in the context of the present work *we kept polarity exactly as it was* in Dini & al. (2018)[3]. We just mention the achieved results on polarity only detection which were a precision of 0.78185594 and a recall of 0.54541063 (F-

---

[3] Some previous works on unsupervised polarity lexicon acquisition for sentiment analysis were done in (Castellucci et al., 2016; Basili et al., 2017)

measure: 0.6425726). These numbers are important because in our approach a positive match is always given by a positive match of polarity *and* a correct entity identification (in other words a perfect entity detection system could achieve a maximum of 0.64 precision).

4.1 Entity Matching

Entity matching was achieved by manually associating two seed words to each Semeval entity (RESTAURANT, FOOD, DRINK, etc.) and then applying the following algorithm:

- Associate each entity to the average vector of the seed words (**e-vect.** E.g. evect(FOOD)=avg(vect(*cuisine*),vect(*pizza*)).
- If a syntactic cause is found by the grammar (as in "I liked the **meal**")assign it the entity associated to the closest e-vect.
- Otherwise compute the average vector of n words surrounding the opinion trigger and assign the entity associated to the closest e-vect.

With n=35 we obtain precision= 0.47914252, recall= 0.4888 and F-measure=0.3998.

5 Integrating terminology

A possible path to improve results in entity assignment can be found in the usage of "synonyms" in the computation of the set of e-vect. These can again be obtained from W2VR by selecting the *n* closest world to the average of the seeds and using them in the computation of the e-vect. Expectedly, the value of *n* can influence the result as shown in Figure **1**.
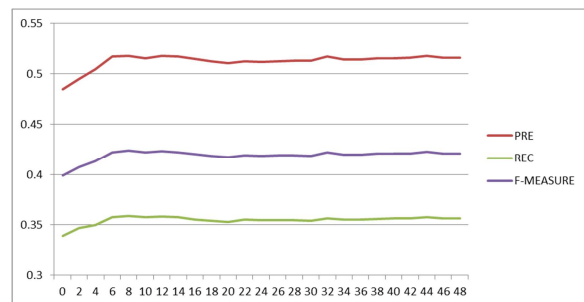

Figure 1. Changes of measures accoring to different top N closest worlds (without terminology).

We notice that best results are achieved by using a set of closest world around 10: after that threshold the noise caused by "false synonyms" or associated common words causes a decay in

the results. We also notice that overall the results are better than the original seed-only method, as now we obtain precision: 0.51 recall: 0.35 F-measure: 0.42. Here the positive fact is not only a global raise of the f-measure, but the fact that this is mainly caused by an increased precision, which according to Dini et al. (2018) is *the* crucial point in POC level applications.

As a way to remedy to the noise caused by an unselective use of the *n* closest words coming from W2VR we decide to explore an approach that filters them according to the words appearing as *terms* in a terminology obtained from unsupervised terminology extraction system. To this purpose we adopted the software TermSuite (Cram & Daille, 2016) which implements a classic two steps model of identification of term candidates and their ranking. In particular TermSuite is based on two main components, a UIMA Tokens Regex for defining terms and variant patterns over word annotations, and a grouping component for clustering terms and variants that works both at morphological and syntactic levels (for more details cf. Cram & Daille, 2016). The interest of using this resource for filtering results from W2VR is that "quality word" lists are obtained with the adoption of methods fundamentally different from W2V approach and heavily based on language dependent syntactic patterns.

We performed the same experiments as W2VR expansion for the computation of e-vect, with the only difference that now the top *n* must appear as closest terms in W2VR *and* as terms in the terminology (The W2VR parameters, including corpus are described in section 4; the terminology was obtained from the same corpus about restaurants). The results are detailed in Figure **2**.
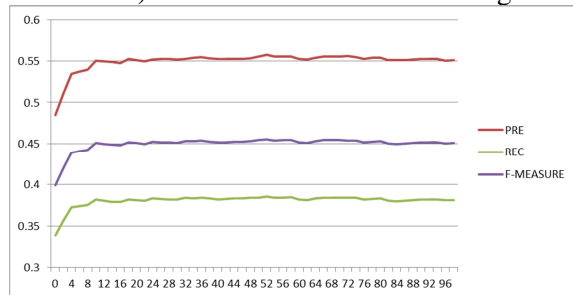

Figure 2. Different measures with top N words filtered with terminology.

We notice that all scores increase significantly. In particular at top *n*=10 we obtain P=0.550233483, R=0.381750288 and F=0.450762752, which represents a 5% increase (in F-measure) w.r.t. the results presented in Dini et al. (2018).

6   Conclusions

Many improvements can be conceived to the method presented here, especially concerning the computation of the vector associated to the opinionated windows, both in terms of size, directionality and consideration of finer grained features (e.g. indicators of a switch of topic). However our future investigation will rather be oriented towards full-fledged ABSA, i.e. taking into account not only Entities, but also Attributes. Indeed, if we consider that the 45% F measure is obtained on a corpus where only 66% sentences were correctly classified according to the sentiment and if we put ourselves in a Semeval perspective where entity evaluation is provided with respect to a "gold sentiment standard" we achieve a F-score of 68%, which is fully acceptable for an almost unsupervised system.

References

Ehsan Amjdian, Diana Inkpen, T.Sima Paribakht and Farahnaz Faez. 2016. Local-Global Vectors to Improve Unigram Terminology Extraction. *Proceedings of the 5th International Workshop on Computational Terminology,* Osaka, Japan, Dec 12, 2016, 2-11.

Marianna Apidianaki, Xavier Tannier and Cécile Richart. 2016. Datasets for Aspect-Based Sentiment Analysis in French. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016).* Portorož, Slovenia.

Roberto Basili, Danilo Croce and Giuseppe Castellucci. 2017. Dynamic polarity lexicon acquisition for advanced Social Media analytics. *International Journal of Engineering Business Management*, Volume 9, 1-18.

David M. Blei, Andrew Y. Ng and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *The Journal of machine Learning research*, Volume 3, 993–1022.

M. Teresa Cabré Castellví, Rosa Estopà Bagot, Jordi Vivaldi Palatresi. 2001. Automatic term detection: a review of current systems. In Bourigault, D. Jacquemin, C. L'Homme, M-C. 2001. *Recent Advances in Computational Terminology*, 53-88.

Giuseppe Castellucci, Danilo Croce and Roberto Basili. 2016. A Language Independent Method for Generating Large Scale Polarity Lexicons. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16)*, Portoroz, Slovenia, 38-45.

Damien Cram and Béatrice Daille. 2016. TermSuite: Terminology Extraction with Term Variant Detection. *Proceedings of the 54th Annual Meeting of*

the Association for Computational Linguistics—System Demonstrations, Berlin, Germany, 13–18.

Luca Dini, Paolo Curtoni and Elena Melnikova. 2018. Portability of Aspect Based Sentiment Analysis: Thirty Minutes for a Proof of Concept. Submitted to: *The 5th IEEE International Conference on Data Science and Advanced Analytics*. DSAA 2018, Turin.

Luca Dini, André Bittar, Cécile Robin, Frédérique Segond and M. Montaner. 2017. SOMA: The Smart Social Customer Relationship Management. *Sentiment Analysis in Social networks. Chapter 13*. 197-209. DOI: 10.1016/B978-0-12-804412-4.00013-9.

Luca Dini and André Bittar. 2016. Emotion Analysis on Twitter: The Hidden Challenge. *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016*, Portorož, Slovenia, 2016.

Arnaud Ferré. 2017. Représentation de termes complexes dans un espace vectoriel relié à une ontologie pour une tâche de catégorisation. *Rencontres des Jeunes Chercheurs en Intelligence Artifcielle (RJCIA 2017)*, Jul 2017, Caen, France.

Aitor García-Pablos, Montse Cuadros and German Rigau. 2017. W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis. *Expert Systems with Applications, (91):* 127-137. arXiv:1705.07687v2 [cs.CL], 18 jul 2017.

Athanasios Giannakopoulos, Diego Antognini, Claudiu Musat, Andreea Hossmann and Michael Baeriswyl. 2017. Dataset Construction via Attention for Aspect Term Extraction with Distant Supervision. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*

Tomáš Hercig, Tomáš Brychcín, Lukáš Svoboda, Michal Konkol and Josef Steinberger. 2016. Unsupervised Methods to Improve Aspect-Based Sentiment Analysis in Czech. *Computación y Sistemas, vol. 20*, No. 3, 365-375.

David Jurgens and Keith Stevens. 2010. The S-Space package: An open source package for word space models. *Proceedings of the ACL 2010 Systel Demonstrations*, 30-35.

Noriaki Kano, Nobuhiku Seraku, Fumio Takahashi, Shinichi Tsuji, 1984. Attractive Quality and Must-Be Quality. *Hinshitsu*: *The Journal of the Japanese Society for Quality Control*, 14(2) : 39-48.

Emilie Merdy, Juyeon Kang and Ludovic Tanguy. 2016. Identification de termes flous et génériques dans la documentation technique : expérimentation avec l'analyse distributionnelle automatique. *Actes de l'atelier "Risque et TAL" dans le cadre de la conférence TALN*.

Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*

Tomas Mikolov, Wen-tau Yih and Geoffrey Zweig. 2013b. Linguistic regularities in continuous space word representations. *Proceedings of NAACL-HLT 2013*, 746–751.

Dan R. Olsen. 2015. *The Lean Product Playbook: How to Innovate with Minimum Viable Products and Rapid Customer Feedback.* John Wiley & Sons Inc: New York, United States.

Jeffrey Pennington, Richard Socher and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543.

Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clecq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeny Kotelnikov, Nuria Bel, Salud M. Jiménez-Zafra, Gülşen Eryiğit. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*. San Diego, USA.

Marco A. Valenzuela-Escárcega, Gustave V. Hahn-Powell and Mihai Surdeanu. 2015. Description of the Odin Event Extraction Framework and Rule Language. arXiv:1509.07513v1 [cs.CL], 24 Sep 2015, version 1.0, 2015.

Lars Witell, Martin Löfgren and Jens J. Dahlgaard. 2013. Theory of attractive quality and the Kano methodology – the past, the present, and the future. *Total Quality Management & Business Excellence*, (24), 11-12:1241-1252.

Gerhard Wohlgenannt, Filip Minic. 2016. Using word2vec to Build a Simple Ontology Learning System. *Proceedings of the ISWC 2016 co-located with 15th International Semantic Web Conference (ISWC 2016)*. Vol-1690. Kobe, Japan, October 19, 2016

Dionysios Xenos, Panagiotis Theodorakakos, John Pavlopoulos, Prodromos Malakasiotis, Ion Androutsopoulos. 2016. AUEB-ABSA at SemEval-2016 Task 5: Ensembles of Classifiers and Embeddings for Aspect Based Sentiment Analysis. *Proceedings of SemEval-2016*, San Diego, California, 312–317.

Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP),* 1746–1751. arXiv:1408.5882