Combining Probabilistic Language Models for Aspect-Based Sentiment Retrieval

Lisette García-Moya, Henry Anaya-Sánchez, and
Rafael Berlanga-Llavori

Universitat Jaume I, Spain
`{lisette.garcia,henry.anaya,berlanga}@uji.es`

Abstract. In this paper, we present a new methodology aimed at retrieving relevant product aspects from a collection of customer reviews, as well as the most salient sentiments expressed about them. Our proposal is both unsupervised and domain independent, and does not relies on NLP techniques such as parsing or dependence analysis. In our experiments, the proposed method achieves good values of precision. It is also shown that our approach is capable of properly retrieving the relevant aspects and their sentiments even from individual reviews.

Keywords: Sentiment analysis, aspect retrieval, language models

1 Introduction

One of the most relevant applications of sentiment analysis is the aspect-based summarization [1, 3]. Broadly, given a collection of opinion posts about a product or service, this task is aimed at obtaining the most relevant opined aspects (also called features) along with their most relevant sentiment information expressed by customers (usually an opinion word and/or a polarity score).

Aspect-based summarization usually is composed of three main tasks: aspect identification, sentiment classification, and aspect rating. *Aspect identification* is focused on extracting the set of aspects concerning the product from the reviews. Most approaches to this task mainly rely on NLP techniques such as parsing or dependency analysis [4, 3]. Thus, these approaches heavily depend on (domain dependent) extraction patterns, and do not account for aspect relevance. The *sentiment classification* task consists in determining the opinions about the product aspects and/or their polarities, whereas *aspect rating* leverages the relevance of aspects and their opinions to properly present them to the users. These tasks have been usually formulated as classification problems. Unfortunately, it is infeasible to obtain sufficient training examples for all the products and all the potential aspects regarded in opinion posts. Thus, supervised approaches have been mainly focused on obtaining sentiment lexicons and detecting the subjectivity of sentences containing them.

In this paper, we address the aspect-based summarization task by introducing an unsupervised and domain independent methodology based on a combination of stochastic language models for retrieving the most relevant product aspects from a collection of customer reviews, as well as their most relevant sentiments.

2   A Model for Aspect-Based Sentiment Retrieval

Given a collection of customer reviews about a specific product and a free-text document d, which can be either a subcollection of reviews or an individual review, our goal is to obtain a ranking of the relevant aspects of the product together with their most relevant sentiments as expressed by customers.

For this purpose, we consider that the set of relevant aspects discussed in d can be modeled by a probabilistic language model that assigns higher probability values to words defining aspects.

In the context of customer reviews, opinion words (e.g "good", "bad", etc.) are usually utilized to express sentiments about the different aspects of a product. This causes the review texts to reflect some entailment relationship from opinion words to aspect words. In this way, we consider the use of an (stochastic) entailment-based self-translation model between the words in d to reveal a probability distribution of words that approaches the language model of aspects expressed in d from a general probabilistic model of opinion words. Thus, we firstly regard a unigram language model of aspects P defined as follows:

$$P(w_i) \propto \sum_{j=1}^{n} t(w_i|w_j)\, Q(w_j) = \left(T \cdot \begin{pmatrix} Q(w_1) \\ \vdots \\ Q(w_n) \end{pmatrix} \right)_i \tag{1}$$

where $V = \{w_1, \ldots, w_n\}$ represents the vocabulary of d, the n-by-n matrix $T = \{t(w_i|w_j)\}_{1 \leq i \leq n, 1 \leq j \leq n}$ represents the entailment-based self-translation model of words from d, and Q is a generative model of opinion words (i.e. Q assigns to each word w its likelihood of being an opinion word, namely $Q(w)$).

This unigram language model of aspects can be used to generate an arbitrary sequence of words $s = w_{i_1} \ldots w_{i_r}$ by considering $P^*(s) = \prod_{t=1}^{r} P(w_{i_t})^{1/r}$.

2.1   Estimation of the Entailment-Based Self-Translation Model Between Words

To estimate the entailment model T, we rely on the basic entailment-based self-translation model $\{\mathrm{p}(w_i|w_j)\}_{1 \leq i,j \leq n}$ such that $\forall j \in \{1, \ldots, n\}$, $\{\mathrm{p}(w_i|w_j)\}_{1 \leq i \leq n}$ represents the distribution of words in V according to their occurrence in those local contexts from d (e.g. N-grams of length 5) in which w_j occurs. Let $\mathcal{T}$ be the n-by-n matrix such that $\mathcal{T}_{i,j} = \mathrm{p}(w_i|w_j)$.

We regard several versions of T defined from different powers of $\mathcal{T}$, i.e. $T = T(k) = \mathcal{T}^k$ (for $k \geq 1$). These models correspond to different smoothing degrees of the basic entailment model $\mathcal{T}$. The greater the value of k the greater the smoothing of the model. Our hypothesis is that the posterior probability of aspects in $T(k)$ can be increased as k grows, and therefore aspect retrieval can also be improved.

Relying on $\mathcal{T}$, we also rank sentiment words with respect to the sequence $s = w_{i_1} \ldots w_{i_r}$ by regarding the score $R(w) = \prod_{t=1}^{r} \mathrm{p}(w_{i_t}|w)^{1/r} Q(w)$.

2.2   The Generative Model of Opinion Words

Considering that opinion words entail similar posterior distribution of words in a stochastic entailment model such as $\mathcal{T}$, we rely on a kernel-based density estimation approach for defining the generative model of opinion words Q. Thus, given a predefined set of (general-domain) opinion words $\{u_1, \ldots, u_m\}$ [1], we define $Q(w) = \frac{1}{m} \sum_{i=1}^{m} k(w, u_i)$, where $k(w, u_i)$ is the gaussian kernel $k(w, u_i) = \exp\left(-0.5h(g(w), g(u_i))^2/\sigma^2\right)$ in which h represents the geodesic distance between distributions [2], $g(w)$ is the posterior distribution of words from the collection of reviews conditioned in word w (defined in the same way as the conditional distributions from the basic entailment model $\mathcal{T}$), and σ is a predetermined distribution width.

2.3   Refining the Unigram Language Model of Aspects

Since some product aspects can be expressed by means of multi-word phrases (e.g. "*plug and play*", "*signal to noise ratio*", "*on/off button*", etc.), the unigram language model P might assign high probability values to meaningless words (e.g. determiners, prepositions, etc.), and therefore the quality of the generative model P^* can be affected. To overcome this problem, we propose to refine model P by considering that it can be expressed as a mixture of two unigram distributions: (i) the refined unigram model of aspects P', and (ii) a background unigram model P_{bg} of the source language of the reviews (e.g. English). That is, $P(w_i) = \lambda P'(w_i) + (1 - \lambda)P_{bg}(w_i)$. We estimate P' by means of an EM routine aimed at minimizing the cross entropy $-\sum_{i=1}^{n} P(w_i) \log(\lambda P'(w_i) + (1 - \lambda)P_{bg}(w_i))$ given an estimation of P_{bg} from the COCA corpus. [2]

3   Evaluation and Conclusions

To evaluate our approach, we rely on a collection of customer reviews from five products widely tested in the literature. [3] This collection has been manually annotated at the sentence level with the relevant aspects referred to in the text.

The retrieval performance is measured in terms of **precision@N**. Table 1 shows the results obtained for each product of the collection. As it can be seen, our method obtains overall good results despite of being an unsupervised approach applied on free-text reviews. The best results are obtained from the sequence-based model. This was expected since many relevant aspects correspond to word sequences. Also, the performance improves with smoother models (higher k).

Figure 1 shows the top ranked aspects obtained for some individual reviews together with their top scored sentiments. Notice that our method is able to identify the most relevant aspects and their sentiments from both positive and negative opinion reviews.

---

[1] http://www.cs.uic.edu/ liub/FBS/opinion-lexicon-English.rar
[2] http://corpus.byu.edu/coca/
[3] http://www.cs.uic.edu/ liub/FBS/CustomerReviewData.zip

Table 1. Retrieval performance obtained for different values of k (sequences are generated with length $r \leq 3$ and they are validated by their occurrence in the reviews).

	Improved unigram model														
	Apex			**Canon**			**Creative**			**Nikon**			**Nokia**		
N	k=1	k=3	k=20	k=1	k=3	k=20	k=1	k=3	k=20	k=1	k=3	k=20	k=1	k=3	k=20
10	0.70	0.70	0.70	0.80	**0.90**	**0.90**	0.70	**0.80**	**0.80**	0.80	**0.90**	**0.90**	0.60	**0.70**	**0.70**
20	0.45	0.60	**0.65**	0.80	**0.90**	**0.90**	0.70	**0.85**	**0.85**	0.60	0.75	**0.80**	0.55	**0.75**	**0.75**
30	0.43	**0.70**	**0.70**	0.69	**0.93**	**0.93**	0.63	0.77	**0.83**	0.53	0.77	**0.80**	0.53	0.70	**0.80**
	Sequence-based generative model														
	Apex			**Canon**			**Creative**			**Nikon**			**Nokia**		
N	k=1	k=3	k=20	k=1	k=3	k=20	k=1	k=3	k=20	k=1	k=3	k=20	k=1	k=3	k=20
10	0.90	0.90	**1.00**	**1.00**	0.90	0.90	0.80	0.80	0.80	1.00	1.00	1.00	**0.80**	0.70	**0.80**
20	0.75	**0.85**	**0.85**	0.80	**0.85**	**0.85**	**0.85**	0.80	0.75	1.00	1.00	1.00	0.75	0.80	**0.85**
30	0.73	0.80	**0.83**	0.83	**0.87**	**0.87**	**0.77**	0.73	**0.77**	**1.00**	0.97	0.97	0.83	0.83	0.83

Fig. 1. Top ranked aspects and their sentiments retrieved for some reviews.

Apex AD2600 DVD player: review6	Aspects	Sentiments
Their customer service sucks. No way to contact their customer service. Very bad quality.	quality	bad
	customer_service	suck
Canon G3: review13	**Aspects**	**Sentiments**
Love my new g 3. Just received this camera two days ago and already love the features it has. Takes excellent photos. Night mode is clear as day. I have not played with all the features yet, but the camera is easy to use once you get used to it. The only drawback is the viewfinder is slightly blocked by the lens. However, using the lcd seems to eliminate this minor problem. Overall it is the best camera on the market. I give it 10 stars!	camera	best, easy
	photo	excellent
	viewfinder	drawback, slightly, block
	lcd	use, eliminate
	lens	slightly, block
	star	give
	night_mode	clear
	market	best, camera
Creative Lab: review4	**Aspects**	**Sentiments**
	sound	great, good
The sound is the thing! For me the sound of the music is what is important. The zen xtra sounds great with good headphone and well recored mp3s. Navigation is a bit tricky and takes getting used to but the included pdf instructions cover the operation pretty well. At 8 ounces it is pretty light but not as light as the ipod. The battery is replaceable (a real plus) and there is a leather case (another real plus). I suppose that you could jog with it but it seems to be made for the enjoyment of music. So far the software for the pc works easily. By the way, it looks nice also.	music	important, enjoyment
	headphone	good, well, great
	pdf_instruction	include, cover
	software	work, pc
	leather_case	real
	pc	work, software, easily
	operation	pretty, well, cover
	zen_xtra	great, sound
	ounce	pretty, light
	battery	replaceable
	recored_mp3	well
	ipod	light
	navigation	tricky, bit

These results validate the usefulness of the proposed model to address aspect-based summarization. As future work, we plan to rate individual customer reviews according to relevancy.

References

1. Carenini, G., Ng, R., Pauls, A.: Multi-document summarization of evaluative text. In: Proc. of EACL 2006. pp. 305–312 (2006)
2. Dillon, J., Mao, Y., Lebanon, G., Zhang, J.: Statistical Translation, Heat Kernels, and Expected Distance. In: Proc. of the 23rd Conference on Uncertainty in Artificial Intelligence (2007)
3. Yu, J., Zha, Z., Wang, M., Chua, T.: Aspect ranking: identifying important product aspects from online consumer reviews. In: Proc. of ACL 2011. pp. 1496–1505 (2011)
4. Zhang, L., Liu, B., Lim, S.H., O'Brien-Strain, E.: Extracting and ranking product features in opinion documents. In: Proc. of COLING-2010. pp. 1462–1470 (2010)