pablos_2018_w2vlda_almost_unsupervised_system_for_aspect_based_sentiment_analysis

Year

2018

Author(s)

Aitor Garcia-Pablos and Montse Cuadros and German Rigau

Title

W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis

Venue

Expert Systems with Applications

Topic labeling

Partially automated (W2VLDA) and manual (Baselines)

Focus

Primary

Type of contribution

Novel

Underlying technique

(MaxEnt) based classification model, Manual labeling

Topic labeling parameters

\

Label generation

For baselines

The authors examined each topic and manually set a label according to their judgment.

W2VLDA

W2VLDA only requires a minimal domain aspects […] configuration […], which consists on defining a single seed aspect-term for each desired domain aspect, plus a single positive seed word and a single negative seed word independent of the domain.
These features are the only domain and language dependent information required by W2VLDA

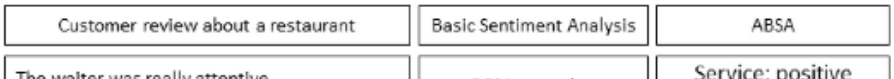

Fig. 1. An example of classical Sentiment Analysis vs. Aspect Based Sentiment Analysis.

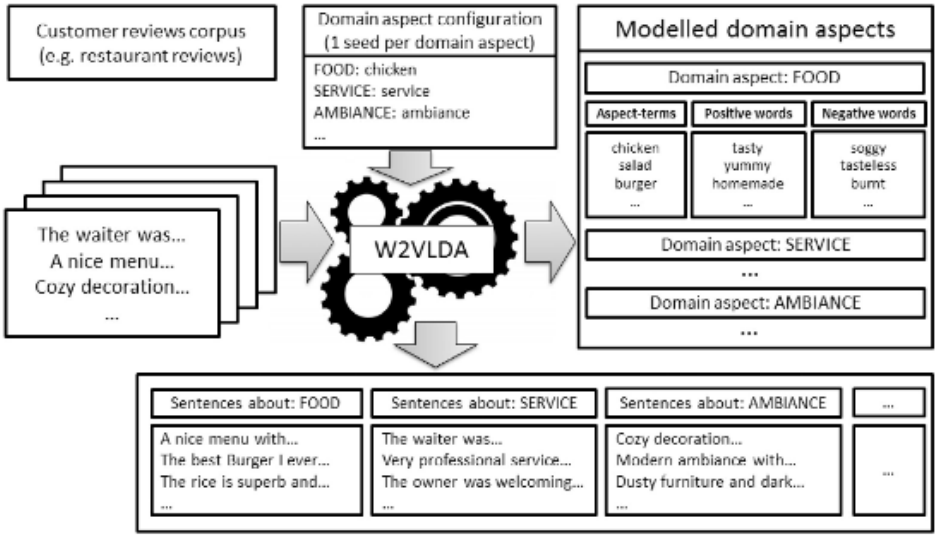

Fig. 2. A schema of W2VLDA. The input is an unlabelled corpus of a particular domain and its domain aspects specification. Domain aspects are split into three word distributions: aspect-terms, positive words and negative words. Sentences are modelled by domain aspect and polarity.

W2VLDA provides a set of labelled topics according to the domain aspects defined in the initial configuration step, so no manual topic inspection and labelling are required.

Conceptually, each domain aspect keyword is used as a topic label for the three topics (Aspect-terms, positive words, negative words) generated by the model.

Table 2
Resulting domain aspect word distributions for English in two different domains. The domain aspects are automatically split into three different word distributions: aspect terms, positive words and negative words.

Language:domain	Domain aspect	Aspect-terms	Positive words	Negative words
English: restaurant reviews	Food	Chicken, beef, pork, tuna, egg, onions, shrimp, curry	Moist, goat, smoked, seared, roasted, red, crispy, tender	Undercooked, dry, drenched, overcooked, soggy, chewy
	Service	Staff, workers, employees, chefs, hostess, manager, owner	Helpful, polite, knowledgeable, efficient, prompt, attentive	Inattentive, rude, unfriendly, wearing, making, packed
	Ambiance	Lighting, wall, interior, vibe, concept, ceilings, setting, decor	Modern, beautiful, chic, nice, trendy, cozy, elegant, cool	Bad, loud, uninspired, expensive, big, noisy, dark, cramped
English: electronic devices reviews	Warranty	Warranty, support, repair, service, answer, center, policy	Worked, lucky, owned, big, exchange, extended, longer	Called, contact, broken, faulty, defective, expired, worthless
	Design	Plastic, wheel, style, handle, pocket, design, exterior, wheels	Adjustable, clean, good, versatile, attractive, lightweight, stylish	Ugly, odd, awkward, tight, felt, weird, cute, stupid, flimsy
	Price	Money, store, item, bucks, price, regret, deal, gift	Paying, reasonable, penny, worth, delivered, stars, inexpensive,	Disappointed, paid, cheaper, skeptical, pricey, overpriced

Motivation

Topic modeling

LDA-based
(W2VLDA - extended to include aspect-term and opinion-word separation and sentiment polarity classification for each defined domain aspect)
Baselines: LocLDA (Brody & Elhadad, 2010) and ME-LDA (Zhao et al., 2010).

Topic modeling parameters

α: 50/T
δ: T
β: 0.01
(T being the nr of topics)
Nr of topics (T): 14

Nr of iterations: 500
burn-in period: 100 iterations
sampling lag: 10 iterations

Nr. of topics

14

Label

For baselines.
Single or multi-word label manually provided by the authors

W2VLDA
Single word domain aspect label passed to W2VLDA as part of the domain aspect configuration

Label selection

\

Label quality evaluation

\

Assessors

\

Domain

Paper: Opinion mining
Dataset: Restaurant reviews

Problem statement

Supervised approaches for Aspect Based Sentiment Analysis achieve good results for the domain and language they are trained on, but manually labelling data to train supervised systems for all domains and languages is very costly and time consuming.
In this work, we describe W2VLDA, an almost unsupervised system based on topic modelling that, combined with some other unsupervised methods and a minimal configuration step, performs aspect category classification, aspect-term and opinion-word separation and sentiment polarity classification for any given domain and language.

Corpus

Dataset 1

Origin: Citysearch New York Ganu, Elhadad, and Marian (2009)
Nr. of documents: 52264
Details:

- restaurant reviews labelled with domain-related aspects (e.g., food, staff, ambience) in English.
- All reviews present in the system were extracted over the course of one week in 2006.

Dataset 2

Origin: Various popular customer review websites

Nr. of documents: 64730

Details:

- Restaurant reviews in English, Spanish, French and Dutch.

Dataset 3

Origin: SemEval-2016 (task 5)

Nr. of documents:

Details:

Document

Dataset 1 - 2

Reviews contain structured metadata (star rating, date) along with text. Typically reviews are small; the average user review has 5.28 sentences

Pre-processing

\

```
@article{pablos_2018_w2vlda_almost_unsupervised_system_for_aspect_based_sentiment_analysis,
  abstract = {With the increase of online customer opinions in specialised websites and social networks, automatic systems to help organise and classify customer reviews by domain-specific aspect categories and sentiment polarity are more needed than ever. Supervised approaches for Aspect Based Sentiment Analysis achieve good results for the domain and language they are trained on, but manually labelling data to train supervised systems for all domains and languages is very costly and time consuming. In this work, we describe W2VLDA, an almost unsupervised system based on topic modelling that, combined with some other unsupervised methods and a minimal configuration step, performs aspect category classification, aspect-term and opinion-word separation and sentiment polarity classification for any given domain and language. We evaluate its domain aspect and sentiment classification performance in the multilingual SemEval 2016 task 5 (ABSA) dataset. We show competitive results for several domains (hotels, restaurants, electronic devices) and languages (English, Spanish, French and Dutch).},
```

```
  author = {Aitor Garc{\'\i}a-Pablos and Montse Cuadros and German Rigau},
  date-added = {2023-03-20 21:46:57 +0100},
  date-modified = {2023-03-20 21:46:57 +0100},
  doi = {https://doi.org/10.1016/j.eswa.2017.08.049},
  issn = {0957-4174},
  journal = {Expert Systems with Applications},
  keywords = {Opinion mining, Aspect Based Sentiment Analysis, Almost unsupervised, Multilingual, Multidomain},
  pages = {127-137},
  title = {W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis},
  url = {https://www.sciencedirect.com/science/article/pii/S0957417417305961},
  volume = {91},
  year = {2018}}
```

#Thesis/Papers/Initial