Aspect Extraction from Online Consumer Reviews with WordNet-Guided Continuous-Space Language Models

Completed Research Paper

Jie Tao
Fairfield University
Fairfield, CT, USA, 06824
jtao@fairfield.edu

Lina Zhou
The University of North Carolina at Charlotte
Charlotte, NC, USA, 28223
lzhou8@uncc.edu

Conor Feeney
Fairfield University
Fairfield, CT, USA, 06824
conor.feeney@student.fairfield.edu

Abstract

Online consumer reviews contain rich yet implicit information regarding consumers' preferences for specific aspects of products/services. Extracting aspects from online consumer reviews has been recognized as a valuable step in performing fine-grained analytical tasks (e.g. aspect-based sentiment analysis). Extant approaches to aspect extraction are dominated by discrete models. Despite explosive research interests in continuous-space language models in recent years, these models have yet to be explored for the task of extracting product/service aspects from online consumer reviews. In addition, previous continuous-space models for information extraction have largely overlooked the role of semantic information embedded in texts. In this study, we propose an approach of aspect extraction that leverages semantic information from WordNet in conjunction of building continuous-space language models from review texts. The experiment results with online restaurant reviews demonstrate that the WordNet-guided continuous-space language models outperform both discrete models and continuous-space language models without incorporating the semantic information. The research findings have important implications for understanding consumer preferences and improving business performances.

Introduction

Online review platforms allow consumers to freely express their opinions on products and services in form of free text. These textual reviews contain opinions and previous experiences of existing customers toward various aspects of the products/services. Thus, they play an integral role in the decision making processes of prospective consumers. However, the large volumes of unstructured review texts hinder not only the ability of consumers to fully leverage the power of other peers' experiences and opinions, but also that of businesses to improve their product/service performance. Automatic extraction of aspects from review text has a great potential to provide fine-grained support to better facilitate consumers and businesses in their decision making process.

Aspect is a generic term that refers to attributes, components, and related concepts of a product/service as a whole or as a part. There has been a stream of research in aspect extraction from online consumer reviews (Hu et al. 2016; Kang and Zhou 2017; Poria et al. 2016). Existing methods for aspect extraction can be classified into statistical and rule-based categories. One major strength of the statistical methods lies in its ability to learn from data with little or no human supervision. However, completely ignoring human knowledge can be ineffective. In contrast, the rule-based methods are able to leverage human knowledge in learning aspects from text. On the other hand, such methods may not be adaptable or generalizable to different contexts. To address the above-mentioned limitations of existing methods, we introduce an approach to aspect extraction from online consumer reviews in this study.

The proposed approach learns keywords of aspects from continuous language space of online consumer reviews. The method design introduces several unique advantages: First, it is semi-supervised which only requires a small set of manually selected feature expressions, and thus is easy to apply. Second, it embeds word sense disambiguation into feature learning, which improves not only the efficiency of feature learning by reducing feature space but also the outcome of feature learning. Thirdly, it is capable of learning informal expressions of features, which are ubiquitous in online consumer reviews.

We evaluate the proposed method using a large dataset of online restaurant reviews. The evaluation consisted of both direct and indirect methods. The direct method compare the aspect keywords extracted by using the proposed method against human-engineered keywords; and the indirect method measure the effect of using the extracted keywords of aspects in predicting some key performance measures such as review rating and useful vote. To support the evaluation, we develop a time series analysis model based on a deep learning technique. The experiment results demonstrate that the keywords extracted by the WordNet-guided continuous-space language models outperform other alternative methods.

The rest of the article is organized as follows. Section 2 reviews the related literature on aspect extraction, continuous-space language models, and semantic similarity. Section 3 presents the proposed analytical approach and its main components. In Section 4, we describe the experiment setting and report evaluation results. We also discuss the insights learned from the evaluation in Section 5. Finally, Section 6 concludes the paper with a summary of contributions.

Related Work

Aspect Extraction from Online Consumer Reviews

Aspect extraction is an important task for understanding online consumer reviews, as consumers more often express their preferences over different aspects of products and/or services in their reviews (Kang and Zhou 2017). Extant approaches to aspect extraction can be categorized into statistical-based and rule-based methods.

The statistical-based methods typically utilizes shallow textual statistics (e.g. word frequencies) to identify domain-specific keywords. Siering et al. (2018) utilize frequent words (frequency > 2.5%) in online consumer reviews and human intelligence (i.e. domain knowledge) to identify keywords related to different aspects of airline services. Other methods in this category acquire a domain-specific list of keywords by leveraging domain knowledge structures (i.e. ontologies) or empirical results. Peñalver-Martinez et al. (2014) utilize classes (as seed aspects) from an existing movie ontology to compute semantic similarity between terms appearing in online reviews and the seed aspects in identification of aspects from movie reviews. Similarly, Lau et al. (2014) utilize a fuzzy product ontology to extract features for mining opinions from online reviews. To incorporate the contextual information from online review texts into aspect extraction, recent studies employ topic models such as Latent Dirichlet Allocation (LDA). Duric and Song (2012) propose a Hidden Markov Model based LDA (HMM-LDA) model to extract aspects from movie reviews; and Qiao et al. (2017) design a domain-oriented LDA model to extract product defects from online consumer reviews. The other stream of aspect extraction approaches is rule-based, which utilize pre-defined rules to identify (implicit or explicit) patterns from online reviews. For instance, Archak et al. (2011) propose a linguistic pattern based approach using Parts-of-Speech (POS) to identify product aspects from online reviews. Hu et al. (2016) utilize sentence level linguistic patterns and semantic similarity to extract hotel aspects from consumer reviews. Recently, Kang and Zhou (2017) propose a syntactic dependency based approach enhanced with two-stage pruning strategies to extract product aspects from the online reviews of both search and experience goods.

The statistical modeling approaches are dominant in this area. However, they require a robust domain-specific knowledge structure and a large corpus of annotated text corpus to yield quality results. Both of them remain as scarce resources. In addition, traditional topic models (e.g. LDA-based) belong to discrete text representation models – e.g. Bag-of-Word models (C. Li et al. 2017), and consequently they are incapable of capturing the contextual information. This along with the unsupervised nature of these methods often lead to relatively low-quality results and high computational costs. On the other hand, rule-based approaches require human intervention in developing the rules, and the resulting rules may not be generalizable across different domains.

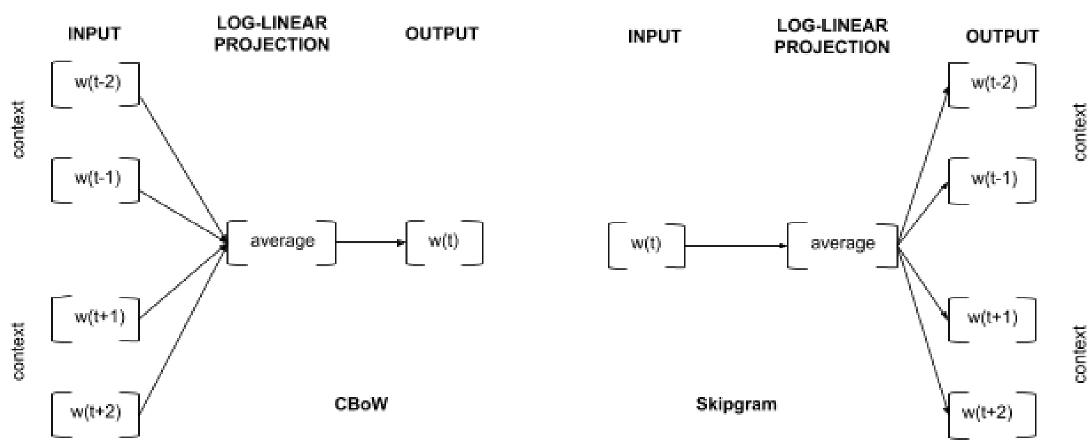


Figure 1. Architectures of CBoW and Skipgram Models

Continuous-space Language Models

Continuous-space language models are based on neural networks, which represent words and phrases in a corpus as high-dimensional, real-valued vectors (Mikolov et al. 2013). Extant studies suggest that continuous-space language models can better capture the contextual information than discrete models. One of its common core is the word2vec model (Mikolov et al. 2013), and continuous bag-of-words (CBoW) and skipgram models are two major types of continuous-space language models. The CBoW model uses the context to classify the target word, while the skipgram models uses the target word to classify the context. Both models utilize a fix-sized sliding window over text to incorporate the contextual information of any word in the text. Figure 1 depicts the architectures of the two types of continuous-space language models.

However, basic continuous-space language models do not capture *syntactic* and *semantic* information in the texts. Researchers have proposed to incorporate *syntactic* information (e.g. POS tags) into continuous-space language models to acquire domain-specific sentiment lexicons from financial reports (Tsai et al. 2016). Nevertheless, aspect extraction is a distinct research topic from topic extraction in that the former faces more significant semantic ambiguities than the latter, among others. The proposed method, however, is incapable of capturing the *semantic* information in the extracted keywords.

Semantic Similarity and Word Sense Disambiguation

Word sense disambiguation refers to an intelligent process, in which a word is given an explicit meaning given certain context (Navigli 2009). The measurement of semantic similarity is one key step in word sentence disambiguation.

Semantic similarity is a measure of the semantic relatedness between terms, such as 'is-a' or 'a-kind-of', which has been widely applied in Natural Language Processing (NLP) and other related domains (Tao et al. 2015). Different semantic similarity metrics have been developed in the literature (Leacock and Chodorow 1998; Lin 1998; Resnik 1995), which can be categorized into *corpora-based* and *knowledge-based* methods. A detailed survey of different metrics can be found in (Sánchez et al. 2012). *Corpora-based* metrics rely on the co-occurrence of a pair of terms within the document corpus, while *knowledge-based* metrics map the terms to a formal knowledge structure such as WordNet (Princeton-University 2012) or other domain ontologies. The knowledge-based measures often reflect the collection of human intelligence and outcome of empirical validation, and accordingly they can better ensure results quality. For instance, WordNet builds in well-defined semantic structure, and is a widely used source for word sense disambiguation tasks. The knowledge-based similarity metrics supported by WordNet are more preferable in this work because they rely on knowledge networks rather than (enormous) document corpus or (implicit) external knowledge (Ge and Qiu 2008). In this study, we adopt a WordNet-based measure, namely WuP similarity (Wu and Palmer 1994), in pruning the extracted aspect keywords via word sense disambiguation. Details of the metric is discussed in the next section.

The Proposed Approach

The proposed approach to extracting keywords of aspects from online consumer reviews consists of three main components, namely *text preprocessing*, *candidate keyword extraction*, and *WordNet-guided pruning*. In the remainder of this section, we focus on the introduction to the last two components. For the purpose of incorporating semantic information in the aspect extraction process, we follow the research procedure as shown in Figure 2.

Text Preprocessing

The preprocessing component involves standard text preprocessing tasks, such as sentence tokenization, tokenization/lemmatization, POS tagging, and stop word removal. Lemmatization rather than stemming is performed for two main reasons. One is that different morphs of the same word may have different POSs/senses. The other is that it is easy to evaluate lemmas as keyword extraction results. The POS tag assigned to individual words allows us to enrich the information for learning candidate keywords in a continuous-space language model. Common stop words (e.g. 'a', 'the') are removed because these words carry little meaning in the text. In addition, we also remove non-word tokens and non-English tokens.

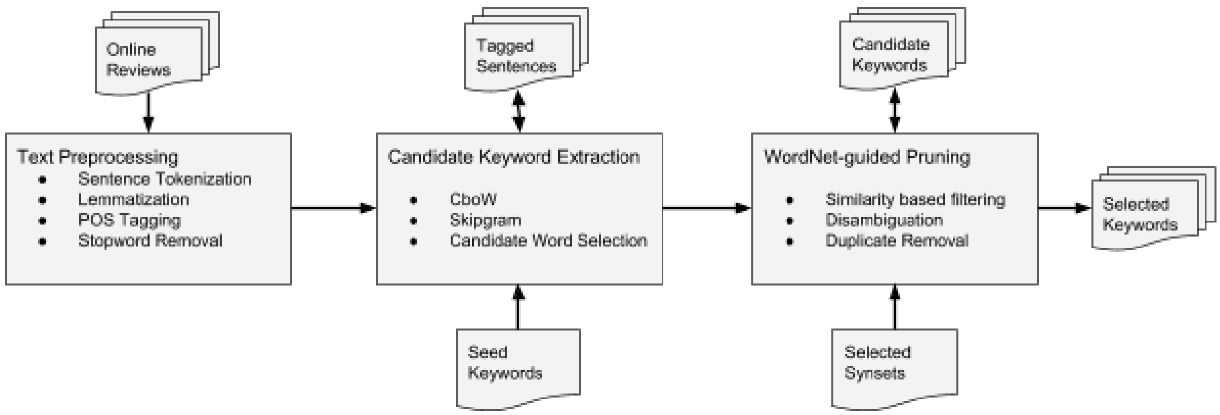


Figure 2. Main Components of the Proposed Method

Candidate Keyword Extraction

The candidate keywords are extracted using continuous-space language models, which incorporate contextual and syntactic information from the text. CBoW and skipgram models have their own unique strengths. For instance, skipgram models can better capture semantic information in general (Li et al. 2017); and CBoW models require less time to train (Tsai et al. 2016). Thus, we make the design decision that incorporates the results from both types of models.

During the process of learning word embedding, we pair up word tokens and POS tags. In addition, the method is seeded by a small set of keyword-POS pairs for each aspect. For each aspect, we learn the top-K most similar token-POS tag pairs to the seed keyword-POS pairs in the word embedding. The similarity is measured using the cosine similarity function (see eq. (1)).

$$\cos(\vec{w}_x, \vec{w}_y) = \frac{\vec{w}_x \cdot \vec{w}_y}{||\vec{w}_x||\,||\vec{w}_y||} = \frac{\sum_{i=0}^{n-1} w_x^i w_y^i}{\sqrt{\sum_{i=0}^{n-1}(w_x^i)^2}\sqrt{\sum_{i=0}^{n-1}(w_y^i)^2}} \quad (1)$$

where w_x^i and w_y^i are the i^{th} element in n-dimensional word vectors $\vec{w}_x$ and $\vec{w}_y$ generated from either the CBoW or the skipgram model, respectively. The cosine similarity of word vectors helps ensure the contextual relatedness of candidate words.

WordNet-guided Pruning

To improve the quality of extracted keywords, we introduce WordNet-guided pruning as a critical post-processing step. WordNet has been widely used in traditional word sense disambiguation research. Nevertheless, the role of WordNet has not been explored in the context of continuous-pace language models. For instance, recent lexicon extraction methods for continuous-space language models (e.g. (Tsai et al. 2016)) do not address the issue of extracting semantic information from text corpus. This study aims to fill the gap by incorporating semantic information from WordNet.

In this component, we utilize WuP similarity– which measures the normalized distance between a pair of concepts and their Least Common Subsume (LCS), as shown in eq. (2). The main rationale is that WuP relies on the relative depth in WordNet instead of absolute depth as used in path similarity. This not only makes the former more computationally efficient than the latter, but also enables the comparison across multiple domain knowledge bases (Tao et al. 2015).

$$sim_{WuP}(w_1,w_2)=\frac{2\times depth(LCS,root)}{depth(w_1,LCS)+depth(w_2,LCS)+2\times depth(LCS,root)} \quad (2)$$

In eq. (2), LCS is the farthest shared parent term of a pair of terms (words) according to WordNet; root denotes the root node in WordNet; and depth computes the number of intermediate nodes between a pair of nodes.

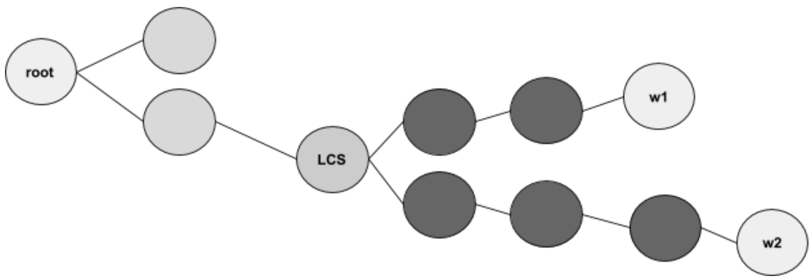


Figure 3. An Example of WuP Similarity Calculation

Figure 3 illustrates how to compute WuP similarity, where $depth(w_1, LCS) = 2$, $depth(w_2, LCS) = 3$, and $depth(LCS, root) = 1$. Based on eq. (2), $\mathrm{sim_{WuP}}(w_1, w_2) \approx 0.2857$. Figure 4 presents the algorithm in pseudo code.

INPUT: Candidate Keywords for feature f: kw_list, where kw_i is a keyword in *kw_list*;
INPUT: All WordNet synsets for each kw_i: $\mathrm{synsets}_{kw_i} = [(syn_{kw_i}^{l})]$
INPUT: list of seed keywords for feature f: seed_list, where seed_i is a word in seed_list;
INPUT: Selected WordNet synset for each seed_i: $\mathrm{syn}_{seed_i} = (w_s, pos_s, sense_s)$
OUTPUT: Finalized keywords for feature f: $\mathrm{kws}^f = [(syn_{kw_i}^{k})]$

FUNCTION WordNet_guided_pruning:

//BEGIN
FOREACH $\mathrm{seed}_i \in$ seed_list: // Iterate through seed keyword list
  FOREACH $\mathrm{kw}_i \in$ *kw_list*: //iterate though keyword list
    //use eq.(2) to calculate sim_{WuP} between selected synset of seed word
    //and each synset of kw_i
    GET $[\mathrm{sim}(\mathrm{syn}_{seed_i}, \mathrm{syn}_{kw_i}^{l})$ FOREACH $\mathrm{syn}_{kw_i} \in \mathrm{synsets}_{kw_i}]$;
SELECT $\mathrm{syn}_{kw_i}^{l=k}$ IF $\mathrm{sim}(\mathrm{syn}_{seed_i}, \mathrm{syn}_{kw_i}^{k}) = \max([\mathrm{sim}(\mathrm{syn}_{seed_i}, syn_{kw_i}^{l})])$;
IF $\mathrm{sim}(\mathrm{syn}_{seed_i}, \mathrm{syn}_{kw_i}^{k}) = \mathrm{sim}(\mathrm{syn}_{seed_i}, \mathrm{syn}_{kw_i}^{m})$:
 //select synset with higher frequency if similarities are of equal value
  SELECT $\mathrm{syn}_{kw_i}^{k}$ IF $freq(syn_{kw_i}^{k}) > freq(syn_{kw_i}^{m})$;
FOREACH $\mathrm{kw}_i^f \in \mathrm{kws}_i^f$:
  IF $\mathrm{kw}_m^f = \mathrm{kw}_n^f$:
    DELETE kw_n^f; //remove duplicate synset
RETURN kws^f
//END

Figure 4. WordNet-guided Pruning Algorithm

Our method incorporates word sense disambiguation by selecting the most appropriate synset from WordNet given a specific context during the pruning process. We apply the WuP similarity to a pair of synsets, specifically we measure the similarity between the synsets of a seed word and all possible synsets of a candidate word. Based on the results, the synset with the largest similarity is selected. If more than one synset share the same similarity value, the synset with highest frequency is selected. The synset of each seed word is pre-determined. Like other WordNet based similarity measures, WuP similarity is unable to measure the similarity between the synsets across different POSs (e.g. a verb and a noun). We address the limitation by developing a heuristic rule as the following: for a pair of synsets $(w_1.pos_1, w_2.pos_2)$, if w_1 appears in the definition of synset $(w_2.pos_2)$, then synset $(w_1.pos_1)$ is selected; and vice versa.

Experiment

Data Preparation

Within the context of this study, we acquire data from Yelp dataset challenge (cf. https://www.yelp.com/dataset/challenge). The dataset contains over 5 million online reviews of 174,000 business over 11 metropolitan areas from 2003 to 2017. Among different types of businesses, restaurant is the most common domain in the dataset and is thus selected as the target domain in this study. The research sample is created by selecting 10 restaurants in the Las Vegas metropolitan area that have received the most online reviews during the selected time period. The dataset consists of 51,311 online reviews, and 484,367 sentences. Besides the textual contents, each review is also associated with star rating (1 to 5 stars), posting date, and vote count such as the review being 'useful'.

There is a lack of an established lexicon of restaurant aspects that are publicly available. We have to prepare the dataset from scratch. One important task concerns determining a set of aspects of restaurants. Drawing on SemEval 2015 annotation guidelines, we select the following five aspects: *food*, *price*, *service*, *location*, and *general*. We focus on the first four aspects in this study.

To construct the ground truth for keyword lexicon of restaurant review aspects, we follow the procedure discussed in Loughran and McDonald (2011). First, we build CBoW and Skipgram models of online restaurant reviews. Then, we select the most frequent 5,000 words from each of the two models and merge them. Finally, we manually screen and map the selected words to the four aspects. The resulting lexicon contains 486 keywords for food, 251 keywords for price, 316 for service, and 397 for location.

Feature	Seed Keyword Lists
Food	wine, dessert, dish, delicious, sandwich, soda, seafood, taste, sushi, menu
Price	coupon, expensive, pay, cheap, fee, value, cash, cost, overprice, discount
Service	wait, staff, server, friendly, slow, waiter, attentive, attitude, host, rude
Location	clean, neighborhood, décor, spot, atmosphere, furnishing, interior, dim, environment, design

Table 1. Sample Seed Keywords for Restaurant Aspects

Seed Word Selection

The proposed method requires a small set of seed words for each aspect. In view that individual words can be ambiguous, we need to determine the sense(s) for each word. This is accomplished by selecting the appropriate synsets from WordNet, which groups English words into sets of synonyms called synsets (Princeton-University 2012). For instance, the synset corresponding to 'eat'

– a keyword of the 'food' aspect - is 'eat.v.01'. We select 20 seed keywords for each of the four aspects. A list of sample seed keywords is shown in Table 1.

Baseline Methods

To evaluate the performance of our proposed approach to extracting aspects from online consumer reviews, we employ two types of baseline methods: discrete language models and alternative continuous-space language models with different configurations.

We choose Latent Dirichlet Allocation (LDA) as a representative of the discrete language model (EXT_LDA). In this setting, 200 topics are learned from our dataset based on discrete bag-of-words using LDA; and the top 50 keywords are selected for each topic. In addition, we assign each of the top 50 LDA topics to the aspect that has the highest similarity between its seed words and selected keywords for the topic based on the cosine distance function (see eq. (1)).

The two baseline continuous-space language models have the following configurations.

- One baseline model (EXT_CON) extracts aspects by using *contextual* information only in building word embedding models.
- The other baseline model (EXT_POS) incorporates syntactic information (POS tags) into building word embedding models.
- Like our proposed method, both baseline continuous-space language models extract keywords based on a list of seed words using the cosine similarity function. For each seed word, we select 50 words with the highest cosine similarity scores. All methods that are under comparison are summarized in Table 2.

Configuration	Embedding Type	Foundation	Description
EXT_LDA	Discrete Bag-of-Words	Co-occurrence	Using LDA topics to extract keyword based on seed keywords using cosine similarity
EXT_CON	CBoW & skipgram	Contextual	Using continuous word vectors to extract keywords based on the seed word list using cosine similarity
EXT_POS	CBoW & skipgram	Syntactic	Using continuous word vectors and syntactic information to extract keyword based on the seed word list based on cosine similarity
EXT_SEM (Proposed Method)	CBoW & skipgram	Semantic	Using continuous word vectors, syntactic information, and semantic information to extract keyword based on the seed word list based on cosine similarity

Table 2. Model Configurations in This Study

All continuous space language models are implemented via deep neural networks in Python. The window size for all models is set to 5 – which means the context of any target word consists of two words before and after it. We also set the dimensionality of the word vectors to 300. In order to improve the quality of learned word vectors, we also implement techniques such as negative sampling and rare word removal while training the continuous-space language models.

Evaluation Methods and Metrics

We conduct both direct and indirect evaluations of the proposed approach. The direct evaluation involves a comparison of aspect keywords extracted against those of the ground truth. The evaluation metrics include the number of unique words, the ratio of overlapping keywords (keywords appearing in more than one aspects), and accuracy. We define the accuracy (*acc*) of the extracted keywords set (kws_i, in the form of synsets) in eq. (3).

$$acc(kws_i) = \frac{N(kws_i \cap kws_g)}{MAX(N(kws_i), N(kws_g))} \times 100\% \tag{3}$$

In eq. (3), $N(kws_i \cap kws_g)$ is the number of common keywords in extracted keywords set (kws_i) and the ground truth keyword set (kws_g); and $MAX(N(kws_i), N(kws_g))$ is the number of distinct keywords in the long set between the kws_i and kws_g.

The indirect method involves assessing the effects of extracted keywords on the prediction of some target performance measures of online reviews, including *review rating* and *useful votes*. Using predictive models in which mentions of extracted feature keywords is common to evaluate aspect/lexicon extraction in prior related studies (Deng et al. 2017; Siering et al. 2018; Tsai et al. 2016). The two variables are selected because of their practical importance: 1) review rating can serve as a surrogate of business performance, and 2) useful votes can indicate the merit of online reviews to other consumers. Similar variables are used in related previous studies (Floyd et al. 2014; Hong et al. 2017). To build predictive models, we aggregate the occurrences of aspect keywords first at the sentence level and then at the review level. We define a sentence as mentioning an aspect if the sentence contains any of the extracted keywords of the aspect. For instance, sentence 'I would recommend restaurant X because of the wide selection of great cuisine' mentions 'food' aspect because it contains 'selection' and 'cuisine', both of which are keywords of 'food' aspect. The raw counts of mentions are then normalized by the number of sentences in a review.

Unlike regression models used in previous related studies, we choose multivariate time series analysis as the predictive models to address one key feature of online reviews. This is because the measurements of review rating and useful votes (selected target variables) are subject to the influence of posting date and ratings of preceding reviews of the same product/service (Hu et al. 2013). In addition, we implement an advanced time series analysis technique based on deep learning, namely Long Short-Term Memory (LSTM) networks, as the modeling technique. LSTM network is a special type of recurrent neural network, which memorizes the persistent information from one state (e.g. a time spot) to another, and also captures very long dependencies among input signals (Kraus and Feuerriegel 2017). It is a supervised learning approach. Compared to traditional linear models, LSTM networks can model time series with multiple input variables and learn the nonlinear relationship(s) between the inputs and the target variable.

In view of the degree of data sparsity, we build time series models at the weekly level. In other words, the measures of all input and target variables are aggregated by week before building predictive models. To support the model development, we develop a Python program to create a stepwise multivariate dataset. Specifically, we use inputs variables x_i^{t-1} along with y^{t-1} to predict y^t, where x_i^{t-1} is the value of input variable x_i at week (*t-1*); and y^{t-1} and y^t is are the values of a target variable (i.e., star rating or 'useful' votes) at week *t-1* or *t*, respectively.

We split the data into training and testing data at the ratio of 80:20. Specifically, we use the data collected from the first 528 weeks (80%) for training, and the data from the remaining 131 weeks

(20%) for testing. We select Root Mean Squared Error (RMSE) as both the loss function and the validation metric – the model with the lowest loss and RMSE is the best model. RMSE is defined in eq. (4):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad (4)$$

where n is the number of time series in the training/testing sample, y_i is either star rating or 'useful' votes, and $\hat{y}_i$ is the predicted value of y_i.

In this study, all LSTM models for the time series analysis follow the same configuration for the network architecture: one input layer, two LSTM layers (with a dropout layer after each), and a fully connect layer with softmax activation to generate the output (predictions). Each LSTM layer has 64 neurons; and each dropout layer has a dropout rate of 0.5. We also apply L2-regularization to the learned weight vectors. By applying the dropout layers and L2-regularization, we limit the chance of model overfitting. Each model is trained over 50 epochs.

Results and Discussion

Evaluation Results

Table 3 lists the top-10 extracted keywords for word *sushi_NOUN* across all four methods. This sample results demonstrate that the proposed approach yields cleaner and richer results than the baseline methods.

Setting	Top-10 Extracted Keywords
EXT_LDA	curry, chunk, pick, Thai, term, plate, particular, light, pretty, seafood
EXT_CON	ratio, curry, broth, shellfish, spaghetti, varied, station, raw, cold, Alaskan
EXT_POS	Mexican_ADJ, carvery_NOUN, taco_NOUN, crawfish_NOUN, steamed_ADJ, steam_VERB, lamb_NOUN, fresh_ADJ, leg_NOUN, mat_NOUN
EXT_SEM	sashimi_NOUN, paella_NOUN, seafood_NOUN, taco_NOUN, pasta_NOUN, fresh_ADJ, roll_NOUN, dumpling_NOUN, roll_VERB, eel_NOUN

Table 3. Extracted Keywords for Sushi_NOUN across Different Methods

Table 4 presents the results of direct evaluation of all four methods. We make the following main observations from the table.

- EXT_POS produces the most unique words across all four configurations (2,970 unique words, 3,712.5% over the SEED list (which contains 80 terms) across all four features). After investigating randomly selected samples from the results of all four configurations, we discover that this difference is attributed to the different POS tags of the same word.
- Our proposed approach has the lowest overlapping ratio (4.72%) among all four configurations. In other words, the results of EXT_SEM contain the least keywords appearing in more than one aspect. Overlapping aspects within the boundary of the same linguistic unit (e.g. the same sentence) has been an issue in aspect based sentiment analysis research. The results in Table 5 suggest that our proposed approach helps alleviate the overlapping problem.
- We use the ground truth of 1,450 terms to evaluate the accuracy of the term extraction results. The overlapping ratio (i.e. ratio of one term belonging to multiple aspects) in the

ground truth is 9.02%. By using contextual information obtained from the continuous-space language models (EXT_CON), the accuracy of aspect extraction is improved to 51.08% from 43.27% (LDA). Adding syntactic information (i.e. EXT_POS) further increases the accuracy to 62.33%. Finally, by incorporating semantic information (WordNet-based semantic similarity based filtering), our proposed method boosts the accuracy to 73.65%.

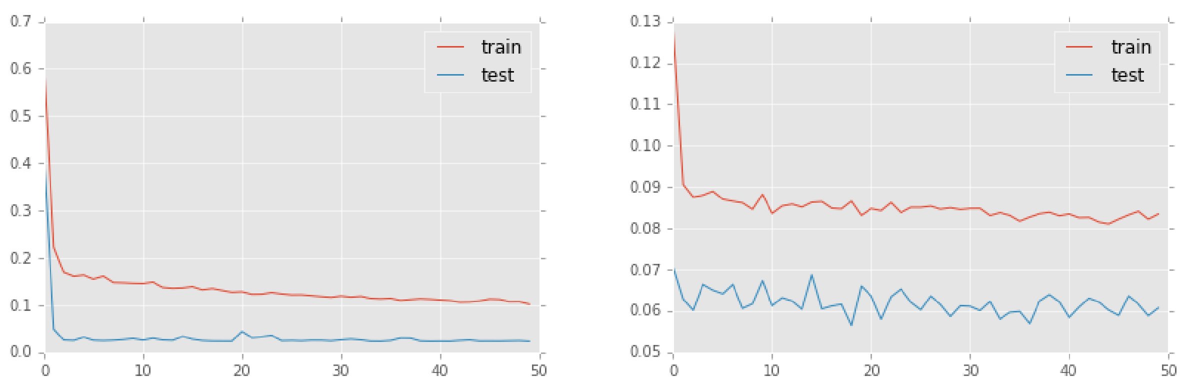
Configuration	# of Unique Words	Overlapping Ratio	Accuracy
EXT_LDA	1,988	23.29%	43.27%
EXT_CON	2,915	17.75%	51.08%
EXT_POS	**2,970**	11.88%	62.33%
EXT_SEM	1,782	**4.72%**	**73.65%**

Table 4. Direct Evaluation Results of All Four Configurations

Target	Star Rating		'useful' Votes	
Configuration	Loss	RMSE	Loss	RMSE
EXT_LDA	0.0236	0.161	0.0549	0.413
EXT_CON	0.0233	0.159	0.0563	0.424
EXT_POS	0.0232	0.157	0.0569	0.412
EXT_SEM	**0.0231**	**0.151**	**0.0539**	**0.412**

Table 5. Indirect Evaluation Results of All Four Configurations

Table 5 reports the results of indirect evaluation, which is the deep learning (i.e. LSTM networks) based time series analysis toward star rating and 'useful' votes. The boldface values in Table 5 denote the best performance among all four configurations. The methods are ranked in a descending order of predictive performance in the following order: EXT_SEM (the proposed approach), EXT_POS, EXT_CON, EXT_LDA. The results suggest that the contextual and syntactic information captured by the continuous-space language models, as well as the semantic information embedded in the proposed approach, all contribute to the prediction performance. Such findings also provide indirect support that our proposed approach is superior in extracting aspects from the review texts. Figure 5 depicts how loss declined while both predictive models are trained.



(a) Star Rating ***(b) 'Useful' Votes***

Figure 5. Training Histories of Predictive Models Built with EXT_SEM

Discussion

In this section, we discuss several insights with respect to aspect extraction from online (restaurant) reviews, which may invite future research.

The results suggest that: 1) the continuous-space language models are more valuable than discrete language models (e.g. LDA on discrete Bag-of-Words); and 2) in addition to capturing contextual and syntactic information in the corpus using the continuous-space language models, applying WordNet-guided pruning as implemented in this study further improves the performances of aspect extraction.

With continuous-space language models, a more specific seed word may lead to better extraction results. For instance, the top-10 extracted keywords for seed word 'wine_NOUN' using the proposed approach are: riesling_NOUN, chardonnay_NOUN, champagne_NOU, bordeaux_NOUN, draft_NOUN, beer_NOUN, sangria_NOUN, cabernet_NOUN, tap_NOUN, and pint_NOUN; while the top-10 extracted keywords for 'food_NOUN' are: slop_NOUN, stuff_NOUN, fare_NOUN, atmosphere_NOUN, product_NOUN, item_NOUN, meal_NOUN, dish_NOUN, ingredient_NOUN, and delicacy_NOUN. In comparison, the list of keywords extracted for wine shows better quality that for food. Thus, we suggest selecting specific rather than general words in engineering seed words for aspect extraction from online consumer reviews,.

In the context of aspect extraction, we discover that the CBoW model leads to better keywords, compared to the corresponding skipgram model. Similar observations were made in the literature (Poria et al. 2016; Tsai et al. 2016). Nevertheless, the results our proposed approach show that the results skipgram models still have values. Our proposed approach offers a pathway to merging results from both types of models.

The performance of proposed approach may benefit from incorporating other textual features such as linguistic patterns (Poria et al. 2016), rule-based methods (Kang and Zhou 2017), and dependency relations (Deng et al. 2015).

It is worth-noting that in this study, we do not differentiate the strength of different keywords for representing an aspect. For instance, even though both 'selection' and 'cuisine' appear in a sentence, we still count it as one mention of the feature 'food'. It might be interesting to conduct a robustness test to verify whether the strength contributes to aspect extraction in future research.

Conclusion

Online consumer reviews contain rich yet implicit information regarding the reviewers' preferences toward the different aspects of products/services. Extracting aspects from online consumer reviews has been recognized as a valuable step in performing different analytical tasks (e.g. aspect-based sentiment analysis). In this study, we propose an approach that leverages semantic information from WordNet to prune keywords of aspects extracted with continuous-space language models. This research makes multifold research contributions. First, the proposed approach proposed is semi-supervised, which requires only a small seed set of aspect keywords. Second, embedding word sense disambiguation into syntactically-enhanced continuous-space language models in extracting aspect keywords, offers a wealth of contextual, syntactic, and semantic information in improving the measurement of relatedness. Third, by applying this approach to social media contents, in particular online consumer reviews, we learn informal expressions of aspects in a certain domain. The proposed approach can also be used to construct a robust domain-specific knowledge structure (i.e. an ontology) from online reviews. We evaluate the proposed approach using a large set of Yelp reviews on restaurants. The extracted aspects are evaluated against a manually engineered ground truth, as well as in time series models predicting review measures that reflect business performance and consumer perceptions. In both evaluation steps, the proposed approach outperforms the baseline models. The evaluation results demonstrate the value of incorporating semantic information in aspect extraction tasks – which facilitates understanding consumers' preferences embedded in online reviews.

ACKNOWLEDGEMENTS

This research is supported in part by the National Science Foundation under Grant No. SES-152768, CNS-1704800. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Archak, N., Ghose, A., and Ipeirotis, P. G. 2011. "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science* (57:8), pp. 1485–1509 (doi: 10.1287/mnsc.1110.1370).

Deng, S., Sinha, A. P., and Zhao, H. 2017. "Adapting sentiment lexicons to domain-specific social media texts," *Decision Support Systems* (94), Elsevier B.V., pp. 65–76 (doi: 10.1016/j.dss.2016.11.001).

Deng, S., Sinha, A., and Zhao, H. 2015. "Resolving Ambiguity in Sentiment Classification: The Role of Dependency Features," *the 25th Workshop on Information Technology and Systems (WITS 2015)* (8:2), p. 4.

Duric, A., and Song, F. 2012. "Feature selection for sentiment analysis based on content and syntax models," *Decision Support Systems* (53:4), Elsevier B.V., pp. 704–711 (doi: 10.1016/j.dss.2012.05.023).

Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., and Freling, T. 2014. "How online product reviews affect retail sales," *Journal of Retailing* (90:2), pp. 217–232 (available at https://ac-els-cdn-com.proxy1.cl.msu.edu/S0022435914000293/1-s2.0-S0022435914000293-main.pdf?_tid=4e574d8a-ab0c-11e7-b192-00000aacb362&acdnat=1507345680_f0b31dcb92677431117061b7ba7991e0).

Ge, J., and Qiu, Y. 2008. "Concept Similarity Matching Based on Semantic Distance," in *2008 Fourth International Conference on Semantics, Knowledge and Grid*, IEEE, December, pp. 380–383 (doi: 10.1109/SKG.2008.24).

Hong, H., Xu, D., Wang, G. A., and Fan, W. 2017. "Understanding the determinants of online review helpfulness: A meta-analytic investigation," *Decision Support Systems* (102), pp. 1–11 (doi: 10.1016/j.dss.2017.06.007).

Hu, H.-W., Chen, Y.-L., and Hsu, P.-T. 2016. "a Novel Approach To Rate and Summarize Online Reviews According To User-Specified Aspects," *Journal of Electronic Commerce Research* (17:2), pp. 132–152.

Hu, N., Koh, N. S., and Reddy, S. K. 2013. "Ratings Lead You To The Product , Reviews Help You Clinch It : The Dynamics and Impact of Online Review Sentiments on Products Sales The Mediating Role of Online Review Sentiments on Product Sales," *Decision Support Systems* (57), pp. 42–53.

Kang, Y., and Zhou, L. 2017. "RubE: Rule-based methods for extracting product features from online consumer reviews," *Information and Management* (54:2), Elsevier B.V., pp. 166–176 (doi: 10.1016/j.im.2016.05.007).

Kraus, M., and Feuerriegel, S. 2017. "Decision support from financial disclosures with deep neural networks and transfer learning," *Decision Support Systems* (104), Elsevier B.V., pp. 38–48 (doi: 10.1016/j.dss.2017.10.001).

Lau, R. Y. K., Li, C., and Liao, S. S. Y. 2014. "Social analytics: Learning fuzzy product ontologies for aspect-oriented sentiment analysis," *Decision Support Systems*, Elsevier B.V. (doi: 10.1016/j.dss.2014.05.005).

Leacock, C., and Chodorow, M. 1998. "Combining local context and WordNet similarity for word sense identification," in *WordNet: An electronic lexical database*, MIT press, pp. 265–283.

Li, C., Duan, Y., Wang, H., Zhang, Z., Sun, A., and Ma, Z. 2017. "Enhancing Topic Modeling for Short Texts with Auxiliary Word Embeddings," *ACM Transactions on Information Systems* (36:2), pp. 1–30 (doi: 10.1145/3091108).

Li, Y., Wei, B., Liu, Y., Yao, L., Chen, H., Yu, J., and Zhu, W. 2017. "Incorporating Knowledge into neural network for text representation," *Expert Systems with Applications* (96), Elsevier Ltd, pp. 103–114 (doi: 10.1016/j.eswa.2017.11.037).

Lin, D. 1998. "An information-theoretic definition of similarity," in *Proceeding ICML '98 Proceedings of the*

Fifteenth International Conference on Machine Learning, pp. 296–304 (available at http://webdocs.cs.ualberta.ca/~lindek/papers/sim.pdf).
Loughran, T., and Mcdonald, B. 2011. "When is a Liability not a Liability ? Textual Analysis , Dictionaries , and 10-Ks," *Journal of Finance* (66:1), pp. 35–65.
Meijer, K., Frasincar, F., and Hogenboom, F. 2014. "A semantic approach for extracting domain taxonomies from text," *Decision Support Systems* (62), Elsevier B.V., pp. 78–93 (doi: 10.1016/j.dss.2014.03.006).
Mikolov, T., Chen, K., Corrado, G., and Dean, J. 2013. "Efficient Estimation of Word Representations in Vector Space," (doi: 10.1162/153244303322533223).
Navigli, R. 2009. "Word sense disambiguation: A Survey," *ACM Computing Surveys* (41:2), pp. 1–69 (doi: 10.1145/1459352.1459355).
Peñalver-Martinez, I., Garcia-Sanchez, F., Valencia-Garcia, R., Rodríguez-García, M. Á., Moreno, V., Fraga, A., and Sánchez-Cervantes, J. L. 2014. "Feature-based opinion mining through ontologies," *Expert Systems with Applications* (41:13), pp. 5995–6008 (doi: 10.1016/j.eswa.2014.03.022).
Poria, S., Cambria, E., and Gelbukh, A. 2016. "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowledge-Based Systems* (108), Elsevier B.V., pp. 42–49 (doi: 10.1016/j.knosys.2016.06.009).
Princeton-University. 2012. "About WordNet.," (available at http://wordnet.princeton.edu/).
Qiao, Z., Zhang, X., Zhou, M., Wang, A., and Fan, W. 2017. "A Domain Oriented LDA Model for Mining Product Defects from Online Customer Reviews," in *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 1821–1830.
Resnik, P. 1995. "Using Information Content to Evaluate Semantic Similarity in a Taxonomy," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (Vol. 1).
Sánchez, D., Batet, M., Isern, D., and Valls, A. 2012. "Ontology-based semantic similarity: A new feature-based approach," *Expert Systems with Applications* (39:9), pp. 7718–7728 (doi: 10.1016/j.eswa.2012.01.082).
Siering, M., Deokar, A. V., and Janze, C. 2018. "Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews," *Decision Support Systems* (107), Elsevier B.V., pp. 52–63 (doi: 10.1016/j.dss.2018.01.002).
Tao, J., El-gayar, O. F., Deokar, A. V., and Chang, Y. 2015. "Term Extraction and Disambiguation for Semantic Knowledge Enrichment : A Case Study on Initial Public Offering (IPO) Prospectus Corpus," in *Proceedings of HICSS 48' -- 2015 48th Hawaii International Conference on System Sciences*, pp. 3719–3728 (doi: 10.1109/HICSS.2015.448).
Tsai, M.-F., Wang, C.-J., and Chien, P.-C. 2016. "Discovering Finance Keywords via Continuous-Space Language Models," *ACM Transactions on Management Information Systems* (7:3), pp. 1–17 (doi: 10.1145/2948072).
Wu, Z., and Palmer, M. 1994. "Verbs semantics and lexical selection," in *Proceedings of the 32nd annual meeting on Association for Computational Linguistics (ACL '94)*, pp. 133–138 (available at http://dl.acm.org/citation.cfm?id=981751).