Sentiment Augmented Attention Network for Cantonese Restaurant Review Analysis

Rong Xiang
csrxiang@comp.polyu.edu.hk
The Hong Kong Polytechnic University
Hong Kong

Ying Jiao
naomi.jiao@connect.polyu.edu.hk
The Hong Kong Polytechnic University
Hong Kong

Qin Lu
csluqin@comp.polyu.edu.hk
The Hong Kong Polytechnic University
Hong Kong

ABSTRACT

Online reviews written in Cantonese style are widely utilized by native Cantonese speakers and a large amount of Cantonese reviews are available on the Internet. However, only few studies on Cantonese sentiment analysis are reported as there is a serious lack of resources including annotated corpora and adequate lexical collections. In this work, we present a novel approach for sentiment analysis of Cantonese style text by incorporating sentiment knowledge into the attention mechanism in the state-of-the-art deep learning based Long Short-Term Memory network, referred to as the sentiment augmented attention (LSTM-SAT). A restaurant review dataset is first collected from Openrice, a popular restaurant review website mostly written in Cantonese style with naturally annotated rating labels. We then extract a Cantonese sentiment lexicon based on an automatic construction method to obtain both the sentiment terms and their polarities using sentiment scores of the review text. The automatically obtained terms can then be used to augment a manually obtained small Cantonese sentiment lexicon. Furthermore, we propose a novel method to incorporate lexical knowledge in the sentiment lexicon to the attention layer as the prior knowledge in an LSTM model to further highlight the importance of sentiment words. Experimental results show that our automatically constructed Cantonese sentiment lexicon helps improve coverage and this type of sentiment knowledge can be a semantically meaningful information in deep learning models. This information indeed serves as effective information as our proposed LSTM-SAT shows a significant improvement on the performance of sentiment classification.

KEYWORDS

Cantonese sentiment analysis, restaurant review analysis, neural networks, attention mechanism

ACM Reference Format:
Rong Xiang, Ying Jiao, and Qin Lu. 2019. Sentiment Augmented Attention Network for Cantonese Restaurant Review Analysis. In *Proceedings of WISDOM '19: Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM '19).* ACM, New York, NY, USA, 9 pages.

WISDOM '19, August 4, 2019, Anchorage, Alaska

1 INTRODUCTION

Online consumer shopping behavior had changed dramatically in the era of Web 2.0. The massive volume of online customer reviews (OCRs) provides a wealth of information about customer interests and opinions. The importance of review text analysis is becoming increasingly significant [8]. Due to the interactivity and publicity, OCRs not only strongly influence users' purchase decisions[7], but also present an informative and valuable feedback channel for business managers to evaluate and improve product quality[6]. However, it is usually very time-consuming for both end-users and entrepreneurs to grasp sentiment information over different products because the scale of OCRs keeps increasing on the blooming electronic shopping platforms. Sentiment analysis (SA), therefore, can alleviate the burden and provides a feasible and valuable way to automatically scan through the reviews and help classify them into different sentiment polarities with strength indications.

Cantonese is a Chinese dialect widely used by over 70 million people in southeastern of Mainland China, Hong Kong SAR and overseas Chinese communities[20]. Snow indicated that native Cantonese speakers usually prefer to write comments in Cantonese in informal settings especially the Internet forums[2]. However, due to its dialect specific vocabulary and the use of unique Cantonese characters, it is generally quite difficult for other Chinese readers to understand text written in Cantonese in casual occasions and social media platforms. Given the popularity of Cantonese writings in online social media, which we refer to as **Cantonese style text** (**Cantonese text** for short), Cantonese sentiment analysis is in urgent need if businesses want to overcome the communication gap using sentiment classification techniques. This will also enable non-Cantonese speakers to capture the sentiment expressed in Cantonese review text.

Compared to English and other Asian languages such as Mandarin and Japanese, Cantonese sentiment analysis is much less investigated [14, 20, 21]. The first and foremost problematic issue is the lack of benchmark data. Datasets for Cantonese sentiment classification can hardly be found from previous work. Another issue is the lack of adequate Natural Language Processing (NLP) resources for Cantonese such as Cantonese specific lexicons for general use and Cantonese sentiment lexicons for sentiment analysis. The lack of necessary language related resources, therefore, makes NLP methods less robust in providing analytically informative syntax and semantics capabilities.

As one type of language resources, sentiment lexicon refers to a collection of lexicon units constructed with labels associated with sentiment orientations. Cantonese sentiment words such as "好味"

(*tasty*), "唔錯" (*not bad*) and "唔得" (*not good*) are Cantonese unique lexical units which have strong sentiment evidence and acquired knowledge on these lexical terms can facilitate Cantonese sentiment classification tasks. The use of appropriate lexicons has proven to be useful for sentiment classification tasksmedhat2014sentiment, despite the scarcity of Cantonese sentiment lexicons.

Motivated by the crucial need on sentiment analysis for Cantonese text and the extremely inadequate availability of language resources for Cantonese sentiment analysis, this work focuses on exploring the use of online restaurant review data available in Hong Kong and Guangdong Province. Openrice[1], a famous catering website with over one million registered restaurants and over four million users, is used as the review text source to be studied for this sentiment analysis task. We first collect a set of review comments which naturally contains rating labels from 1-star to 5-star of the restaurants. A Cantonese sentiment lexicon is then extracted using an automatic method to augment a small manually obtained Cantonese sentiment lexicon. The automatic method makes use of naturally annotated text labels to predict the sentiment scores of the extracted lexical terms. A Cantonese dataset as well as the enlarged sentiment lexicon can both contribute as Cantonese language resources not only for Cantonese sentiment classification but also for other NLP related tasks including Cantonese text. We also propose a novel method to incorporate the acquired Cantonese lexical resource into the current state-of-the-art attention-based deep learning neural network. The proposed method, referred to as LSTM-SAT, is implemented on the Long Short-Term Memory network with the sentiment augmented attention for Cantonese sentiment classification. Evaluation results show that our proposed sentiment lexicon can be well augmented in LSTM-SAT, resulting a 1.9% improvement of Cantonese sentiment classification task.

The rest of this paper is organized as follows. Section 2 introduces related work. Section 3 explains details of sentiment lexicon construction and LSTM-SAT structure. Section 4 gives performance evaluation. Section 5 concludes our work with directions for future work.

2 RELATED WORK

Despite the widely used Cantonese style writing in the social media, only a limited number of Cantonese sentiment classification studies were proposed. The investigated languages in previous studies for sentiment analysis tasks mainly focus on English, Mandarin and Japanese[3, 9, 14, 20]. Although the models of these studies are also applicable for Cantonese sentiment classification, the problematic issues attribute to ensure syntax and semantic techniques well-matched with Cantonese characteristics. We present relevant studies in two parts.

2.1 NLP in Cantonese

Many previous studies investigated Cantonese language processing problems by contrasting with Mandarin. As a variant of Chinese, Cai indicated that the phenomenon of language nodes which passed activation to all the lemmas was common in both Cantonese and Mandarin[3]. A linguistic comparative study was provided by Gang from the aspect of syllables. It is pointed out that Cantonese syllables exhibit less variation than Mandarin[21]. Lee also compared Cantonese and Mandarin in several linguistic differences including grammar, register coverage, plural marker for personal nouns, agentless passive and possessive constructions[14]. Based on the evaluation, it clearly indicates that the NLP techniques should be well-established for Cantonese.

Research conducted by Chen et al. contributed to building up the Part-of-Speech (POS) tagging system and the sentiment word dictionary for Cantonese[5]. A Cantonese POS tagging system was trained with the hierarchical hidden Markov model (HHMM) algorithm proposed by Zhang et al. with Cantonese corpora. The method for Cantonese sentiment word dictionary construction is to use simple grammar rules to extend the manually selected opinion words[30]. However, no down-steam task was conducted to demonstrate the effectiveness for POS tagging system nor the sentiment word dictionary. Beside POS tagging, Fung [9] showed a dictionary-based word segmentation method with the advantage of automatically identifying Cantonese words in spoken Cantonese texts.

Zhang et al. incorporated Naïve Bayes and Support Vector Machines into sentiment classification of Cantonese restaurant reviews collected from a famous Hong Kong restaurant review website Openrice[31]. They manually labeled sentiment polarity of the reviews and selected 1,800 reviews (900 positive and 900 negative) as their dataset. A 93% accuracy is obtained for well-constructed features including unigrams, bigrams and trigrams in their classifiers. Considering the small volume of their study, the 1800 instance dataset is not suitable for sentiment classification benchmarks since many models will be risky of over-fitting problem. A combined approach for the multiple-domain Cantonese sentiment classification was introduced by Ngai et al.[20]. They combined three popular classifiers, including Support Vector Machine with Naïve Bayes features (NBSVM)-based, convolutional neutral network (CNN)-based, and lexicon-based classifiers. Sentiment words were manually constructed for the lexicon-based classifier. Although a small coverage in sentences attributed by the sentiment lexicon, this method can result in a remarkable improvement in three experimental datasets, Weibo, Facebook and Dianping.

2.2 Sentiment Analysis

Recently, neural network-based methods have greatly improved the performance of sentiment analysis. Typical models include Convolutional Neural Networks (CNN)[23], Recursive Neural Network (ReNN) [24], and Recurrent Neural Networks (RNN)[11]Long-Short Term Memory model (LSTM) is introduced by adding a gated mechanism to keep long term memory [26] which is more suited for text understanding. Attention based LSTM (LSTM-AT) is proposed to highlight semantically important words and sentences [29]. Most attention models are built from local context. Other methods build attention models using external knowledge, such as user/product information[4] and cognition grounded data[15, 16].

In the aspect of lexicon-based approach, sentiment lexicons are usually utilized as dictionaries of opinion words with sentiment labels to identify sentiment of text[25]. To collect sentiment lexicons, manual and/or automatic methods can be applied. VADER

[1]https://www.openrice.com/zh/hongkong

was proposed by Hutto and Gilbert which generated reliable lexical sentiment features and then combined those features to five rules to determine the sentiment of documents[10]. The Sentiment Orientation CALculator (SO-CAL) system can help to produce the sentiment lexicon manually with high-accuracy, yet low-coverage[17]. Khan et al. use an automatic method to produce the SentiWordnet[13].

In previous studies, there are two main directions to incorporate lexicon-based methods and machine learning approaches. One the one hand, two weighted classifiers can be linearly integrated into one system. Sentiment classifications mainly use the lexicon-based approach or the machine learning approach as well as the hybrid approach using the combination of the two approaches [18]. Andreevskaia and Bergler [1] presented a system consisting of the ensemble of two classifiers with precision-based weighting which obtains significant gains in accuracy and recall over corpus-based classifiers and lexicon-based systems taken individually. On the other hand, it is also feasible to incorporate lexicon knowledge into learning algorithms. Sentiment lexicon can be used directly in a rule-based approach [10] which indicates the sentiment degree. Wilson et al. [28] and Melville et al.[19] demonstrate sentiment dictionary can improve linear classifier-based classifiers. Jovanoski et al. [12] also indicate that logistic regression models can be benefited from sentiment lexicon. Sentiment lexicon contributes to neural network models as well. A remarkable task utilizing sentiment lexicons is done by Teng et al. [27], which assigns sentiment score of a sentence by using a weighted sum of prior sentiment scores of negation words and sentiment words to further improve the performance of neural networks. Qian et al. [22] propose to apply three linguistic regularizers to sentiment classification. Zou et al. [32] adopt a mixed attention mechanism to further highlight identified sentiment lexicon in attention layer.

3 DESIGN PRINCIPLES

In this work, we first obtain a Cantonese review dataset from Openrice. Based on this corpus, a Cantonese sentiment lexicon is extracted with corresponding sentiment score. These sentiment terms along with their scores are then incorporated to the attention layer of an LSTM model to highlight the importance of sentiment words.

3.1 Acquisition of Sentiment Lexicon

To collect data from the Internet, python crawlers are used to obtain Cantonese restaurant reviews from the dining website Openrice which allows users to post their ratings of restaurants as well as review comments. Openrice is originated in Hong Kong. The majority of Openrice users are native Cantonese speakers and the reviews are generally written in Cantonese style. In this work, 250,805 Cantonese style reviews are collected. Users must give ratings to the restaurant when writing their review. Thus, ratings can be used as naturally annotated labels for the review text. User ratings are integers from 1 (very negative) to 5 (very positive).

In our research group, we have a manually constructed small sentiment lexicon with over two thousand lexicon terms produced by native speakers. Even though manual construction generally can obtain more accurate lexical knowledge, the size of this lexicon is so small that it would have coverage problem for any machine learning algorithm. Therefore, we need to augment it. As restaurant reviews are more or less domain specific, we choose to augment the lexicon using a data driven approach to first obtain lexical candidates and then assign sentiment scores to certain types of these candidates to obtain the sentiment lexicon.

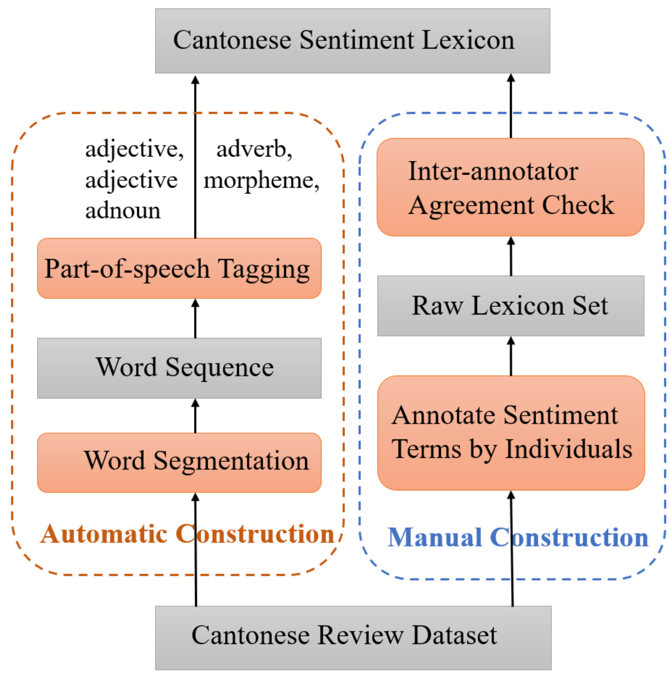

Figure 1: Acquisition of Cantonese Sentiment Lexicon

Sentiment Lexicon Extraction Figure 1 shows the general framework of our method for obtaining the Cantonese sentiment lexicon. The manually constructed Cantonese sentiment lexicon was obtained from fifty native speakers who identified Cantonese sentiment terms by marking words from general domain text. For quality control, each text was read by a number of persons for cross validation. Sentiment terms including words, phrases or clauses were annotated by at least two people reading the same text material. The Manual Sentiment Lexicon (MSL) contains 2,308 terms.

To enlarge the volume of the sentiment lexicon, an automatic collecting algorithm is proposed using the Chinese word segmentation and POS tagging tool Jieba[2]. Documents in Openrice reviews are first segmented to word sequences. After getting the POS tags, the words marked as "a" (adjective), ad (adverb), "ag" (adjective morpheme), and "an" (adnoun) are selected as sentiment lexicon candidates as they are descriptors and more potentially sentiment-related. 15,926 candidates are collected and referred to as the Auto-Sentiment Lexicon (ASL). After removing 194 overlapped terms in both MSL and ASL, 18,040 Cantonese terms are integrated into the combined sentiment lexicon (COMB).

Although the Chinese segmenter Jieba has the capability to provide syntactic information, no segmenter can identify all out of

[2]https://github.com/fxsjy/jieba

vocabulary words. Therefore, we are aware that some Cantonese specific words may not be identified using this automatic approach. Nonetheless, this method can still help to identify more sentiment terms and experiments show that lexicon augmented by this method can improve performance, which will be clarified later.

Sentiment Scores For each extracted sentiment term, the corresponding numerical value of the sentiment needs to be determined. Let t denote an extracted sentiment term, the sentiment score, denoted as $S(t)$, needs to be either predicted or estimated. In this work, we propose to obtain $S(t)$ based on the assumption that positive words appear more frequently in positive reviews, and vice versa. As our review text is linked to restaurant ratings which are mandatory when review text is submitted, we make use of review text and the rating to obtain the predicted sentiment score of terms.

Given the frequency of t in the reviews rated $i(i \in [1, 5])$, denoted as $F_i(t)$, the coarse sentiment score $R(t)$ can be computed with a weighted mean as defined below:

$$R(t) = \frac{\sum i * F_i(t)}{\sum F_i(t)}. \tag{1}$$

The distribution of rating classes is usually unbalanced and this is also true for our data. To handle the imbalance issue properly, the sentiment scores should be re-weighted with respect to the dataset. A linear transformation is conducted to adjust the $R(t)$ to obtain the sentiment score $S(t)$ of t as defined below:

$$S(t) = \alpha * R(t) + \beta, \tag{2}$$

where the balance factor α is introduced to drive an adjusted score of $R(t)$. β is a constant value used to shift the distribution.

3.2 LSTM-SAT

Based on the sentiment lexicon and corresponding sentiment scores collected, this section elaborates the mechanism of our proposed Sentiment Augmented Attention Network under the deep learning model of LSTM. LSTM-based models currently outperform most of the other deep learning algorithms in sentiment classification tasks. Based on LSTM, attention mechanism based on local context information is proposed to give more emphasis to words which have more semantically relevant information in a sentence, which has achieved state-of-the-art performance[29]. However, attention mechanism is not generally considered to have the ability to detect sentiment affiliation of words. In this work, we propose to use a similar mechanism to grant LSTM models the ability to also highlight the importance of sentiment terms. LSTM-SAT is proposed to benefit from both analytic power of semantic information in its context and the prior lexical sentiment knowledge. The injected additional sentiment knowledge as well as the current attention mechanism can make the deep learning process more interpretable.

In sentiment classification, given D as the collection of n instances. Each instance d_i ($i \in 1, 2, ..., n$) is one piece of review text document with a corresponding rating label y_i. In sentiment analysis, the label typically indicates a binary class or a numerical value to indicate the sentiment polarity. The learning objective is to train a model for correctly mapping a d_i to y_i.

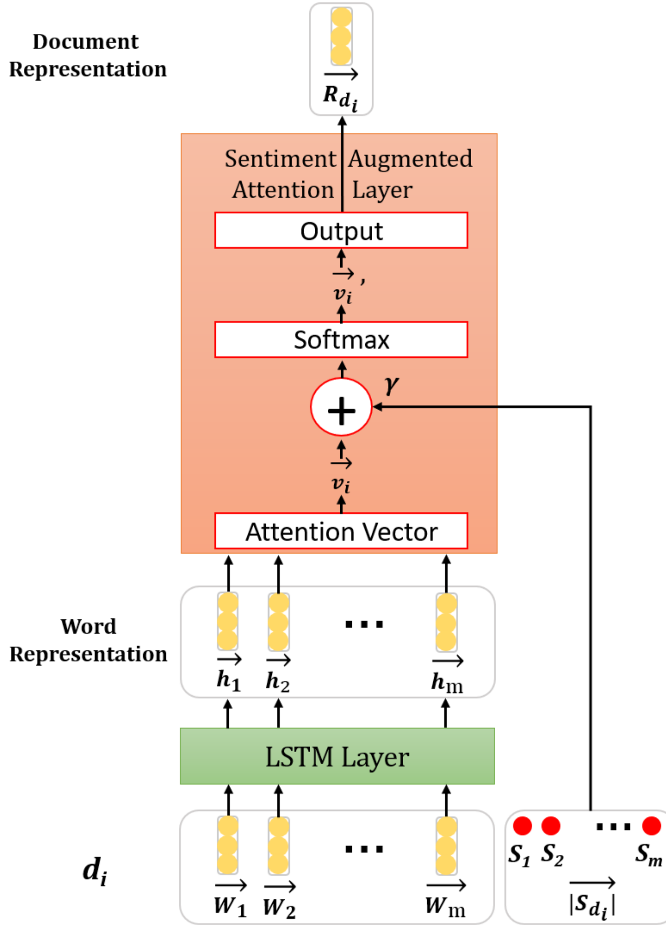

Figure 2: Framework of LSTM-SAT

In the pre-processing phase of Cantonese text, each review document d_i is first segmented to a sequence of tokens, denoted as w_j ($j \in 1, 2, ..., m$). In deep learning models using word embedding as representation, an embedding layer is added to map each token to a vector, denoted as $\overrightarrow{w_j}$. Meanwhile, another sentiment vector $\overrightarrow{S_{d_i}}$ for d_i with the same length of token sequence is constructed as follows. For each w_j, if w_j is in the given sentiment lexicon $SLEX$, $w_j \in SLEX$, its corresponding sentiment score is used for s_j in $\overrightarrow{S_{d_i}}$. For those w_j which are not in $SLEX$, their s_j values are assigned to zero, indicating that there is no prior sentiment knowledge for these words.

The framework of the proposed LSTM-SAT is depicted in Figure 2. Similar to a typical learning process of an LSTM model, the word embedding vectors $\overrightarrow{w_j}$ are trained through the LSTM layer to obtain the hidden sequential representative vectors $\overrightarrow{h}_1, \overrightarrow{h}_2 ... \overrightarrow{h}_m$. In our framework, a similar attention mechanism proposed by [29] is included. Semantically differential contribution of words is highlighted and results in $\overrightarrow{v_j}$. Our LSTM-SAT further incorporates S_{d_i} into the attention layer with a sentiment influence factor γ to produce a sentiment augmented attention layer before generating the document representation.

In our proposed LSTM-SAT, a coefficient vector $\overrightarrow{U}$ is introduced as an informative representation of the words in a network memory, evaluating the significance of each word w_j. The representation of a word $\overrightarrow{h_j}$ and the corresponding word-level context vector $\overrightarrow{U}$ is integrated to obtain an attention weight v_j as follows:

$$v_j = \frac{\exp(\overrightarrow{U} \cdot \overrightarrow{h_j})}{\sum_j \exp(\overrightarrow{U} \cdot \overrightarrow{h_j})}. \tag{3}$$

The lexicon-based sentiment vector $\overrightarrow{S_{d_i}}$ is then incorporated to the attention layer. Let $\overrightarrow{v_i}'$ denote the updated $\overrightarrow{v_i}$ which consists of a series of v'_j ($j \in 1, 2, ..., m$). The updating process can be computed according to the following formula:

$$\overrightarrow{v_i}' = softmax(\overrightarrow{v_i} + \gamma * |\overrightarrow{S_{d_i}}|). \tag{4}$$

Note that sentiment score is element-wisely added to the attention weight and γ is an algorithm parameter to control how much sentiment should influence the adjustment of attention weight in the LSTM model. when $\gamma = 0$, the LSTM-SAT deteriorates to LSTM-AT, removing all sentiment knowledge for adjustment.

The weights in attention layer are normally considered to indicate the intensity which should be a non-negative value. However, sentiment is normally represented by polarity which can be either positive or negative. To make the formula work, the numerical polarity vector $\overrightarrow{S_{d_i}}$ needs to be transformed into a degree vector of intensity $|\overrightarrow{S_{d_i}}|$.

Then Softmax function is applied to transform values in the updated attention matrix to make the sum of $\overrightarrow{v_i}'$ to be 1. Result of the Softmax is used as the weight for word representation to calculate the weighted average of word representation to obtain document representation. The updated document representation $\overrightarrow{R_{d_i}}$ of d_i can be generated as a weighted sum of the word vectors given below.

$$\overrightarrow{R_{d_i}} = \sum_j (v'_j * \overrightarrow{h}_j) \tag{5}$$

The generated $\overrightarrow{R_{d_i}}$ contains both semantic information and sentimental prior knowledge which can be used by deep learning based classifier.

3.3 Objective Functions

Since the classes in Openrice are numerical values, the final output is the regression value of the training model. After generating $\overrightarrow{R_{d_i}}$, a fully connected layer is used to calculate the predicted value p_i of i-th document according to the following formula:

$$p_i = softmax(\mathbf{W} * \overrightarrow{R_{d_i}} + \overrightarrow{b}), \tag{6}$$

in which, $\mathbf{W}$ and $\overrightarrow{b}$ are the parameter matrix and vector in the fully connected layer.

Then the objective of sentiment analysis is to optimize L according to Formula 7. RMSE (root-mean-square error) is used to measure the difference between y_i and p_i as shown below.

$$L = \sqrt{\sum (y_i - p_i)^2} \tag{7}$$

4 PERFORMANCE EVALUATION

To evaluate the effectiveness of our lexicons MSL, ASL, COMB as well as the sentiment augmented attention network LSTM-SAT, the dataset obtained from Openrice is used for validation. Experiments are conducted on fine tuning of sentiment scores, the effect of different sentiment lexicons, as well as LSTM-SAT to be compared to baseline models. Several sample cases are also provided to analyze the impact of introducing sentiment lexicons as prior knowledge in real data.

4.1 Dataset

Stratified sampling is used to selected 60,000 review text and their corresponding rating labels from the complete dataset. In this dataset, comments using Mandarin, English and other languages are excluded. The length of comment is limited to 250 characters. After stochastic shuffling, 90% of dataset is used as the training set, while the other 10% reviews are used as the testing set. The longest review has a length of 249 characters while the shortest one has 4 characters. Table 1 describes the dataset for each rating class. It is easy to see from Table 1 that the rating distribution is quite skewed where the majority of ratings are 4 with 48.2% and the smallest proportion is only 2.3%, less than 1/20 of the majority. However, in terms of review text length, it is quite evenly. In other words, people tend to justify equally for any rating they give. This is good as the amount of information expressed in the reviews of all five classes should be similar.

Rating	Proportion	Average Length	Median Length
1	2.3%	173.6	184
2	8.5%	169.8	174
3	27.6%	170.8	175
4	48.2%	175.1	181
5	13.4%	178.1	185
Total	100.0%	173.8	179

Table 1: Features of Each Rating Class

4.2 Sentiment Analysis

To assess the performance of our proposed LSTM-SAT, sentiment classification task is conducted. The six baseline models are used for comparison to LSTM-SAT. All of the models are listed below. In the implementation, every model is tuned for their respective parameters based on the dataset.

- **NB**: Naive Bayes is the basic model using bag-of-word representation as the feature vector. Multinomial naive Bayes is adopted for multi-label classifier. The parameter *max_features*, a threshold value to limit the vocabulary to terms of certain frequency, is tuned to give the best result for this dataset.
- **LR**: Logistic Regression uses the mean of word embedding to generate sentence representation. Since the dataset is

unbalanced, the parameter *class_weight* is fed to the classifier according to Table 1.

- **SVM**: Support Vector Machine introduces a kernel function to the classifier that uses a sentence feature vector. The mean of word embedding is utilized to generate the sentence representation. The kernel function used here is *'rbf'*.
- **CNN**: Convolutional Neural Network uses a convolution layer to capture features of adjacent words. Each word is mapped to a 300 dimension vector. The final sentiment label is classified with a perceptron.
- **LSTM**: LSTM is a typical RNN architecture with a gated mechanism. LSTMs were developed to handle exploding and vanishing gradient problems when training traditional RNNs. The hyper-parameters *learning rate* and *dropout rate* are the most important parameters to affect the performance of the model.
- **LSTM-AT**: LSTM-AT uses LSTM with attention mechanism to re-weight important words before the fully connected layer.
- **LSTM-SAT**: LSTM-SAT is our proposed model. Based on the attention mechanism, MSL, ASL and COMB are separately used in the attention layer to highlight the identified sentiment terms. The LSTM-SAT model with the three respective sentiment lexicons are denoted as LSTM-SAT (MSL), LSTM-SAT (ASL), and LSTM-SAT (COMB), respectively. Considering the distribution of Openrice classes, the parameter α and β introduced in Formula 2 are experimentally adjusted to be optimized to 1.0 and -3.5 so that the sentiment score can be equally distributed.

The performance of sentiment analysis is measured by accuracy and RMSE described in Formula 7 and 8. To calculate accuracy, we use the following notations: for each class k, TP_k = True positive; FP_k = False positive; TN_k = True negative; FN_k = False negative.

$$accuracy = \frac{\sum (TP_k + TN_k)}{\sum (TP_k + TN_k + FP_k + FN_k)} \tag{8}$$

Algorithm	Accuracy	RMSE
NB	51.5%	NA
LR	53.1%	NA
SVM	56.9%	NA
CNN	57.4%	0.341
LSTM	57.9%	0.352
LSTM-AT	58.9%	0.342
LSTM-SAT (MSL)	60.2%	<u>0.334</u>
LSTM-SAT (ASL)	<u>60.5</u>%	0.337
LSTM-SAT (COMB)	**60.8**%	**0.328**

Table 2: Performance of Sentiment Analysis; Overall best is marked bold; second best is underlined

Table 2 shows the performance of the six baselines and three variants of our proposed model in terms of accuracy and RMSE. It is obvious that the accuracy results of all deep learning models outperform other machine learning models. This clearly shows that deep learning models have better performance because the embedding based document representation can capture more features of the text. Among all the deep learning models which do not use explicit sentiment knowledge, the LSTM-AT model is the best performer. Its gain in performance compared to CNN as well as LSTM as the attention mechanism is an important mechanism to give more emphasis on semantically more meaningful words. When sentiment lexicons are incorporated into LSTM-AT, the performance of our proposed LSTM-SAT shows a significant improvement. LSTM-SAT (ASL) using our automatically acquired sentiment lexicon results in 0.3% higher accuracy than LSTM-SAT (MSL) using manually constructed sentiment lexicon. This indicates that without using expensive acquired manual methods, our automatic sentiment lexicon construction approach can provide effective prior knowledge for LSTM-AT. Benefited from the union of ASL and MSL, LSTM-SAT (COMB) has the best performance to obtain an increase of accuracy by LSTM-AT at 1.9%.

4.3 Impact of Different Sentiment Lexicons

This section examines the impact of the sentiment lexicons from two perspectives. Firstly, since LSTM-SAT is based on the idea of adding sentiment knowledge into the attention layer, it is important to find out what would be an appropriate value for the sentiment influence factor γin Formula 4. Secondly, the details of the performance differences with different sentiment lexicon acquired are carefully compared. We are particularly interested in the quality of our automatically acquired sentiment lexicon.

Table 3 shows a set of experiments conducted for the three different lexicons, LSTM-SAT (MSL), LSTM-SAT (ASL) and LSTM-SAT(COMB), respectively. The LSTM model parameter settings are tuned at the learning rate = 0.001 and the dropout rate = 0.05. The range of γ in the experiment is from 0.1 to 0.7.

Table 3 shows that accuracy reaches the peak when $\gamma = 0.4$ for all three lexicons, MSL and ASL as well as COMB. When γ equals to 0.3 and 0.5, the performance is close. However, accuracy decreases significantly when γ moves away from 0.4 in both directions. This indicates that there indeed exists an optimal point for the sentiment influence factor. From Formula 4, we can see that semantic information and sentiment information are being combined linearly in the attention layer where γ serves as a coefficient to weigh in the sentiment score. When sentiment influence factor is around 0.4, the semantic information highlighted by the attention mechanism strikes a balance with the augmented sentiment knowledge.

In general, attention mechanism provides an extra layer to include more knowledge in deep learning models. The original attention model includes semantic information which is in principle implicit knowledge as most of the attention models are statistic based using local context. Here, we propose a method to incorporate explicit sentiment knowledge into deep learning models and the knowledge has good theoretic basis. In other words, we are able to demonstrate how an explainable knowledge can be incorporated into deep learning models.

Observed from Table 3, the consistency among three sentiment lexicons is quite good. LSTM-SAT (COMB) outperforms the other two methods in every γ setting which is expected. When LSTM-SAT (MSL) and LSTM-SAT (ASL) are compared, we can see that in general, ASL gives better accuracy, but worse RMSE compared

Sentiment Lexicon Set	y	accuracy	RMSE
MSL	0.1	59.2%	0.341
MSL	0.2	59.7%	0.339
MSL	0.3	59.6%	0.341
MSL	0.4	**60.2%**	**0.334**
MSL	0.5	59.8%	0.340
MSL	0.6	59.3%	0.344
MSL	0.7	59.1%	0.342
ASL	0.1	59.1%	0.346
ASL	0.2	59.6%	0.344
ASL	0.3	60.3%	0.340
ASL	0.4	**60.5%**	0.337
ASL	0.5	60.1%	**0.335**
ASL	0.6	59.5%	0.340
ASL	0.7	59.2%	0.342
COMB	0.1	59.3%	0.345
COMB	0.2	59.5%	0.344
COMB	0.3	60.4%	0.336
COMB	0.4	**60.8%**	**0.328**
COMB	0.5	60.5%	0.331
COMB	0.6	59.4%	0.338
COMB	0.7	59.1%	0.348

Table 3: Performance of Sentiment Lexicon; Overall best result is marked bold ; group best is marked bold with underline; second best of each group is underlined

to that of MSL, mostly likely because ASL has better coverage. To demonstrate this, Table 4 shows the coverage of sentiment terms by different lexicons in the dataset used.

Sentiment lexicon set	MSL	ASL	COMB
Average coverage rate	13.4%	18.3%	24.1%

Table 4: Coverage of Sentiment Lexicon

The average coverage rate in Table 4 is defined as the average proportion of sentiment terms contained in the sample documents. Although MSL as manually engineered sentiment knowledge is more reliable, its coverage is significantly lower than that of ASL and COMB. Even though ASL can be quite noisy, the structure of LSTM-SAT can mitigate the effect of noise in ASL. Positive influence brought by the high coverage rate of ASL obviously exceeds the negative effect of noise in LSTM-SAT that leads to a better good performance.

4.4 Case Study

To have a deeper insight of our proposed LSTM-SAT model, we use three randomly selected instances to demonstrate the effect of LSTM-SAT using sentiment augmented lexicon compared to LSTM-AT. Their true labels and the predicated labels by LSTM-AT and LSTM-SAT are listed after the respective review text.

E1 上星期/ 剛從/ 美國/ 公幹/ 回來/ ，/ 非常/ 掛住/ 我/ 至愛/ 韓菜/ ，所以/ 前天/ 就/ 去/ 左/ 尖沙咀/ 金/ 巴利/ 道/ 嘅/ “ / 石牆/ 道/ ” / ，/ 嘆下/ 美味/ 韓國菜/ 。/ 我/ 最愛食/ 韓式/ 炸/ 鮮人參/ ，/ 用/ 天扶良/ 做法/ ，/ 炸/ 得/ 鬆/ 脆/ ，/ 鮮/ 人參/ 好爽口/ ，/ 甘中/ 帶/ 甜/ ，/ 幾好/ 食/ ，/ 值得/ 介紹/ 俾/ 大家

(I just returned from the business trip to America last week and missed my favorite Korean food very much. Therefore, I went to Tsim Sha Tsui to enjoy the delicious Korean dishes. My favorite one is the Korean-style fried ginseng. It is specially cooked in Tempura style which makes it taste crunchy, refreshing, tangy with a hint of sweetness. This dish is so tasty and it is worthy of my recommendation.)

Ground Truth: 5 LSTM-AT: 4 LSTM-SAT: 5

Figure 3: LSTM-SAT case study 1

E1 is a positive instance which our proposed LSTM-SAT correctly predicted, yet LSTM-AT has missed a bit. The sentiment terms contained in the review document include “美味” (*delicious*), “脆” (*crunchy*), “爽口” (*refreshing*), “甜” (*sweet*), “幾好” (*very*) , “值得” (*worth*), and “介紹”(*recommendation*). Figure 3 shows the heat graph of LSTM-AT and LSTM-SAT. The first row refers to the segmented text while the second and the third rows show the normalized attention weights of LSTM-AT and LSTM-SAT, respectively. Figure 3 shows that before sentiment scores are incorporated (LSTM-AT), these seven sentiment words are less significant within the sentence, especially for “脆” (*crunchy*) and “值得” (*worth*). In our proposed LSTM-SAT, on the other hand, the attention values of those positive sentiment terms are intensified. Benefited from the prior knowledge reflected by their sentiment scores, the prediction result of LSTM-AT is correctly shifted from 4 to 5 in LSTM-SAT.

E2 雞扒/ 烤餅/ 其實/ 幾好/ 食/ 塊扒/ 係/ 去/ 左/ skin / 的/ 有/ dd / 辣/, / 幾/ 入味/ 而/ 我/ 平日/ 係/ 唔/ eat / salad / 醬/ 的/ 都/feel / 佢/ 個/ 千島醬/ 同/ d / 料/ match / 得/ 好好/ 呀/ ~ / d / / 生菜/ 又/ 爽/ $ / 14.5 / / 都/ ok / / 。/ 既/ 唔/ 係/ 話/ 好/ 抵不過/ 可以/ 接受/ lor

(Grilled chicken scone is quite delicious. The chicken is skinned and a bit hot with spice, but very tasty. Although I usually do not eat with the salad sauce, the Thousand Island sauce compliment the chicken wonderfully. The simple salad with lettuce is also refreshing. The $ 14.5 extra for the salad is ok, not super, but acceptable.)

Ground Truth: 4 LSTM-AT: 3 LSTM-SAT: 4

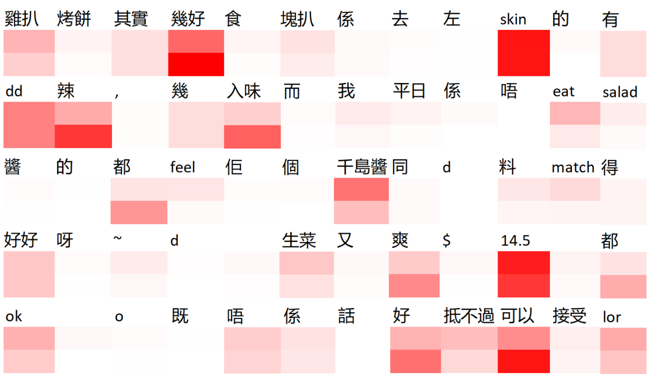

Figure 4: LSTM-SAT case study 2

E2 is also a positive example. The difference between these two examples is that some English words and digits are used in the E2. Figure 4 shows that instead of detecting sentiment words like "入味" (*tasty*) and "爽" (*refreshing*), the context-based attention mechanism highlights information in different written systems such as "*skin*" and "*14.5*", which are not related to sentiment expressions in this case. In LSTM-SAT, attention values on the sentiment words are largely increased. The other weights such as "*skin*" and "*14.5*" are relatively decreased after normalization. Adjustment of LSTM-SAT results in a more reasonable attention distribution for sentiment analysis and thus contributes to a better prediction result.

E3 米芝連/ 一星/ 雲/ 吞/ 麵/ , / 不及/ 坊/ 間/ 出名/ 雲吞/ 麵/ 鋪/ . / 干/ 炒/ 牛河/ 太油/ , / 牛肉/ 有/ 梳打/ 粉味/ , / 芽菜/ , / /, / 及/ 洋/ 比例/都/ 恰當/ , / 比/ 以往/ 在/ 舊/ 鋪/ 差/ . / 收茶/ 錢/ 及/ 加一/ , / 更加/ 可笑/ . /有/ 邊間/ 會/ 收/ 加一/ 呢/ ? / 麵/ 不夠/ 爽/ , / 湯不香/, / 雲吞還/ 可以/ . / 價錢/ 比/ 其它/ 名鋪/ 貴/ , / 物非/ 所值/.

(The wonton noodles in this 1-star Michelin restaurant is worse than other famous wonton noodles restaurants. The fried beef noodle is too greasy. Beef with rice noodle can taste the soda powder in it. The proportion of garnish including bean sprout, chives, and onion is appropriate but not as good as when the restaurant was in its old location. There is also the charge for tea and service fee. No tea house charges for service fee, this is ridiculous. The noodle is not crunchy, and the soup has no flavor. The wonton is just alright though. The food is more expensive than other popular restaurants, not worthy the high price.)

Ground Truth: 3 LSTM-AT: 3 LSTM-SAT: 2

E3 is a negative case and the predication of LSTM-AT model is correct whereas LSTM-SAT gives a more aggressive lower score. Figure 5 shows that both two attention methods put much weight on "米芝連" (*Michelin*), "一星" (*1-star*), "可笑" (*ridiculous*) and "不香" (*the soup has no flavor*). LSTM-SAT identifies and further highlights some negative words such as "差" (*bad*), "不夠" (*not enough*) and "貴" (*expensive*), resulting in a less favorable class 2. We should not neglect the fact that the incorrectness by LSTM-SAT is partially caused by the inconsistency of the text to the rating.

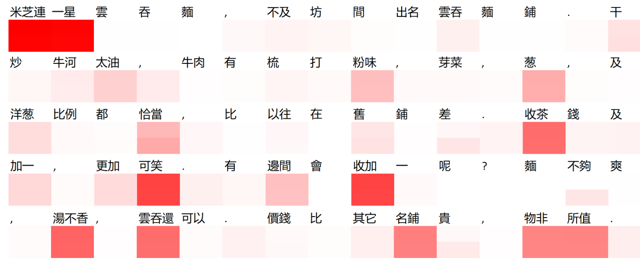

Figure 5: LSTM-SAT case study 3

The text description in E3 is quite negative but the user still gives a more generous rating. In this sense, LSTM-SAT actually gives a more reasonable class rating to reflect the sentiment of the review than ground truth label.

The above three examples show that our model has the ability to effectively highlight sentiment information. Incorporating the lexicon with sentiment scores into the attention layer can make LSTM benefit from prior sentiment knowledge and the result is more semantically interpretable.

5 CONCLUSION

This paper presents our work to use a data driven approach to build a Cantonese lexicon with sentiment scores and contribute to attention-based deep learning neural network for a Cantonese sentiment classification task. We first use an automatic method to augment the manually constructed Cantonese sentiment lexicon. According to the label distribution in a large corpus, automatically obtained sentiment terms are then augmented with predicted sentiment scores. Our proposed model LSTM-SAT leverages on the current state-of-the-art deep learning model with attention mechanism by incorporating sentiment scores into its attention layer. In general, attention mechanism provides an extra layer to include more knowledge in deep learning models. The original attention model includes semantic information mostly using local context. Here, we propose a method to incorporate explicit sentiment knowledge into deep learning models and the knowledge has theoretic basis. In other words, we can demonstrate how an explainable knowledge can be incorporated into deep learning models.

Experimental results show that our automatically constructed Cantonese sentiment lexicon helps to improve coverage and serve as sentiment knowledge with semantically informative meaning. Benefited from this information, our proposed LSTM-SAT shows a significant improvement on the performance of Cantonese sentiment classification.

Future work covers two directions. One is to refine automatic sentiment lexicon extraction method to include some validation mechanism to reduce noise. The other is to explore approaches of employing user related information which may further enhance the performance of our models.

6 APPENDICES

The datasets are available in
https://github.com/Christainx/Openrice_Cantonese

ACKNOWLEDGMENTS

The work is partially supported by the research grants from Hong Kong Polytechnic University (PolyU RTVU) and GRF grant (CERG PolyU 15211/14E, PolyU 152006/16E).

REFERENCES

[1] Alina Andreevskaia and Sabine Bergler. 2008. When specialists and generalists work together: Overcoming domain dependence in sentiment tagging. *Proceedings of ACL-08: HLT* (2008), 290–298.

[2] Dana Scott Bourgerie. 2006. Cantonese as a Written Language: The Growth of a Written Language. by Don Snow. *Journal of Asian Pacific Communication* 16, 2 (2006), 351–353.

[3] Zhenguang G Cai, Martin J Pickering, Hao Yan, and Holly P Branigan. 2011. Lexical and syntactic representations in closely related languages: Evidence from Cantonese–Mandarin bilinguals. *Journal of Memory and Language* 65, 4 (2011), 431–445.

[4] Huimin Chen, Maosong Sun, Cunchao Tu, Yankai Lin, and Zhiyuan Liu. 2016. Neural Sentiment Classification with User and Product Attention. EMNLP.

[5] Jian Chen, Yu Liu, Guangyi Zhang, Yi Cai, Tao Wang, and Huaqing Min. 2013. Sentiment Analysis for Cantonese Opinion Mining. In *2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies*. IEEE, 496–500.

[6] Christy MK Cheung and Dimple R Thadani. 2012. The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision support systems* 54, 1 (2012), 461–470.

[7] Geng Cui, Hon-Kwong Lui, and Xiaoning Guo. 2012. The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce* 17, 1 (2012), 39–58.

[8] Raffaele Filieri. 2015. What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research* 68, 6 (2015), 1261–1270.

[9] Roxana Fung and Brigitte Bigi. 2015. Automatic word segmentation for spoken Cantonese. In *2015 International Conference Oriental COCOSDA held jointly with 2015 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE)*. IEEE, 196–201.

[10] Clayton J Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

[11] Ozan Irsoy and Claire Cardie. 2014. Opinion Mining with Deep Recurrent Neural Networks.. In *EMNLP*. 720–728.

[12] Dame Jovanoski, Veno Pachovski, and Preslav Nakov. 2016. On the impact of seed words on sentiment polarity lexicon induction. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 1557–1567.

[13] Farhan Hassan Khan, Usman Qamar, and Saba Bashir. 2017. A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet. *Knowledge and information Systems* 51, 3 (2017), 851–872.

[14] John Lee. 2011. Toward a parallel corpus of spoken Cantonese and written Chinese. In *Proceedings of 5th International Joint Conference on Natural Language Processing*. 1462–1466.

[15] Yunfei Long, Lu Qin, Rong Xiang, Minglei Li, and Chu-Ren Huang. 2017. A Cognition Based Attention Model for Sentiment Analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 473–482.

[16] Yunfei Long, Rong Xiang, Qin Lu, Chu-Ren Huang, and Minglei Li. 2019. Improving attention model based on cognition grounded data for sentiment analysis. *IEEE Transactions on Affective Computing* (2019).

[17] Mateus Tarcinalli Machado, Thiago AS Pardo, and Evandro Eduardo Seron Ruiz. 2018. Creating a Portuguese context sensitive lexicon for sentiment analysis. In *International Conference on Computational Processing of the Portuguese Language*. Springer, 335–344.

[18] Diana Maynard and Adam Funk. 2011. Automatic detection of political opinions in tweets. In *Extended Semantic Web Conference*. Springer, 88–99.

[19] Prem Melville, Wojciech Gryc, and Richard D Lawrence. 2009. Sentiment analysis of blogs by combining lexical knowledge with text classification. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1275–1284.

[20] EWT Ngai, MCM Lee, YS Choi, and PYF Chai. 2018. Multiple-Domain Sentiment Classification for Cantonese Using a Combined Approach. (2018).

[21] Gang Peng. 2006. Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics* 34, 1 (2006), 134.

[22] Qiao Qian, Minlie Huang, Jinhao Lei, and Xiaoyan Zhu. 2016. Linguistically regularized lstms for sentiment classification. *arXiv preprint arXiv:1611.03949* (2016).

[23] Richard Socher, Jeffrey Pennington, Eric H Huang, Andrew Y Ng, and Christopher D Manning. 2011. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 151–161.

[24] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, Vol. 1631. Citeseer, 1642.

[25] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics* 37, 2 (2011), 267–307.

[26] Duyu Tang, Bing Qin, and Ting Liu. 2015. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 1422–1432.

[27] Zhiyang Teng, Duy-Tin Vo, and Yue Zhang. 2016. Context-Sensitive Lexicon Features for Neural Sentiment Analysis.. In *EMNLP*. 1629–1638.

[28] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 347–354.

[29] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

[30] Hua-Ping Zhang, Hong-Kui Yu, De-Yi Xiong, and Qun Liu. 2003. HHMM-based Chinese lexical analyzer ICTCLAS. In *Proceedings of the second SIGHAN workshop on Chinese language processing-Volume 17*. Association for Computational Linguistics, 184–187.

[31] Ziqiong Zhang, Qiang Ye, Zili Zhang, and Yijun Li. 2011. Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications* 38, 6 (2011), 7674–7682.

[32] Yicheng Zou, Tao Gui, Qi Zhang, and Xuanjing Huang. 2018. A Lexicon-Based Supervised Attention Model for Neural Sentiment Analysis. In *Proceedings of the 27th International Conference on Computational Linguistics*. 868–877.