Multi-review opinion summarization via jointly modeling semantic and helpfulness

Group Name: DATA SMILE
Members: Jiangtao, LIU; Xin, Li; Dihan, CHEN; Jialong, CHEN

Background

Abstract

Nowadays, reviews and ratings are playing a more important role in the Internet social media. In this project, we plan to construct a system to analyze Chinese movie reviews as to realize multi-review opinion summarization via jointly modeling semantic and helpfulness. The whole project includes five parts: data garbing, word segmentation, semantic analysis, multi-purpose aggregation and visual presentation. Among them, more attention will be paid to the semantic analysis and multi-purpose aggregation. Several concepts, such as LSH, linear regression, machine learning and unsupervised learning, are combined to help realize several objectives of the opinion summarization, including the cleaning of useless comments, the refining of the key characteristics, the evaluation of review quality and other objectives. Further, since there are many scenes full of reviews and ratings upon different objects, such as foods, books, attractions and Weibo hot topics, our system could be easily transformed to analyze them, which might provide a new perspective on things and could perform much more efficiently upon data of large volume or high velocity.

Motivation

Since the Internet has achieved exponential increasing in these decades, billions of people all over the world benefit from the diversity and immediacy of information brought by the Internet, which serves as an infinite platform for sharing opinions about everything, such as book reviews, video reviews, shop reviews, social hot comments, etc. Meanwhile, people are also getting accustomed to enjoying these reviews and ratings before or after appreciating the contents to help them preview or search for someone with similar interests. Compared with ratings, text reviews, especially long text comments, are usually more reliable and convinced due to the responsibility and interest of the authors, which might provide plenty of valuable information. However, due to the limited reading speed with respect to the massive data in the Internet, it becomes much harder to obtain useful information efficiently without any auxiliary tools. At the same time, a large amount of junk comments induced by commercial interests also influence the quality of the ratings and reviews. To solve the problem of the summarization of these text-based reviews, we plan to build such an online review analysis system via advanced big data algorithms to improve the process, starting from Chinese movie reviews. Multiple review purposes, such as the cleaning of useless comments, the refining of the key characteristics, the evaluation of review quality and other objectives will be implemented.

Further hidden interesting clues are also expected to be dug out via applying the algorithms taught in this course.

Topics related

To achieve the final goal, many subtopics in this course are highly useful for data processing. Two of the most related course topics should be text processing and machine learning. Text processing is the preliminary step for the abstraction, and machine learning is employed to aggregate and summarize the conclusions from raw big data. Besides, locality sensitive hashing and dimensionality reduction might also have a chance to be applied to simplify data processing or enhance the final performance.

Technical details

Deliverables

We intend to submit one movie review sentiment analysis model, which could identify one certain sentence into three possible sentiments. The three emotional tendencies are positive (POS), middle (MID) and negative (NEG), where positive means reviewers recommend that movie, middle means reviewers have neutral emotion or no obvious sentiment and negative means reviewers have bad experience when watching that movie.

To evaluate the accuracy of the Sentiment model, we would like to use precision factor which comes from the following expression:

Precision = (#correctly recognized sentences) / (#all input sentences)

Dataset

The dataset comes from the famous movie town — movie.douban.com, which have millions of movie lovers from mainland China. The whole dataset contains 20 different types of movies that are classified by official Douban website. Each type includes around 1000 popular movies, where one movie has at least 2000 reviews. The reviews are divided into three categories following the scored stars when reviewers made these reviews. We assume 4-5 stars is positive attitude, 3 stars is middle attitude and 1-2 stars is negative emotion. Now we have approximately 500 reviews per class of emotions per movie to train our model and another 500 hottest reviews plus 100 newest reviews per movie for testing. We developed our unique crawler to obtain all the reviews from website and build the special dataset ourselves.

The example of basic statistics of the dataset is shown in *Table 1*, which takes The Shawshank Redemption (肖申克的救赎) for example. The dataset contains the corresponding reviews of various movies and each movie has its own five rows of reviews, including 500 positive reviews, 500 middle reviews, 500 negative reviews, 500 hottest reviews and 100 newest reviews. The total number of movies in our dataset is around 10000.

Table 1: Basic statistics of the dataset

Movie Name	*1*	*2*	...

(e.g., 肖申克的救赎)			
POS (500 reviews)	超级喜欢，不看的话人生不圆满	经典的电影	...
MID (500 reviews)	好看但这么高的分有点假了吧	在我的心中，它一直都是最被高估的电影	...
NEG (500 reviews)	真的不喜欢，不好看，没感觉	大众经典我从不感冒，为什么？我欣赏水平不行？	...
HOT (500 reviews)	不需要女主角的好电影	关于希望最有力的诠释	...
NEW (100 reviews)	超级震撼	看完后劲很大的一部经典，强推！	...

Work review & differences

Our project is to capture the semantic features and the helpfulness of the review and determine the product-level / entity-level opinions according to both of the review representations and the review helpfulness. First of all, this project is related to classification-oriented representation learning. [1][4][5][6] employ the convolutional neural networks to learn the sentence/document-level representations. [2] explicitly models the contributions of the sentences within the review via a hierarchical Gated Recurrent Unit (GRU). Different from the above-mentioned works, [3][9] proposes to combine the temporally recurrent structure and convolutional feature extractor, achieving competitive classification results. Recently, the language model pre-training techniques [7][8] have been introduced into the classification models and substantially advance the state-of-the-art for the document classification task.

This project is also related to review helpfulness prediction, whose aim is to automatically assess the helpfulness rating of the reviews. [11] and [12] are the pioneering works exploiting hand-crafted features for the prediction of review helpfulness. [10] further improves the prediction performance based on argument-based features. [13][14] conduct some initial attempts to perform the helpfulness prediction with the help of neural network based models and establish the new state-of-the-art results.

The differences between these problems and the problem investigated in our project are as follows:

1. Our project does not solve these two problems separately but unifies them within a multi-task learning framework for predicting the product-level opinions.

2. The supervision signals required by the helpfulness prediction task are dispensable in our framework. Theoretically, in order to predict product-level opinions more accurately, the model has the tendency to highlight the useful reviews and down-weight those useless.

Applied algorithms

Due to the sparsity and the expensive computational cost brought by one-hot word representation, we plan to employ word embedding techniques [15][16] to map each word to a low-dimensional real-valued vector. As with the learning of review-level semantic representations, we plan to develop an enhanced self-attention networks [17] with the ability of modeling the hierarchical structure within each review. After obtaining a sequence of review representations, we can borrow the recurrent neural networks (RNN) and convolutional neural networks (CNN) to calculate the helpfulness-aware overall representations. The review helpfulness can be learned by following the multi-task learning (MTL) paradigm [18] (when the gold standard helpfulness scores are available) or directly learned from the training data at hand.

Visual presentation

We intend to present our analysis results with the following several points. The main format of our output result will be graph.

Firstly, one training and testing result with the dataset of Top 250 movies will be presented, mainly comparing our calculated ratings with actual ratings, and find out the main bright spot of each movie.

Secondly, several training and testing results with the dataset of movies of different types will be presented, illustrating our function of cleaning of useless reviews and helping evaluate the helpfulness of each review.

Thirdly, several online real-time testing with new movies which are released in November will be presented. This could help us evaluate the actual quality of new movies and perform real-time rating. Further, this application demonstrates our ability to deal with big data with high velocity.

Timeline

The complete project schedule with milestones is presented in *Table 2*.

Table 2: Project schedule with milestones

Schedule	Milestones
Garbing data	Oct. 3rd, 2019
Modeling word segmentation & semantic analysis model with unsupervised machine learning	Oct. 13th, 2019
Model training & test with the dataset of Top 250 movies	Oct. 23th, 2019
Model training & test with the dataset of Top 250 movies	Nov. 3rd, 2019
Online real-time testing with new movies	Oct. 23th – Nov. 23th, 2019
visual presentation	Nov. 30th, 2019

References

[1] Kim, Yoon. "Convolutional Neural Networks for Sentence Classification." Proceedings of EMNLP. 2014.

[2] Yang, Zichao, et al. "Hierarchical attention networks for document classification." Proceedings of NAACL-HLT. 2016.
[3] Tang, Duyu, Bing Qin, and Ting Liu. "Document modeling with gated recurrent neural network for sentiment classification." Proceedings of EMNLP. 2015.
[4] Johnson, Rie, and Tong Zhang. "Supervised and semi-supervised text categorization using LSTM for region embeddings." Proceedings of ICML. 2016.
[5] Johnson, Rie, and Tong Zhang. "Deep pyramid convolutional neural networks for text categorization." Proceedings of ACL. 2017.
[6] Zhang, Xiang, Junbo Zhao, and Yann LeCun. "Character-level convolutional networks for text classification." Proceedings of NeurIPS. 2015.
[7] Howard, Jeremy, and Sebastian Ruder. "Universal Language Model Fine-tuning for Text Classification." Proceedings of ACL. 2018.
[8] Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." Proceedings of NAACL. 2019.
[9] Lai, Siwei, et al. "Recurrent convolutional neural networks for text classification." Proceedings of AAAI. 2015.
[10] Liu, Haijing, et al. "Using Argument-based Features to Predict and Analyse Review Helpfulness." Proceedings of EMNLP. 2017.
[11] Martin, Lionel, and Pearl Pu. "Prediction of helpful reviews using emotions extraction." Proceedings of AAAI. 2014.
[12] Yang, Yinfei, et al. "Semantic analysis and helpfulness prediction of text for online product reviews." Proceedings of ACL. 2015.
[13] Fan, Miao, et al. "Product-Aware Helpfulness Prediction of Online Reviews." Proceedings of WWW. 2019.
[14] Chen, Cen, et al. "Review Helpfulness Prediction with Embedding-Gated CNN." arXiv preprint arXiv:1808.09896. 2018.
[15] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Proceedings of NeurIPS. 2013.
[16] Pennington, Jeffrey, Richard Socher, and Christopher Manning. "Glove: Global vectors for word representation." Proceedings of EMNLP. 2014.
[17] Vaswani, Ashish, et al. "Attention is all you need." Proceedings of NeurIPS. 2017.
[18] Caruana, Rich. "Multitask learning." Machine learning 28.1 (1997): 41-75.