PresentationTrainer: Oral Presentation Support System for Impression-Related Feedback

Shengzhou Yi,[1] Hiroshi Yumoto,[2] Xueting Wang,[1] Toshihiko Yamasaki[1]
[1]Department of Information and Communication Engineering, The University of Tokyo
[2]P&I Information Engineering Co., Ltd. Innovation Business Headquarters
[1]{yishengzhou, xt_wang, yamasaki}@hal.t.u-tokyo.ac.jp
[2]yumoto@pandi.co.jp

Abstract

In order to support the pratice of oral presentation, we developed PresentationTrainer which includes (1) a presentation impression prediction system and (2) a presentation slide analysis system. For the presentation impression prediction system, we proposed two methods, using Support Vector Machine and Markov Random Field, or using multimodal neural network, to predict audiences' impressions for speech videos. For the slide analysis system, we used Convolutional Neural Network and Global Average Pooling to evaluate the design of slides. We then used Class Activation Mapping to provide visual feedback for showing which areas should be modified.

Oral presentation is the most standard format to express ideas or introduce products in many scences. However, there are few efficient tools that can automatically evaluate the presentation. We developed PresentationTrainer, which includes a presentation impression prediction system and a presentation slide analysis system, to evaluate presenters' performance and provide impression-related feedback.

In (Yamasaki et al. 2015) and (Yi, Wang, and Yamasaki 2019), we respectively used statistical machine learning and deep learning to predict the impressions, a presentation could give to the audiences. In (Oyama and Yamasaki 2019), we proposed a method to evaluate slides and to provide a visual feedback to tell presenters which areas of their slides are better to be modified in order to make a better impression.

Presentation Impression Prediction System

We proposed two methods to predict the audiences' impressions based on linguistic feature as well as acoustic feature. One method used Support Vector Machine (SVM) and Markov Random Field (MRF), and the other method used a multimodal neural network and an attention mechanism.

SVM-MRF

We extracted linguistic feature and acoustic feature from the captions and the audio data, respectively. For the linguistic feature, we used multiple word embedding methods, including Bag-of-words, Latent Semantic Indexing, Latent Dirichlet Allocation, skipgram, and surface-level features. We averaged the vector of all words as document embedding. We extracted the acoustic feature by using openSMILE.

We used SVM to predict the impression labels with only one type of document embedding or only acoustic feature at a time. We then used MRF to consider the correlations between different features and even different impressions to relabel the results of SVMs. However, this method can only predict the impressions for complete presentation videos.

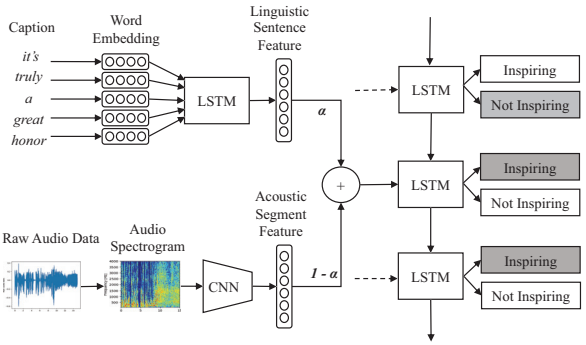

Figure 1: Continous impression prediction

Multimodal Neural Network

We used a multimodal neural network to provide presenters with continuous feedback (Figure 1). We used skip-gram for word embedding, and input the word vectors in each sentence to Long Short-Term Memory (LSTM) in order to get sentence-level linguistic features. We extracted the acoustic features of audio segments, corresponding to each sentence. We took out audio segments and transfered it into frequency domain by Short-Term Fourier Transform. We then used Convolutional Neural Network to extract the deep feature from audio spectrograms as segmental acoustic features.

After we got the sentence-level features, we used an attention mechanism for feature fusion. We then used LSTM to consider the correlations between sentences and predicted audiences' impressions on each sentence. According to the prediction results, the presenters can understand which sentences may leave bad impressions and need to be modified.

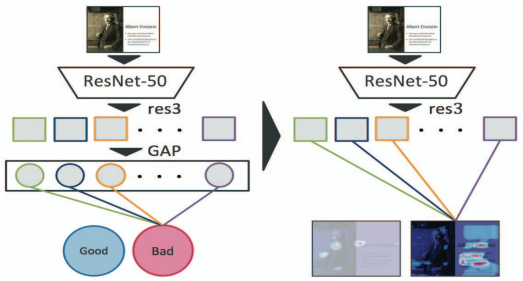

Figure 2: Negative Class Activation Mapping

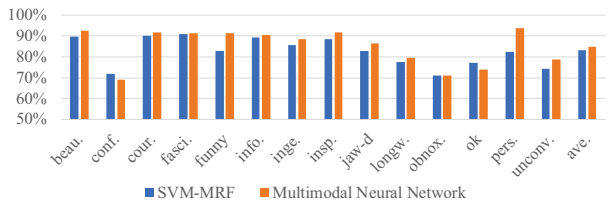

Figure 3: Prediction accuracy for different impressions

Slide Analysis System

Apart from the presenters's performance, the design of slides is also a key element to a successful presentation. We extracted and concatenated image features, structural features, and content features of slides. We used SVM to predict whether the designs are good or not. If users only get a score, they may not be convinced and don't know how to modify their slides. Therefore, we also proposed a visualization method by using Class Activation Mapping (Zhou et al. 2016) to tell presenters which areas of their slides may give negative impressions (Figure 2). We put a Global Average Pooling (GAP) after the Res3 layer of ResNet-50 to learn the weights of class "Positive" and class "Negative". We then gave the weights of "Negative" to the feature maps and got the heatmaps that can show the "Negative" areas.

Experiment

Presentation Impression Prediction

We used the captions and the audio data of 2,445 presentations. Each presentation has 14 impression tags, including *Beautiful*, *Confusing*, *Courageous*, *Fascinating*, etc. These tags are based on all audiences' votes and tell us whether the audiences have corresponding impressions.

The prediction results of complete presentations impressions of two proposals are shown in Figure 3. SVM-MRF achieved higher efficiency but can't predict impressions continuously as multimodal neural network does. Therefore, we predict the impressions of the complete presentation and each sentence by these two methods, respectively.

Slide Analysis

We hired 100 workers to create 1,000 PowerPoint slides in 10 topics, and they gave each slide with a visual clarity rank (1 to 100). We treated top and bottom 30% of the slides as "Positive" and "Negative" samples, respectively.

Figure 4: Presentation impression prediction system

Figure 5: Slide analysis system

We achieved the classification rate of 90.3%. We then further proposed a feedback system that can provide visual clarity scores and point out areas that should be modified.

Demo System

In the demo, we presented our web application of presentation support system. Users only need to upload their presentation videos or presentation slides. Our system will automatically analyze them behind the scences. We first presented our presentation impression prediction system. This system predicted audiences' impressions for each complete presentation from 14 aspects, and it also showed the temporal change of these impressions continuously during the presentation (Figure 4). We then presented our slide analysis system, which gave slides with clarity scores and told us which areas gave audiences negative impressions (Figure 5).

References

Oyama, S., and Yamasaki, T. 2019. Visual clarity analysis and improvement support for presentation slides. In *IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, 421–428.

Yamasaki, T.; Fukushima, Y.; Furuta, R.; Sun, L.; Aizawa, K.; and Bollegala, D. 2015. Prediction of user ratings of oral presentations using label relations. In *Proceedings of the 1st ACM International Workshop on Affect & Sentiment in Multimedia*, 33–38.

Yi, S.; Wang, X.; and Yamasaki, T. 2019. Impression prediction of oral presentation using lstm and dot-product attention mechanism. In *IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, 242–246.

Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; and Torralba, A. 2016. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2921–2929.