# A Review of Sentiment Semantic Analysis Technology and Progress

Yiming Wang, Yuan Rao*, Lianwei Wu
(Lab of Social Intelligence & Complex Data Processing, School of Software,
Xi'an Jiaotong University, Xi'an 710049)

***Abstract*—Sentiment computing promotes some new application opportunities and technique challenges in artificial intelligence of the next generation, and it has become a fascinating research field. In this paper, the conception of sentiment computing with some core elements and feature vectors is defined, and some vital issues are proposed. Based on the theories mentioned above, the subjective content or objective content is classified by some special algorithms in the scenarios of single modal, such as text, image, audio and video data. Furthermore, the method to merge these different kinds of data and to further form the multimodal analysis methods for emotion detection is an important problem, and the fusion strategy is summarized in the paper. Finally, some trends about the sentiment cognition and sentiment generation are analyzed, which provides new ways for further research work.**

***Keywords*—sentiment analysis; multimodal; social sentiment; opinion mining; sentiment generation**

## I. Definition of Sentiment Computing

In general, we focus on subjective experience and external performance of sentiment. In details, subjective experience refers to the individual's feelings of different emotional state, and external performance refers to the behavior and action of different parts of body when the emotional state occurs. In addition, in order to research sentiment expression in physiological wake-up, Picard [1] uses wearable devices to analyze sentiment features under various stress on personal physiology and is the earliest to propose the definition of "sentiment computing". Namely, the sentiment computing is a kind of computational science that is related to sentiment, derived from sentiment, or exerts influence on sentiment. While based on the perspective of the subjective point of view, Liu [2] defines that sentiment computing is the field of study which is used to analyze people's opinions, sentiments, evaluation, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. To sum up, sentiment computing is an interdisciplinary science, fusing cognitive psychology, Natural Language Processing, and multi-modal recognition knowledge, which mainly uses information of objective world to show the tendency, perspectives, stances and attitudes of subjective sentiment. It is a process for sorting and summarizing features such as emotional polarity, intensity, and variation tendency.

## II. Sentiment Computing In Single Modal

The carriers of sentiment are not only some physiological characteristics like complexion, heartbeat and skin conductance, but also texts, sound, images, gestures, etc.. These characteristics are often combined to express a kind of sentiment. However, when we study sentiment, we must begin with a thorough study of sentiment computing under single modal.

### *A. The Sentiment Computing and Measurement of Text*

The text sentiment processing can be divided into three levels: words, sentence and document. For word-level sentiment computing, we often search for the relationship of words in text and that in sentiment lexicons. Thus, the study of building sentiment lexicons has attracted a lot of attention. At present, some famous English sentiment lexicons like GI, BL-Lexicon, MPQA-Lexicon, and Chinese sentiment lexicons like DUTIR emotional lexical ontology, HowNet lexicon and NTU evaluation dictionary have been applied universally. Although these lexicons are highly universal, the scale of words is small. However, how to build domain adapted sentiment lexicons is still a challenge. The special domain sentiment lexicon proposed by Anil et al. [3] can extract better features than the current sentiment lexicon using supervised Dirichlet Distribution and PMI. Wu et al. [4] used the same framework to build a microblog sentiment lexicon which integrates sentiment of sentiment words extracted from microblog and the existing lexicon, and similarity of words extracted from context. However, at least a whole sentence instead some words, can express the whole sentiment and semantic without ambiguity. Poria et al. [5] used Sentic Patterns to convert natural language sentences into a hierarchical structure, which can determine the sentiment polarity of the entire sentence. Fu et al. [6] used a semi-supervised learning method combined with the HowNet lexicon to train a phrase recursive auto encoder (PRAE) to compute emotion under sentence level, which resolves the problem of ordinary recursive self-encoders that they require large quantities of tags. Moreover, the document level sentiment analysis is more in line with practical applications, because the process takes the relationship and position of its sentences into account. Chau et al. [7] used the Stanford parser to construct all sentences of a document into a tree. After data preprocessing, these sentences are sorted by depth-first search algorithm, and then a depth learning algorithm is applied to classify sentiment. Nevertheless, a classifier with good performance in one domain is likely to degrade in other domains. Zhou et al. [8] proposed a topic-related transfer algorithm to learn the information of different domains from a variety of texts, and transfered them into a uniform theme which expresses domain content of different domains' product comments.

Currently, we tend to focus on sentiment analysis of multiple aspect objects from sentiment documents like "this restaurant's food is delicious, but the environment is bad". Liu et al. [9] used POS (part of speech) and chunk features to mark aspect terms. And afterwards, Tang et al. [10] took advantage of context to identify different aspects, and used Deep Memory Network to determine the final opinion of the text. What's more, when we analyze the sentiment content of the text, we mainly consider the subjective content. Subjectivity refers to the

subjective expression tendency of human, it can show people's sentiment tendency for objects intuitively, and objective content is used to obtain and analyze the basic attributes and characteristics of objects. Therefore, when we analyze the sentiment content, we should exclude the noise caused by the description of objective objects and their characteristics. Aston et al. [11] proposed algorithms such as Perceptron, Perceptron with best learning rate, and vote Perceptron to classify the sentiment of tweets in data stream and they can determine the subjectivity or objectivity of a tweet with a low error rate.

### *B. The Sentiment Computing and Measurement of Image*

Image is the most intuitive way to show sentiment. Among features of images, we know that color, texture, shape and contour can express emotion, but shape and contour provide less sentiment information, that's why we focus more on color and texture. Gong et al. [12] further studied other factors that affect sentiment from color, and find that the contribution of hue to emotion is larger compared to chroma and lightness. Texture provides important information for high-level semantic analysis. Classical methods of feature description include GLCM, MRF theory, fractal theory, Tamura feature, wavelet theory, local binary pattern (LBP) and so on. Wang et al. [13] learned from the idea of constituting the gray-level co-occurrence matrix to deal with gray image texture space, and take energy values from three value model matrix as the feature value, to avoid the problem of sparse histogram, and extract related information of texture changes of partial image. However, the research of texture features still has many problems, such as the problem of complex boundary of multi-texture type and the difficulty of distinguishing the texture from the field of image recognition.

On the other hand, different image types can be subdivided into facial expression image, gesture image and ordinary image. Basic procedures are expression preprocessing, feature extraction and sentiment classification. However, there is a large semantic gap between image and sentiment, because image features are more intuitive, but sentiment is more abstract. The same picture when seen by different people will get different sentiment responses because of different background, habits and so on. Methods of facial expression recognition can be divided into the several methods based on geometric feature extraction, model-based, motion and deformation analysis, statistical feature and frequency-domain methods. Shafiq et al. [14] proposed an algorithm to extract the geometric features from a single face image to detect the painful expression. At present, there are many difficulties in facial expression recognition, such as the recognition of subtle facial changes and occluded facial expression, and the influence of illumination, the establishment of face models and so on. What's more, gesture expression is generally divided into body posture expression and gesture expression. However, gesture expression shows little universal, but body gestures provide a lot of help for sentiment expression, for example, rubbing hands represent anxious or nervous. Piana et al. [15] extracted gestures, kinematics and geometric features from the 3D skeleton sequence based on the emotional model of six kinds of classification. After that they used a multi-class SVM to classify, thus achieve sentiment classification based on gesture. In addition, to perform the sentiment analysis of ordinary images such as scenery images, deep CNN is a commonly used classification method. Sun et al. [16] used different methods to extract hierarchical features of images including low-level features, middle features and high-level features, and then used deep CNN to find out sentiment areas from the image with color, structure, composition, content and other features.

### *C. The Sentiment Computing and Measurement of Audio*

Audios contain all sound that people can hear, and some of audios can even mobilize people's sentiment directly. Audio has many features for expressing emotion, like rhythm, sound quality, spectrum, spectral power and so on. It is noteworthy that the cepstrum classification effect in the spectral features is better than the linear spectrum, which is widely accepted. What's more, speech and music are most common in all kinds of audio.

When transmitting semantic information, speech carries a lot of sentiment factors. In different situations, people use different intensity, tone, speed, so the same sentence can arouse different sentiment of listeners. Without even considering the semantic, the sentiment features of speech are mainly rhythmic features, sound quality features and spectral features. At present, people tend to use deep learning algorithms, especially CNN and LSTM to achieve higher accuracy. Anand et al. [17] combine the advantages of CNN and LSTM to achieve sentiment analysis of speech data, and compare the effect in different filter layers of the CNN and get the best result with single filter layer. Trigeorgis et al. [18] also used CNN and LSTM to analyze context-related sentiment relations in speech which is context-dependent. Thus the system can learn from the unprocessed speech signal automatically to obtain the best performance. So it is obvious that music contains almost all kinds of sentiments. In order to make people find music quickly corresponding to their sentiment, many music retrieval systems based on sentiment analysis are developed. Nalini et al. [19] combined MFCC with residual phase (RP) to construct different models for different sentiment, and then classified them with classifiers to identify sentiment content in music. On the other hand, since music and speech both have acoustic characteristics, their sentiment features also have some similarities. Coutinho et al. [20] used deep transfer learning to demonstrate that speech sentiment and music sentiment are similar in the field of acoustic. At present, music sentiment analysis is worthy of being studied in many directions, such as Automatic Text Summarization for music, sentiment analysis combined with music features, music retrieval systems integrated of sentiment and other statistical features and so on.

## III. MIXED SENTIMENT COMPUTING METHODS IN MULTIMODAL

### *A. Research and Application of Bimodal*

With the rapid development of multimedia, people tend to get a lot of information from a variety of modality, multimodal sentiment analysis derives from the bimodal sentiment, which has been widely applied to some domains. For example, audio combined with the text is used to analyze the music with lyrics, images and audio are combined to analyze sentiment content of music albums. In order to achieve sentiment classification of bimodal sentiment, Cai et al. [21] used image CNN, text CNN and text-image fusion CNN to merge sentiment of Twetter text-image co-expression. It is worth noting that input of fusion CNN comes from the last but one layer CNN of image and text, and the last "softmax" layer is finally used to classify two kinds of sentiment polarities. While Chen et al. [22] used two different

classifiers to train cross-domain images and related comments, and then perform their weight to classify sentiment of images. Abburi et al. [23] aimed at analyzing the user's online speech comments, they extracted the MFCC feature in the audio and then used the Gaussian mixture model (GMM) to classify, then used the Doc2vec vector to calculate the text feature and used SVM to classify, finally fuse the two modes in the decision-layer to identify sentiment classification. Zadeh et al. [24] aimed to analyze sentiment content of video image and speech comments, they extracted bag-of-words features set from speech clips and binary features of facial posture (including smile, frown, nod, shake head) from images, then connected the features based on multimodal sentiment lexicon to train them to get sentiment classification. Many other combinations of bimodal fusion sentiment analysis, such as speech and ECG signals, speech and pulse, etc., are worth further exploring.

### *B. Research and Application of Multimodal*

In recent years, many people publish comments on goods on YouTube, Facebook and other social media. It is also necessary for consumers to understand the actual quality and efficacy of goods in this intuitive way. Therefore, the opinion mining in the video is the trend for sentiment computing. So, we need to extract features and integrate them to improve the accuracy of sentiment computing in a variety of modes (text, image, audio) in video. We generally have two fusion methods: feature layer fusion and decision layer fusion. The feature layer fusion refers to combine features from each single mode as an eigenvector, and then uses the classifier to classify sentiment uniformly. In the decision layer fusion, each mode is modeled and classified independently, and the classification results obtained in single mode are finally fused together by selecting an appropriate metric method. The steps of multimodal sentiment analysis under two different fusion strategies are shown in Fig.1.

How do we select these strategies when we fuse multimodal sentiment? Poria et al. [25] conducted comparative experiments of single-mode sentiment analysis, two-mode fusion, feature layer fusion of three modes, decision layer fusion of three modes, and two modes fuse in the feature layer, then fuse the third mode. Their experimental result showed that the more modes we fuse, the higher accuracy of the sentiment recognition we get. Moreover, the fusion effect of the three modes in feature layers is better than that in decision layer, but the processing efficiency of decision layer fusion is significantly better than that in feature layer. The classification strategy that fuses two modes in the feature layer and the third in the decision layer gets the optimal performance. While Siddiquie et al. [26] concluded their work in different directions, their experiment made decision layer fusion (Siddiquie expressed for the late fusion) to further refine into simple late fusion and learning-based late fusion. Late fusion refers to adding decision layer scores obtained by each modal, resulting in a mixed decision score to achieve classification; learning based late fusion refers to training a classifier based on machine learning (Siddiquie uses logistic regression) fusion based method to integrate the decision score which is weighted in each modality to determine the combined score. They contrast the three modes (image, text, and audio) of the video with the two kinds of late fusion and early fusion (feature fusion) with experiments, and proved that the late fusion effect based on learning is the best, followed by simple late fusion, and the early fusion is the worst.

In summary, the feature layer fusion is simpler, but the fusion performance is generally better because the feature layer fusion takes the relationship between the different modes into account. While the decision layer fusion can choose the optimal classifier for each mode, so it may show better classification performance under certain conditions. The latter is better when the decision level fusion is subdivided into the simple decision layer fusion and the learning based decision layer fusion. On the other hand, in terms of processing speed, the feature layer fusion is much slower than the decision layer fusion, because the feature layer combines all features of different modes into a large feature vector matric, which increases the complexity of computing. In order to reduce the complexity of the algorithm, we generally reduce the dimension of its processing, however, how to reduce the dimension and at the same time how to avoid the loss of precision are the problems we are facing.

In addition, in order to apply multimodal sentiment analysis to getting more complex response from sentiment robots, we pay attention to how to generate sentiment content, and how to use them to generate multimodal sentiment expression. For generating sentiment content, Zhou et al. [27] took the consistency of sentiment into account on the basis of the consistency of the generated content, and use the "Encoder-Decoder" frame to get sentiment classification vectors, and further generate sentiment words explicitly and compose sentences for response. In order to express sentiment based on sentiment content related to different modes, Bai et al. [28] used SVM to learn semantic and analyze the relation between words from descriptive sentences and proposed a view model to generate facial sentiment expression. Tsujimoto et al. [29] used RNNRCM to learn from sequences of gesture from four different sentiment expression of the human body, and then generate corresponding emotional gestures. However, how to combine different sentiment expressions in order to generate multimodal sentiment is still awaited in further research.

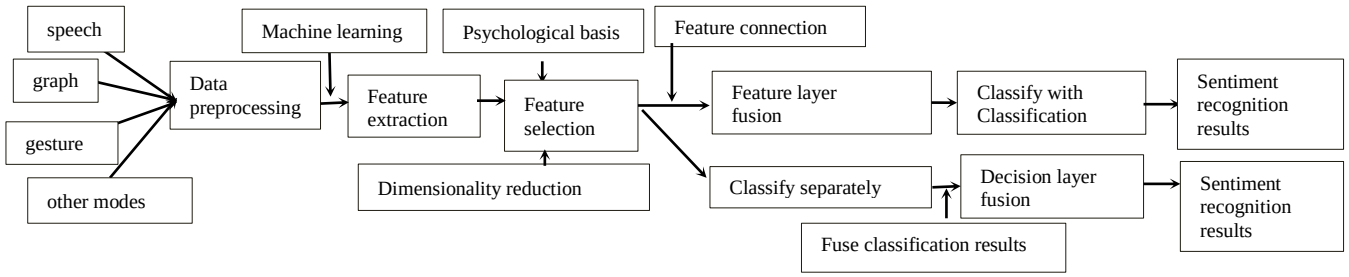


Figure 1. Multimodal feature layer fusion and decision layer fusion steps

## IV. Summary and Prospect

Sentiment computing in the field of artificial intelligence is an important technology and research direction. This paper begins with the basic concept of sentiment computing. Moreover we introduce different methods of sentiment analysis in single mode like text, image, and audio. Finally, we discuss two main sentiment fusion methods to analyze how to gain a better accuracy when classifying sentiment under multimodal, and how to generate multimodal sentiment. With the continuous improvement of technology and application, on the one hand, sentiment computing solves problems in different areas like pattern recognition, context semantics, complaints, irony and other sentiment analysis problems by establishing more effective sentiment cognitive modals. Especially, the development of cognitive science takes the human sentiment into different scenes, thus makes people understand the sentiment world accurately, and realizes the prediction of emotional evolution trends. New and effective sentiment computing methods have attracted a lot of attention. Such as the Lifelong Learning method [30] that can improve the ability and accuracy of sentiment computing in cross-domain, cross-cultural and cross-language conditions. Therefore, cognitive science and new Machine Learning methods provide new opportunities and challenges for sentiment computing.

## V. Acknowledgment

This paper is jointly supported by the funds as follows: The Fund of "Integration between Clouds computing and Big Data" from the Ministry of Education (2017B00030); The Fundamental Research Funds for the Central Universities (ZDYF2017006); Shaanxi Tobacco Corporation's Research Fund (ST2017-R011) and Shaanxi Province's Science and Technology Funds (2015XT-21).

## References

[1] R. W. Picard. Affective computing[M],The MIT Press,1997.09.

[2] B. Liu. Sentiment analysis and opinion mining[J]. Synthesis Lectures on Human Language Technologies, 2016, 30(1):152-153.

[3] A. Bandhakavi, N. Wiratunga, D. Padmanabhan, S. Massie. Lexicon based feature extraction for emotion text classification[J]. Pattern Recognition Letters, 2016.

[4] F. Wu, Y. Huang, Y. Song, S. Liu. Towards building a high-quality microblog-specific Chinese sentiment lexicon[J]. Decision Support Systems, 2016, 87:39-49.

[5] S. Poria, E. Cambria, A. Gelbukh. Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis[C]// Conference on Empirical Methods in Natural Language Processing. 2015.

[6] X. Fu, W. Liu, Y. Xu, L. Cui. Combine HowNet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis[J]. Neurocomputing, 2017,241:18-27.

[7] N. P. Chau, V. A. Phan, M. L. Nguyen. Deep learning and sub-tree mining for document level sentiment classification[J]. 2016 Eighth International Conference on Knowledge and Systems Engineering (KSE). 2016,268-273. ISSN: 978-1-4673-8929-7.

[8] G. Zhou, Y. Zhou, X. Guo, X. Tu, T. He. Cross-domain sentiment classification via topical correspondence transfer[J]. Neurocomputing, 2015, 159(1):298-305.

[9] P. Liu, S. Joty, H. Meng. Fine-grained opinion mining with recurrent neural networks and word embeddings[C]// Conference on Empirical Methods in Natural Language Processing. 2015:1433-1443.

[10] D. Tang, B. Qin, T. Liu. Aspect level sentiment classification with deep memory network[C]// Conference on Empirical Methods in Natural Language Processing. 2016:214-224.

[11] N. Aston, J. Liddle, W. Hu. Twitter sentiment in data streams with perceptron[J]. Journal of Computer & Communications, 2014, 02(3):11-16.

[12] R. Gong, Q. Wang, Y. Hai, X. Shao. Investigation on factors to influence color emotion and color preference responses[J]. Optik - International Journal for Light and Electron Optics, 2017, 136:71-78.

[13] X. Wang, D. Hou, M. Hu, F. Ren. Dual-modality emotion recognition based on composite spatio-temporal features[J]. Journal of Image and SGraphics | J Image Graph,2017,22(01):39-48.
（王晓华, 侯登永, 胡敏, 任福继． 复合合时空特征的双模态情感识别[J]． 中国图象图形学报, 2017, 22(01):39-48.）

[14] S. Shafiq, H. Tauseef, M. A. Fahiem, S. Farhan. An algorithm for facial expression based automatic deceptive pain detection[J]. Pakistan Journal Of Science, 2017,69(1), 69-74.

[15] S. Piana, A. Staglianò, F. Odone, A. Verri, A. Camurri. Real-time automatic emotion recognition from body gestures[J]. Computer Science, 2014, 1(1):1-28.

[16] M. Sun, J. Yang, K. Wang, H. Shen. Discovering affective regions in deep convolutional neural networks for visual sentiment prediction[C]// IEEE International Conference on Multimedia and Expo. IEEE, 2016:1-6.

[17] N. Anand , P. Verma. Convoluted feelings convolutional and recurrent nets for detecting emotion from audio data[M]// Technical Report, Stanford University, 2015.

[18] G. Trigeorgis, F. Ringeval, R. Brueckner, et al. Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2016.

[19] N. J. Nalini, S. Palanivel. Music emotion recognition: The combined evidence of MFCC and residual phase[J]. Egyptian Informatics Journal, 2016, 17(1):1-10.

[20] E. Coutinho, B. Schuller. Shared acoustic codes underlie emotional communication in music and speech[J]. PLoS ONE .2017,12(6):e0179289.

[21] G. Cai , B. Xia. Convolutional neural networks for multimedia sentiment analysis[C]// CCF Conference on Natural Language Processing and Chinese Computing, Springer-Verlag New York. Inc. 2015, pp. 159–167.

[22] M. Chen, L. Zhang, X. Yu, Y. Liu. Weighted co-training for cross-domain image sentiment classification[J]. Journal of Computer Science and Technology, 2017, 32(4):714-725.

[23] H. Abburi, M. Shrivastava, S. V. Gangashetty. Improved multimodal sentiment detection using stressed regions of audio[C]// TENCON 2016 - 2016 IEEE Region 10 Conference. IEEE, 2016:2834-2837.

[24] A. Zadeh, R. Zellers, E. Pincus, L.P. Morency. Multimodal sentiment intensity analysis in videos: Facial gestures and verbal messages[J]. IEEE Intelligent Systems, 2016, 31(6):82-88.

[25] S. Poria, E. Cambria, A. Gelbukh. Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis[C]// Conference on Empirical Methods in Natural Language Processing. 2015.

[26] B. Siddiquie , D. Chisholm , A. Divakaran. Exploiting multimodal affect and semantics to identify politically persuasive web videos[C]// ACM on International Conference on Multimodal Interaction, ACM, 2015, pp. 203–210 .

[27] H. Zhou, M. Huang, T. Zhang, X. Zhu, B. Liu. Emotional chatting machine: Emotional conversation generation with internal and external memory[J]. 2017.

[28] Y. Bai, X. Tang, Y. Zhu, D. Cai. A sentiment analysis method for facial expression generation in human-robot interactive communication[C]// International Conference on Virtual Reality and Visualization. IEEE, 2015:97-102.

[29] T. Tsujimoto, Y. Takahashi, S. Takeuchi, Y. Maeda. RNN with Russell's circumplex model for emotion estimation and emotional gesture generation[C]// Evolutionary Computation. IEEE, 2016:1427-1431.

[30] Z. Chen, B. Liu. Topic modeling using topics from many domains, lifelong learning and big data[J]. Icml,2014, 32: 703-711.