Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/18/20230982

Edge Impulse-based automated machine learning for comedy performance evaluation

Shibo Xing

Department of Telecommunications Engineering and Management, Beijing University of Posts and Telecommunications,Beijing, 100032, China

2020213202@bupt.cn

Abstract. The evaluation of comedic performances presents a challenge due to the subjectivity of humor and various performance factors. This paper proposes a potential solution to this problem by utilizing artificial intelligence (AI), particularly automated machine learning (AutoML), to develop an automatic comedy show scoring system using a diverse Chinese acoustic dataset and the Edge Impulse platform. The proposed method involves analyzing complete video recordings of performances by the renowned Chinese comedy troupe, Deyunshe, to assess audience responses and critical acoustic elements that contribute to a comedic performance's success. The study collected five comedic clips, totaling approximately 50 minutes, segmented and labeled them for training purposes. This study systematically explored various configurations using Edge Impulse to optimize accuracy and minimize loss, yielding encouraging results that demonstrate the potential of AI-driven scoring systems. The study identifies several avenues for future research, including enhancing the quality and quantity of training data, refining classification algorithms, and exploring alternative machine learning techniques to further improve accuracy and loss rates. Moreover, the research highlights the potential benefits of integrating real-time scoring systems, which could facilitate audience engagement and refining comedic material. This innovative approach demonstrates the potential to develop a more sophisticated, accurate, and reliable method of evaluating comedic performances, offering valuable insights into audience preferences, enabling the production of high-quality comedic content, and fostering a more objective evaluation of talent in the competitive world of comedy, which could significantly benefit the entertainment industry.

Keywords: comedy quality evaluation, machine learning, AutoDL.

1. Introduction

Comedy has long been a popular form of entertainment, yet quantitatively scoring comedic performances remains a challenging task. A key aspect of comedy is the use of humor, which is often derived from wordplay, satire, irony, absurdity, or observational commentary. It is highly subjective and dependent on individual tastes, cultural context, and shared experiences. This subjectivity makes it challenging to evaluate and score comedy in a consistent and objective manner. Quantitative scoring of comedy is essential for competitions, talent shows, and industry awards. However, it is fraught with challenges due to the subjective nature of humor and the various factors that can influence a performance's comedic value. These factors include the comedian's delivery, timing, audience

DOI: 10.54254/2755-2721/18/20230982

engagement, and the content's relevance to the specific audience.Traditionally, comedy has been scored by human judges, whose evaluations may be influenced by personal biases, cultural backgrounds, and individual tastes. This subjectivity can lead to inconsistencies and unfair scoring, potentially disadvantaging some performers and skewing the results.

Artificial intelligence (AI)-driven approaches offer a promising solution to the challenges of quantitative comedy scoring, which have been successfully employed in various fields, such as natural language processing [1], image recognition [2], and sentiment analysis [3]. These successes suggest that AI could be similarly applied to comedy, allowing for a deeper understanding of the nuances of humor and the factors that contribute to a successful comedic performance. Furthermore, AI models can be continually trained and improved, making them adaptable to changing comedic trends and cultural contexts [4].

Machine learning, a core component of AI, refers to the development of algorithms that enable computers to learn and adapt through experience. These algorithms identify patterns within data and make predictions or decisions without explicit programming [5]. AI has been applied successfully in numerous fields, including medicine, autonomous vehicles, agriculture, and more [6, 7]. In medicine, AI techniques have been used for diagnosing diseases, predicting patient outcomes, and optimizing treatment plans [8]. Autonomous vehicles rely on AI algorithms for navigation, obstacle detection, and decision-making [9]. In agriculture, AI-powered solutions have been developed for crop monitoring, yield prediction, and pest control [10].

Automated machine learning (AutoML) is a subfield of AI that aims to automate the process of selecting and tuning machine learning models for specific tasks [11]. AutoML techniques can search for optimal model architectures, hyperparameters, and preprocessing techniques, reducing the need for expert intervention and accelerating the development of AI solutions.AutoML has garnered significant attention due to its numerous advantages. These include the ability to reduce the time and effort required for model selection and tuning, which has the potential to significantly improve the efficiency of machine learning (ML) workflows. Additionally, AutoML increases the accessibility of AI to non-experts, allowing them to leverage ML techniques without the need for specialized technical knowledge. Furthermore, AutoML presents the possibility of discovering novel model architectures that may outperform human-designed models, leading to potential advancements in the field of AI [12].

This paper suggests using AutoML to create an automatic comedy show scoring system, identifying the best machine learning model and features to quantify humor while potentially discovering novel relationships with other factors. This study collected a diverse Chinese acoustic dataset and utilized the Edge Impulse platform for processing, analyzing, and feature extraction. The platform facilitated efficient experimentation, leading to an optimal method. The results demonstrate the approach's reliability, accuracy, and broader application potential, highlighting the benefits of integrating cutting-edge technology with curated datasets for innovation and discovery.

2. Method

2.1. Overview of the method

The primary objective of this paper is to develop an automatic scoring system for evaluating comedic performances. This innovative approach involves analyzing complete video recordings of performances by the renowned Chinese comedy troupe, Deyunshe, in order to assess audience responses and acoustic elements that are critical to the success of a comedic performance. The proposed data acquisition process involves analyzing audience facial expressions and capturing various auditory elements, including laughter and background noise, to determine the overall comedic impact of the performance.

To accomplish this, this study adopted a triple classification system to manually label the audience's expressions as laughing, non-laughing, or possibly bored. Concurrently, it extracted laughter and background noise from the comic's recording for training purposes. The combination of visual and auditory data facilitated the development of a more comprehensive and accurate scoring system for comedy shows.

DOI: 10.54254/2755-2721/18/20230982

2.2. Dataset preparation

This study meticulously collected five comedic clips, totaling approximately 50 minutes in duration. These clips were then segmented into 5-second intervals, amassing around 500 training samples. The clips containing laughter and applause were assigned a "y" label, while the remaining clips received an "n" label. However, the training set proved insufficient, and merely shortening the clip length to 3 seconds did not effectively resolve the issue. For instance, using 5-second segments could lead to inaccurate segmentation and labeling of clips containing mixed laughter and applause.

In response to these challenges, this study devised an improved method that involved segmenting clips into 3-second intervals while retaining the original manually labeled 5-second dataset. The "no" labels were maintained without alterations, whereas for "yes" labels, Edge Impulse automatically captured the first three seconds. This strategy enabled the researchers to extract the most representative 3-second segments of laughter and applause. Furthermore, they integrated laughter and applause datasets from external sources, labeled them as "yes," and incorporated them into the training process.

2.3. Edge impulse-based comedy show scoring system

The systematic exploration of various configurations using Edge Impulse has paved the way for further research and refinement of the automatic comedy show scoring system. Future studies may focus on enhancing the quality and quantity of training data, refining the classification algorithms, and exploring alternative machine learning techniques to further improve accuracy and loss rates. Additionally, expanding the scope of the research to include other comedy genres and cultural contexts could yield valuable insights into the universality and adaptability of the proposed automatic scoring system.

Another potential avenue for future research is the incorporation of audience feedback, such as surveys or post-show interviews, to supplement the data acquired through facial expressions and audio cues. This additional layer of information could help calibrate the scoring system more accurately by accounting for individual tastes and preferences in humor. Moreover, researchers could investigate the integration of real-time scoring systems, enabling comedians and producers to gauge audience reactions during live performances. Such a system could be particularly beneficial for refining comedic material and facilitating audience engagement, ultimately leading to more successful and enjoyable comedy shows.

Additionally, the exploration of multimodal data fusion, which combines various data sources, such as audio, video, and textual content, could further enhance the accuracy of the automatic comedy show scoring system. For instance, the analysis of textual data, such as transcripts or closed captions, could offer valuable insights into the types of jokes, delivery styles, and pacing that resonate most with audiences. Incorporating these findings into the scoring algorithm would contribute to a more nuanced and robust evaluation of comedic performances.

In conclusion, this study on developing an automatic scoring system for comedy shows, particularly Deyunshe performances, has laid the groundwork for a more sophisticated and accurate method of evaluating comedic success. By systematically exploring various configurations using Edge Impulse and incorporating both visual and auditory data, it has demonstrated the potential of this approach for a wide range of applications in the entertainment industry. As research in this area continues to advance, the development of a robust and reliable automatic scoring system for comedy shows could greatly benefit the entertainment industry by offering valuable insights into audience preferences and facilitating the production of high-quality comedic content.

3. Results and discussion

During the course of the experimentation with Edge Impulse, a systematic investigation of various configurations was conducted to optimize accuracy and minimize loss. The objective was to establish a highly effective approach that would yield valuable insights for future advancements in research. The performance of various models can be observed in Table 1.

DOI: 10.54254/2755-2721/18/20230982

Table 1. The performance of various models.

Type	Testing Accuracy	Testing Loss
2D Training	72.4%	0.55
1D MFCC	81.9%	0.38
1D MFE	84.6%	0.30
1D MFE 200 Epoch	90%	0.22
1D MFE 300 Epoch	90%	0.22

Initially, the 2D training was employed, which produced mixed results. The accuracy rate stood at 72.4%, and the loss was recorded at 0.55. Recognizing the potential for further improvement, this study shifted the focus to 1D Mel-Frequency Cepstral Coefficients (MFCC) configuration. This change led to a notable increase in accuracy to 81.9%, while simultaneously reducing the loss to 0.38. Then, this study continued refining the approach by incorporating Mel Frequency Cepstral (MFE) into the model. This adaptation resulted in a substantial enhancement of accuracy, reaching 84.6%, accompanied by a decrease in loss to 0.30.

When applying 1D MFE training with 200 epochs, leading to a remarkable improvement in performance. The accuracy rate soared to nearly 90%, and the loss was reduced to a mere 0.22. Interestingly, the same level of performance was observed when implementing 1D training with 300 epochs, maintaining an accuracy rate of approximately 90% and a consistent loss of 0.22.

4. Conclusion

This study employed various models based on the EdgeImpulse platform to evaluate the quality comedy. The experimental results demonstrated the effectivenees of the proposed methods. Furthermore, in order to bolster the credibility and impartiality of the experiment, it is acknowledged that the necessity of curtailing manual labeling, opting instead for the machine learning model to independently extract and assimilate crucial features. By adopting this strategy, the risk of human biases and errors that might inadvertently surface can be mitigated during manual annotation, particularly when confronted with intricate auditory nuances that prove difficult to discern by ear alone.

In summation, by embracing advanced regression methodologies and diminishing the reliance on manual labeling, further study will be better positioned to devise a more resilient and accurate comic entertainment evaluation system. Ultimately, these refinements will play a pivotal role in the ongoing progression of the study, ensuring it remains aligned with state-of-the-art research and technological breakthroughs within the field.

References

[1] Hovy E andLavid J 2010 Towards a ‘science’ of corpus annotation: A new methodological challenge for corpus linguistics International Journal of Translation 22(1) 13-36
[2] Krizhevsky A Sutskever I and Hinton G E 2012 ImageNet classification with deep convolutional neural networks In Advances in neural information processing systems pp. 1097-1105
[3] Pang B Lee L 2008 Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval 2(1-2) 1-135
[4] Bostrom N Yudkowsky E 2014 The ethics of artificial intelligence. In Cambridge handbook of artificial intelligence pp. 316-334Cambridge University Press
[5] Bishop C M 2006 Pattern Recognition and Machine Learning Springer

DOI: 10.54254/2755-2721/18/20230982

[6] YuQ Chang C S Yan J L et al. 2019 Semantic segmentation of intracranial hemorrhages in head CT scans2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) IEEE 112-115.
[7] Dhiman G Juneja S Viriyasitavat W et al. 2022 A novel machine-learning-based hybrid CNN model for tumor identification in medical image processing Sustainability 14(3): 1447
[8] Jiang F et al2017 Artificial intelligence in healthcare: past, present and future Stroke and Vascular Neurology 2(4) 230-243
[9] Bojarski M et al 2016 End to end learning for self-driving cars arXiv preprint arXiv:1604.07316
[10] KamilarisA andPrenafeta-Boldú F X 2018 Deep learning in agriculture: A survey Computers and Electronics in Agriculture 147 70-90
[11] Hutter F et al 2019 Automated Machine Learning: Methods, Systems, Challenges Springer
[12] Zöller M A and Huber M F 2020 Survey on automated machine learning Journal of Artificial Intelligence Research 69 421-441