Classifications of Restricted Web Streaming Contents based on Convolutional Neural Network and Long Short-term Memory (CNN-LSTM)

Jaewon Choi* and Xiuping Zhang

Department of Business Administration
Soonchunhyang University, Asan-si, 31538, South Korea
{jaewonchoi, zhangxiuping}@sch.ac.kr

Received: May 30, 2022; Accepted: July 27, 2022; Published: August 31, 2022

Abstract

The development of social media is beneficial for users to quickly access various types of information online. However, this can be a risky for teenagers under the age of 18 years because they may become exposed to information that is unsuitable for them. Some social media platforms have established age-restricted content policies to prevent teenagers from being exposed to bad information. However, unsuitable content that has not been marked as age-restricted still exists online as a result of the enormous volume of information provided on the Internet and the inability to identify it immediately, among other factors. It is important to classify restricted and unrestricted content to protect teenagers' online safety because teenagers are more likely to be negatively affected by biased and harmful content than adults are. We suggest a strategy for classifying restricted and unrestricted content in this study by examining content comments. We collected and cleaned comments obtained from two datasets (each containing restricted and unrestricted content comments, respectively) from YouTube. Word2vec was used to display comments as vectors, and the classifier was established using convolutional neural network and long short-term memory. Through our findings, we hope make the social media environment more secure to protect the physical and mental health of teenagers.

Keywords: Content classification, Convolutional neural network, Long short-term memory

1 Introduction

The practice of sharing and watching online videos via social media has exploded in popularity in recent years [1]. Video watching is becoming an important part of daily life for social media users [2]. YouTube is one of the largest video-sharing platforms, allowing users to upload, view, rate, and share videos [3]. According to the report from statista, the global YouTube user base is expected to reach 2,854.14 million people by 2025. Furthermore, it was shown that the majority of YouTube users in the United States are between the ages of 15 and 25 [4, 5]. It follows from the above that teenagers under the age of 18 years occupy an important position among YouTube users [6].

While teenagers under the age of 18 years benefit from social media by interacting with and learning from others, they are also at risk of exposure to large amounts of objectionable online content [7], 2011). Social media violence and cyberbullying are on the rise. Video-sharing sites are generally regarded as sites that bring online risks to teenagers under the age of 18 years [6].

Journal of Internet Services and Information Security (JISIS), volume: 12, number: 3 (August), pp. 49-62
DOI:*10.22667/JISIS.2022.08.31.049*

*Corresponding author: Department of Business Administration, 22 Soonchunhyang-ro Shinchang, Chungnam, 420-743 Asan, South Korea, Tel: +82-(0)41-530-4280

YouTube is the most used free video-sharing platform globally and the most utilized social media platform. Therefore, we selected YouTube as the object of this study. On YouTube, restricted content refers to content or videos that contain vulgar language, sexual content, gore/violence, or dangerous activity [8]. In order to prevent teenagers under the age of 18 years from accessing restricted video on the YouTube platform, the platform provides regulations to control restricted content as well as employees who are responsible for evaluating content based on user requests, among other measures. However, a sizable proportion of videos that can be categorized as restricted content are still accessible on the website, and many of them are not labeled, making them accessible to teenagers. This is related to the reasons that the quantity of videos uploaded to the platform is expanding on a regular basis [9]. Meanwhile, it has been discovered that restricted videos are incredibly popular, garnering a large number of views and likes. Teenagers are classed as vulnerable groups owing to their lack of life experience, which makes it difficult for them to distinguish right from wrong. They may be particularly vulnerable to the effects of a video and may imitate the inappropriate behavior displayed in it [10, 11]. Video categorization can assist video-sharing platforms in distinguishing unacceptable videos from acceptable videos, as well as in implementing autonomous video management [12].

Comments reflect the feelings, perceptions, and opinions of viewers about the content they watch [13]. Over than 100 million people interact with videos on YouTube every week, according to the survey of YouTube. They do so by rating, sharing, and commenting on videos. The comments of users can be used to explain the content of a video, and can be valid evidence to declare the video quality, perfection, relevance, and popularity [12, 14] Therefore, users' comments on a video can be highly beneficial for video classification. In this study, we classify restricted content based on YouTube video comments using deep learning methods. The following two questions will be the primary focus of our discussion: (1) How effective is deep learning in classifying restricted content? (2) Is hybrid model building using a convolutional neural network (CNN) and long short-term memory (LSTM) effective for restricted content classification? In this study, convolutional neural network (CNN) and long short-term memory (LSTM) hybrid modeling methods is used to explore video classification, which is helpful to understand the role of deep learning hybrid modeling methods in video classification. Meanwhile, the results of this study are beneficial to improve the accuracy of distinguishing harmful videos from harmless videos, and contribute to creating a safe and healthy Internet environment for teenagers.

The following is the structure of the remainder of this study: Section 2 summarizes related work and discusses how this study differs from previous work. Detailed explanations of the methodology and experimental results are provided in Sections 3 and 4. The suggested model and its findings are described and discussed in Section 5. Section 6 offers the final conclusion.

2 Literature Reviews

Previous studies on video classification have found that machine learning can be used to classify videos by constructing classifiers on the metadata and comments on videos [12]. However, the simplicity of ML functions may reduce their generalization ability. This, combined with limited samples and computational units, makes them unable to handle complex classification problems and express complex functions, and restricts the extent of possible classification [15]. Some studies used metadata for videos to develop naïve Bayes and support vector machine (SVM) classifiers for detecting videos that contain racism, racist slurs, and feminism [16]. Neural network technology is widely used for image processing. Detecting pornographic content using motion information and deep learning architectures; using an architecture based on a CNN to detect inappropriate video scenes in video files by analyzing audio and video features; and a fine-grained approach called KidsGUARD1 were proposed to detect videos that contain violent and pornographic scenes [17, 18, 19]. Pornographic content detection was hindered by

the poor resolution, tiny patch support, and variety in appearance [20]. The analysis of video reviews utilizing long short-term memory (LSTM) methods to identify video content that is detrimental to teenagers has not been addressed before, and this is the focus of our work.

With the help of the CNN and the LSTM, we have developed a text classification system that is hybrid model-building-based. The LSTM model was used to preserve historical information and identify contextual dependencies of text after we employed CNN to extract local characteristics of the text.

2.1 Background of the Deep Learning Model

Deep learning replicates how the human brain works in order to create machine learning models with multiple hidden layers, transforms low-level features into higher-level features to represent attribute categories, and reveals scattered features in data. It creates deep network architectures for extracting features efficiently and has good generalization ability [21]. Many models of deep learning, like as deep neural networks (DNN), LSTM, autoencoder (AE), and CNN have been used in different tasks. Furthermore, the deep learning model has shown remarkable success in the processing of natural language for a number of situations because the feature definition work can be decreased and higher performance in terms of accuracy can be achieved [22, 23].

2.2 Word2vec

As a deep learning model, Word2vec is one of many approaches to word embedding that have been proposed, it is capable of converting text to vector format and relocating words into a new space [24]. In other words, the natural language problem is translated into a form that can be understood and processed by machines. Word2ves is able to regulate the dimensions of feature vectors and handle the situation of multiple dimensions. In addition, as compared to other methods of text representation, the word vector created by the word2vec model contains more contextual semantic information, which is beneficial in the training of neural networks, as previously stated [25]. Each word in word2vec is represented as a vector that is concatenated or averaged with the vectors of the other words in the context. The generated vector is utilized to make context-dependent predictions about further words. Word2vec learns vector representations of words using continuous bag-of-words (CBOW) model and continuous skip-gram model. Figure 1 shows the CBOW and skip-gram models. As seen in the Figure 1, the CBOW model predicts the present word based on context, while the skip-gram model predicts the surrounding words based on the present word [26].

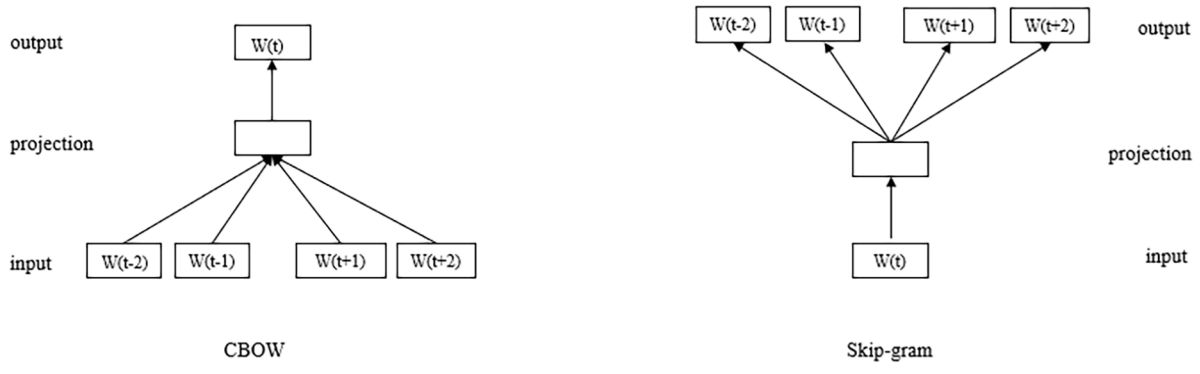

Figure 1: The Method of CBOW and Skip-gram

In previous studies, word2vec was applied to sentiment classification because it can identify the

semantic relationship between words in a document. Zhang et al.[24] utilized word2vec clustering synonyms referencing the same product attributes to verify their ability to extract semantic features, and then SVMperf was used to classify review text. The experimental findings suggest that combining word2vec with SVMperf can achieve satisfactory classification performance. Sun and Chen [27] devised a method for categorizing short texts using word2vec and the Latent Dirichlet Allocation (LDA) topic model in their study. Gibbs sampling was used in their study to train the LDA topic model based on speech weight. The results were vectorized with topic high-frequency words and trained using word2vec, which lengthened the test text. The extended short phrases were categorized using the SVM approach after the features were increased. In accordance with the experimental findings, the proposed strategy in their research has the potential to greatly improve classification performance. According to Xue et al. [28], a novel model based on the semantic orientation pointwise similarity distance (SO-SD) model may be used to assess the emotional inclinations of Weibo posts and other social media communications. They began by segmenting Weibo information into discrete Chinese terms with the use of a word segmentation tool. The word2vec program was trained using a subset of the processed Weibo words, resulting in the production of an extended Weibo emotion lexicon. The distance between words was measured in order to establish which group each word belonged to. The researcher pointed out that word2vec has the potential to capture sentiment information in citations. Jang et al.[29] created a convolutional neural network (CNN) employing two word2vec embedding algorithms, the CBOW model and the skip-gram model, to explore the effect of word2vec on the CNN classification model's outcomes. On real news and tweets, they tested CNN's categorization accuracy using CBOW, skip-gram, and word2vec models. The experimental findings indicated that word2vec considerably enhanced the classification model's accuracy.

2.3 Convolutional Neural Network Model (CNN)

The CNN is a multi-stage trainable neural network architecture developed for classification tasks. Its design was inspired by the human eye system [30]. The CNN was initially used for image processing research and achieved excellent results in this field [20]. Recently, the CNN model was effectively used in text classification [31]. As part of the investigation into the classification of social media communications, Yu et al.[13] used a CNN model on three manually labeled Twitter datasets to examine its effectiveness on the categorization of social media messages. Using the findings, it was discovered that the accuracy of the CNN model was superior to that of SVMs and logistic regression models (LR). Georgakopoulos et al.[32] also proved that CNN was better than traditional text mining algorithms, such as SVMs, naive Bayes, and K-nearest neighbor (KNN), in terms of classification of toxic comments. Wei et al. [33] evaluated the performance of CNN in categorization using real-world data from a variety of legal projects. They compared the effectiveness with an SVM. The findings indicated that the performance of the CNN model was still superior to that of the SVM model, despite the fact that the CNN model was utilized in the experiment without additional optimization. Pei et al. [34] suggested a model named TW-CNN which obtains the semantic characteristics of the short text using LDA and word2vec and then extracts the text features using a CNN for classification. The trials demonstrated that the TW-CNN model has a greater classification accuracy than standard machine learning approaches for text categorization. Furthermore, it has been shown that deep CNN is capable of classifying words without any prior knowledge of the words or the syntactic or semantic structure of a language [35]. This is because the CNN has strong adaptability and can make full use of the convolution filter, which can automatically learn the characteristics suitable for the given task, and no professional knowledge about the structure of the target language is required [36, 23].

The CNN is built primarily of three layers: the input layer, the hidden layer, and the output layer. In the CNN model, the embedding layer is used to address the sparse matrix problem in recommendation

systems. The trained vector values from the embedding layer were sent into the convolutional layer, which extracted the relevant features of the text from the input text. The matrix obtained by the convolution layer was utilized as the input to the polling layer in the next step. The main purpose of the polling layer is to reduce computational complexity. Max-polling is the most used polling technique, which makes use of as much information about the immediate vicinity as possible. The pooling layer outputs the two-dimensional vector, which is flattened and sent to the next layer. Typically, the fully connected layer is the last layer of the CNN. In this study, the CNN model had one embedding layer, three convolutional layers, and one max-polling layer. Figure 2 shows the architecture of the CNN used in this study.

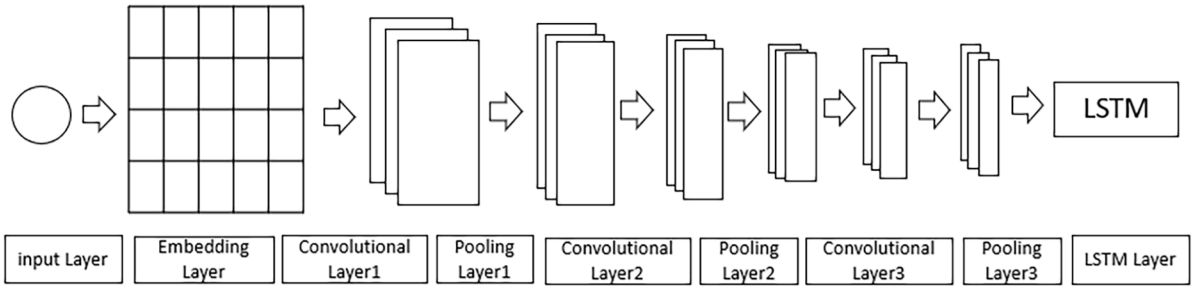

Figure 2: The Architecture of Convolutional Neural Network

2.4 Long Short-Term Memory (LSTM)

The Recurrent Neural Network (RNN) is a kind of neural network that was developed from feed forward neural networks, which are capable of processing variable-length sequences of inputs by using their internal state (memory). Hochreiter [37] proposed the Long Short-Term Memory (LSTM) as a specific sort of RNN. The LSTM seeks to improve the expression of long- and short-time dependency relations, as well as the handling of gradient diffusion and gradient explosion difficulties that are generated by a normal RNN algorithm. The LSTM model has also been shown to be quite powerful in capturing long-distance correlations in sequences of variable lengths, which is another advantage [37]. The LSTM architecture is made up of a cell and three gates: an input gate, an output gate, and a forget gate. Fiure 3 depicts the LSTM architecture.

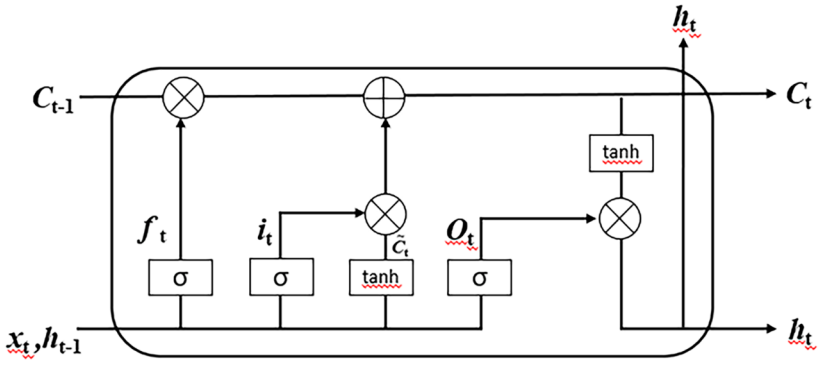

Figure 3: The Architecture of Long Short-Term Memory

The LSTM uses its gates to add, delete, and reset information in each block, which is stored as a

cell state. The importance of new information exceeds the importance of current information [38]. The calculation formula is as follows:

$$i_t = \sigma(W_i h_t - 1 + U_i \alpha_t + b_i) \tag{1}$$

$$f_t = \sigma(W_f h_t - 1 + U_f \alpha_t + b_f) \tag{2}$$

$$\tilde{c}_t = tanh(W_c h_t - 1 + U_c \alpha_t + b_c) \tag{3}$$

$$c_t = f_t * c_t - 1 + i_t * \tilde{c}_t \tag{4}$$

$$o_t = \sigma(W_o h_t - 1 + U_o \alpha_t + b_o) \tag{5}$$

$$h_t = o_t * tanh(c_t) \tag{6}$$

In the equations, σ is the logistic sigmoid function, and $*$ denotes the two vectors' element-wise product. The activation function is denoted by $tanh$. The parameters are represented by the letters W, U, and b. W and U denote the weight matrices between two consecutive layers, while b is the bias between the two successive levels. Furthermore, the input gate, forget gate, cell memory, and output gate are represented by the symbols i_t, f_t, c_t, and o_t , respectively. The input vector at time t is denoted by the symbol α_t. As seen in equation (6), the variable ht not only includes the information represented by α_t, but it also contains the past output state $h_t 1$, which may be used to capture dependencies in a phrase based on previous contextual knowledge.

The LSTM is suitable for processing sequence data because of its unique design structure. Numerous research use it in conjunction with other network architectures to address complicated issues. For the classification challenge of large-scale news texts, Li et al. [39] employed Bi-LSTM-CNN as the learning model. They did not employ a unidirectional LSTM layer to gather context information, but rather a bidirectional LSTM layer. Following that, CNN was utilized to create the contexts for each individual word. The findings revealed that this strategy performs an excellent work to maintain context information while also allowing for a greater variety of word orders. A classification model based on word2vec and LSTM was used to classify patent texts in a study by Xiao et al. [40]. The findings from their analysis show it was discovered that this method outperformed the classification accuracy rates of classification models based on LSTM, KNN, CNN as well as the accuracy rates of models based on CNN and word2Vec. Zhou et al. [41] suggested a cross-language sentiment categorization model based on an attention-based LSTM. This model employs multilingual bidirectional LSTM to represent the source and destination languages' word sequences. Validation was carried out using a benchmark dataset in which Chinese was used as the source language and English was used as the target language, with exceptional results. Liang and Zhang [42] introduced a framework for modeling phrases and documents named AC-BLSTM, which combines an asymmetric convolutional neural network (ACNN) with a bidirectional long short-term memory network (BLSTM) for modeling. When training the model, it is applied to phrases and sentences from the Stanford Sentiment Treebank (SST), although only the sentences are examined during the validation phase. The results demonstrate that the AC-BLSTM model outperforms all other models in the commonly used sentiment classification, question classification, and document classification tasks used by academics.

Astonishing improvements in the area of natural language processing have been made possible by deep learning. Nevertheless, due to CNN's shortcomings in capturing long-term relationships of words in text, some semantic information may be lost, which may result in a reduction in its modeling performance [35]. Furthermore, LSTM may capture long- or short-term relationships, as well as model text sequences of varying lengths. We propose a hybrid model in this study, which incorporates CNN and LSTM to distinguish between restricted and unrestricted content classification.

3 Research Methodology

This study used unstructured data from YouTube as input. After removing stop words and symbols, word2vec converted words into vectors to identify inference between words conveniently. These vector values became the input for the proposed model, which processed the vector values. Finally, accuracy, precision, recall, and F-score were used to verify the classification results. Figure 4 shows the study process.

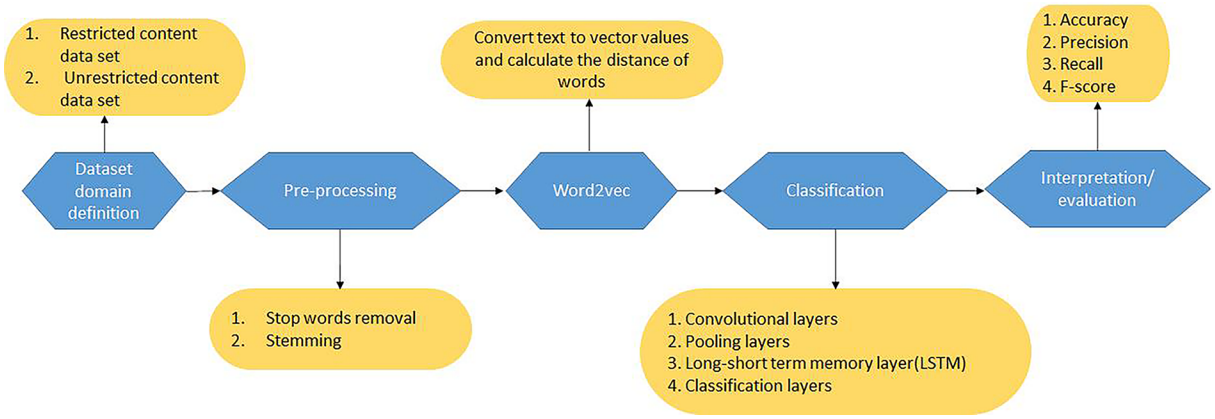

Figure 4: Research Process

3.1 The Proposed Hybrid CNN-LSTM model

A schematic diagram of our hybrid CNN-LSTM model is shown in Figure 5. CNN and LSTM were utilized in this model to get the desired results. We started by converting words into vectors using word2vec, since deep learning is unable to interpret human-written text directly. These vector values were sent into the CNN model as its input data. We used convolutional layers to extract local characteristics from the CNN model's input vector values. Next, the feature vector generated by CNN was used as input for the LSTM model, which was used to extract context-dependent features, and a classifier layer was applied at the end of this process.

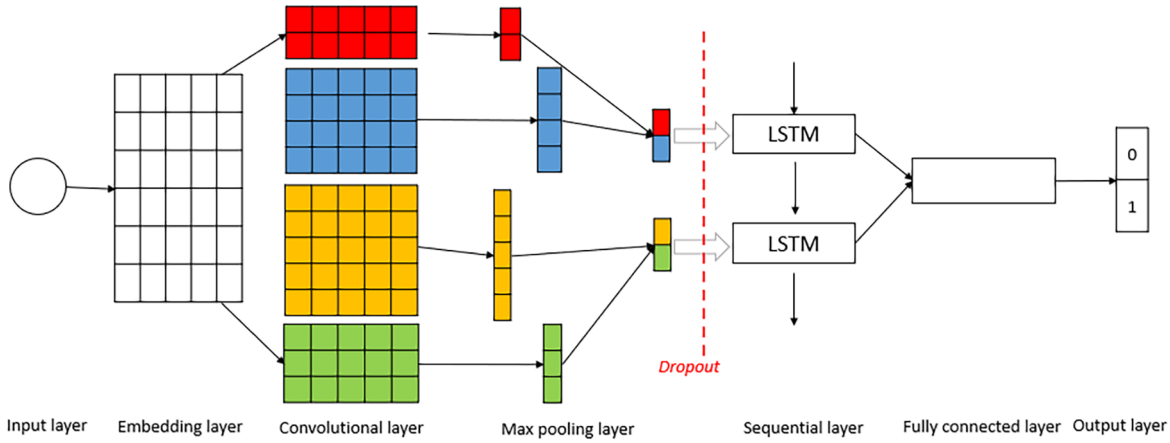

Figure 5: The Architecture of the Proposed Hybrid CNN-LSTM Model

3.2 Embedding Layer

The proposed model's first layer is the embedding layer, which is capable of extracting semantic characteristics from the input and initializing vocabulary word vectors through a pre-trained word vector matrix [43]. The embedding layer is initialized using the embedding extracted from word2vec. We froze the embedding layer to maintain the general meaning of the words. This layer's output is used as the input for the convolutional layers of model.

3.3 CNN Layer

The embedding layer's data is supplied to the CNN module, which extracts local characteristics from it. vi is considered as the d-dimensional word vectors for the ith location of a given word in a headline. Assuming the phrase comprises L words and the CNN's sliding window has a size of k, the word vector in the $jth^{(}jL-1)$ sliding window may be represented as vj,$vj+1$, ..., $vj+k-1$. They can be represented as window vectors, as follows:

$$X_j = [v_i, v_j+1, \ldots, v_j+k-1] \tag{7}$$

where the window vector associated with the current word vj is denoted by the following: $Xj-k+1$, $Xj-k+2$,..., Xj. Then, for each element Yj of the feature map for window vector Xj, the following expression may be used:

$$Y_j = f(X_j \bigodot W + b) \tag{8}$$

where b is the bias, W denotes the convolution kernel, $\bigodot$ is the convolution multiplication, f denotes the activation function, and $relu$ function may enhance the network's learning dynamics and minimize the number of iterations necessary for convergence in deep networks. The $relu$ function is represented by:

$$g(x) = max(0,x) \tag{9}$$

After the convolution layer, max-pooling is used to dramatically reduce the number of features and parameters, as well as the complexity of the computations involved. It is possible to describe the window vector feature matrix as $Y_(j-k+1)$,$Y_(j-k+2)$,...,Y_j , determining the maximum eigenvalue by performing a maximum operation on each row of the matrix. The maximal produced word feature is represented by the following formula:

$$\alpha_j = Max(Y_j-k+1, Y_j-k+2, \ldots, Y_j) \tag{10}$$

3.4 LSTM Layer

However, although CNN can extract local characteristics from text and increase the accuracy of classification, it has limitations when it comes to identifying context dependencies from the text [44]. An exceptionally successful method for collecting long-distance correlations in sequences of variable length has been shown using the LSTM algorithm [37]. The text feature vector produced by the CNN is passed on to the LSTM as input. The following is the formula for calculating the result:

$$i_t = \sigma(W_i h_t - 1 + U_i \alpha_t + b_i) \tag{11}$$

$$f_t = \sigma(W_f h_t - 1 + U_f \alpha_t + b_f) \tag{12}$$

$$\tilde{c}_t = tanh(W_c h_t - 1 + U_c \alpha_t + b_c) \tag{13}$$

$$c_t = f_t * c_t - 1 + i_t * \tilde{c}_t \tag{14}$$

$$o_t = \sigma(W_o h_t - 1 + U_o \alpha_t + b_o) \tag{15}$$

$$h_t = o_t * tanh(c_t) \tag{16}$$

Each of the CNN layer's j output corresponds to an *LSTM* input at time t. The output of the *LSTM* layer is sent into the *softmax* classifier, which then performs classification on the information. The softmax function converts the average of random outcomes into a 0,1 form, and its formula is as follows:

$$P(y = i \mid x, \theta) = e^{o_i} / (\Sigma_{K=1}^{N} e^{o_k}) \tag{17}$$

4 Experiments

It is in this part that the suggested CNN-LSTM model's classification performance is evaluated and the results are analyzed using the training dataset. We will first describe the training and testing datasets, and then we will assess the metrics and outcomes that were obtained.

4.1 Datasets

This study used two datasets from YouTube as corpus. There are 15 categories divided into YouTube channels, and the people and blogs category has been the most deleted and taken down according to ID Africa's survey [45]. We collected comments on videos that were and were not labeled age-restricted in the category of people and blogs. The videos deemed unsuitable for teenagers are tagged with "Age-restricted video (based on Community Guidelines)" on YouTube. First, we manually confirmed the videos that fit our purpose, and then crawled comments. Finally, 81,986 comments were taken from videos labeled as age-restricted, and 77,622 comments were obtained from videos not labeled as age-restricted. The corpus was separated into three portions in this study using a 7:1:2 ratio for the training, verification, and testing sets.

4.2 Experimental Setting

Keras, a prominent Python package, was used to implement the suggested model. In this study, the CNN model consisted of six layers: one embedding layer, three convolutional layers, and three polling layers. Because DNN consists of a large number of parameters, they are prone to overfitting. With the use of dropout, it is possible to randomly remove the feature detector from the network during training and produce less interdependent network elements, resulting in improved performance [46]. Therefore, the dropout technique was used in each convolutional layer to prevent the overfitting of the model. The suggested Hybrid CNN-LSTM model's parameters are listed in Table 1 below, along with their associated values.

5 Results and discussion

On the same corpus, two popular deep learning models (CNN and LSTM) as well as the hybrid model suggested in this work were tested for performance. According to the results, the proposed hybrid model outperforms the other LSTM and CNN models in terms of performance. Model performance was evaluated using conventional evaluation measures such as accuracy, precision, recall, and the F-measure. Figure 6 depicts the outcomes of the study.

The classification system's accuracy is defined as the proportion of correctly classified samples to the total number of samples in each dataset. The accuracy equation can be expressed as $Accuracy = ((TP+TN))/((TP+TN+FP+FN))$. As seen in Figure 6, the hybrid model has an accuracy of 85%,

Table 1: CNN-LSTM Model Parameter Settings

Parameters	Value
Convolutional function	ReLu
Filter windows size	3,5,7
Learning rate	self-adaptive
Dropout	0.2
Embed size	300
Step size	earlystopping
Batch size	50
LSTM hidden vector size	200

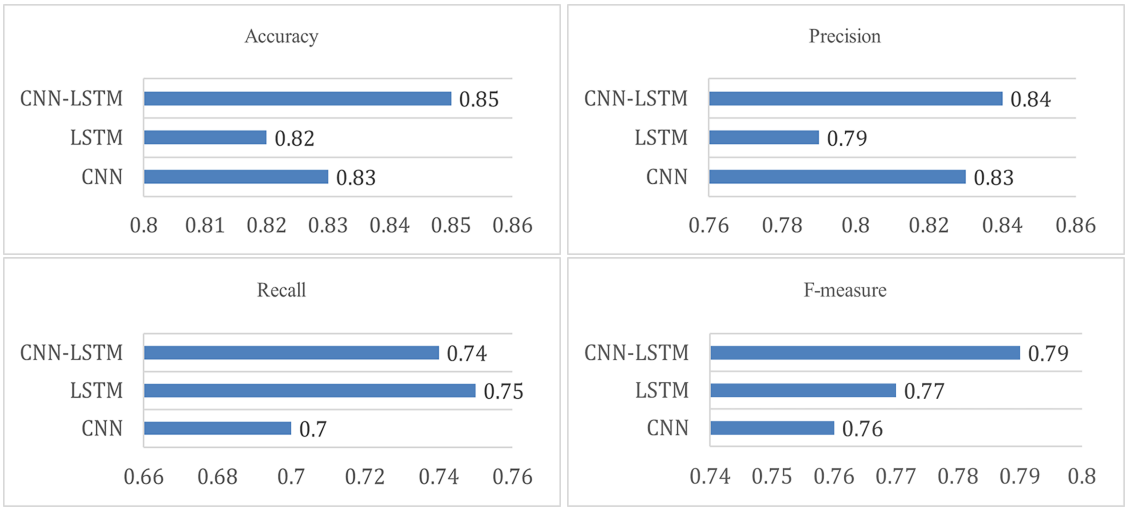

Figure 6: Comparison Results in Accuracy, Precision, Recall and F-measure

which is greater than the accuracy of LSTM and CNN models of 82% and 83%, respectively. Precision means that the positive examples that are judged to be true account for the proportion of all examples that are judged to be true. The equation can be expressed as $Precision = TP/((TP+FP))$. Additionally, the hybrid model's precision was 84%. This is greater than LSTM and CNN models, which have 79% and 83%, respectively. The recall rate was utilized to account for the relationship between the number of positive examples in the classifier and the total number of positive examples. The equation can be expressed as $Recall = TP/((TP+FN))$. Further, the result of the hybrid model reached 74%, that of the typical LSTM model reached 75%, and that of the CNN model reached 70%. The F-measure tries to achieve a better balance between the influence of the accuracy rate and the recall rate, as well as to assess a classifier more comprehensively. Its equation can be expressed as $F-measure = ((2*Precision*Recall))/((Precision+Recall))$. The result of the hybrid model reached 79%, which is higher than the 77% and 76% of typical LSTM and CNN models, respectively. According to the results, the hybrid CNN-LSTM model was proposed to achieve ideal classification performance for classifying restricted and unrestricted content. The hybrid model not only efficiently exploits the sliding window properties of CNN to extract local features, but it also incorporates LSTM to maintain historical information and extract contextual dependencies from text.

6 Conclusion

Technology companies should concentrate on creating a secure online environment for teenagers, who account for the preponderance of social media users, as network technology continues to evolve and social media platforms continue to grow in popularity and growth. We suggested a framework for distinguishing restricted and unrestricted videos in this study using online user comments. The study's findings indicate that the suggested framework may successfully assist social media platforms in identifying content that is inappropriate for teens. This study has the potential to make three contributions. First, this study proposed a framework built using deep learning methods by analyzing and summarizing the theoretical and practical experiences of earlier studies, and the examination of user comments demonstrates the suggested framework's superior performance. This enriches the literature in the field of video classification. Second, in contrast to the practice of using machine learning techniques to analyze comments in earlier studies, this study used a combination of CNN and LSTM to build the framework and compared the classification performance of the framework using a single deep learning technique. The results proved that the framework established by hybrid technology showed excellent classification performance. Third, the results of this study demonstrate that user-generated content can be analyzed to effectively distinguish between restricted and unrestricted videos. Social media platforms must create a safe online environment and improve service quality by analyzing user comments to identify inappropriate content for teenagers in a timely manner.

There are several limitations to this study. First, we only used data from user comments for our study, and future research could consider exploring additional available data types, such as video titles or information about video authors. Second, we only collected data from one category on the YouTube channel. Future research could collect data from other categories to make the dataset more comprehensive. Third, this work focused only on text-based classification algorithms; future research could examine both textual and non-textual aspects for online video categorization.

Declaration of Competing Interest

The authors claim to have no competing financial or personal stakes that could have influenced their work.

Acknowledgements

This study was funded by a research funding from Soonchunhyang University. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2020S1A5A2A01041510)

References

[1] B. Rubenking. Emotion, attitudes, norms and sources: Exploring sharing intent of disgusting online videos. *Computers in Human Behavior*, 96:63–71, July 2019.

[2] A. Ferchaud, J. Grzeslo, S. Orme, and J. LaGroue. Parasocial attributes and youtube personalities: Exploring content trends across the most subscribed youtube channels. *Computers in Human Behavior*, 80:88–96, March 2018.

[3] M. L. Khan. Social media engagement: What motivates user participation and consumption on youtube? *Computers in Human Behavior*, 66:236–247, January 2017.

[4] J. Degenhard. Youtube users in the world 2017-2025, July 2021. https://www.statista.com/forecasts/1144088/youtube-users-in-the-world [Online; Accessed on August 10, 2022].

[5] L. Ceci. Youtube usage penetration in the united states 2020, by age group, March 2022. https://www.statista.com/statistics/296227/us-youtube-reach-age-gender/ [Online; Accessed on August 10, 2022].

[6] S. Livingstone, L. Kirwil, C. Ponte, and E. Staksrud. In their own words: What bothers children online? *European Journal of Communication*, 29(3):271–288, March 2014.

[7] G. S. O'Keeffe, K. Clarke-Pearson, Council on Communications, and Media. The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4):800–804, April 2011.

[8] YouTube. Age-restricted content, 2022. https://support.google.com/youtube/answer/2802167?hl=en [Online; Accessed on August 10, 2022].

[9] C. Southerton, D. Marshall, P. Aggleton, M. L. Rasmussen, and R. Cover. Restricted modes: Social media, content classification and lgbtq sexual citizenship. *New Media & Society*, 23(5):920–938, February 2020.

[10] B. K. Narayanan, M. R. Babu, S. Moses, and M. Nirmala. Adult content filtering: Restricting minor audience from accessing inappropriate internet content. *Education and Information Technologies*, 23(6):2719–2735, May 2018.

[11] T. Janssen, M. J. Cox, M. Stoolmiller, N. P. Barnett, and K. M. ackson. The role of sensation seeking and r-rated movie watching in early substance use initiation. *Journal of youth and adolescence*, 47(5):991–1006, September 2018.

[12] C. Huang, T. Fu, and H. Chen. Text-based video content classification for online video-sharing sites. *Journal of the American Society for Information Science and Technology*, 61(5):891–906, January 2010.

[13] M. Yu, Q. Huang, H. Qin, C. Scheele, and C. Yang. Deep learning for real-time social media text classification for situation awareness–using hurricanes sandy, harvey, and irma as case studies. *International Journal of Digital Earth*, 12(11):1230–1247, February 2019.

[14] K. M. Kavitha, A. Shetty, B. Abreo, A. D'Souza, and A. Kondana. Analysis and classification of user comments on youtube videos. *Procedia Computer Science*, 177:593–598, November 2020.

[15] B. Jang, M. Kim, G. Harerimana, S. U. Kang, and J. W. Kim. Bi-lstm model to increase accuracy in text classification: Combining word2vec cnn and attention mechanism. *Applied Sciences*, 10(17):5841, August 2020.

[16] D. Isa, L. H. Lee, V. P. Kallimani, and R. Rajkumar. Text document preprocessing with the bayes formula for classification using the support vector machine. *IEEE Transactions on Knowledge and Data engineering*, 20(9):1264–1272, September 2008.

[17] M. Perez, S. Avila, D. Moreira, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, and A. Rocha. Video pornography detection through deep learning techniques and motion information. *Neurocomputing*, 230:279–293, March 2017.

[18] J. Mallmann, A. O. Santin, E. K. Viegas, R. R. dos Santos, and J. Geremias. Ppcensor: Architecture for real-time pornography detection in video streaming. *Future Generation Computer Systems*, 112:945–955, November 2020.

[19] J. Cifuentes, A. L. Sandoval Orozco, and L. J. García Villalba. A survey of artificial intelligence strategies for automatic detection of sexually explicit videos. *Multimedia Tools and Applications*, 81(3):3205–3222, January 2022.

[20] S. Gai and Z. Bao. New image denoising algorithm via improved deep convolutional neural network with perceptive loss. *Expert Systems with Applications*, 138:112815, December 2019.

[21] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y.Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected topics in applied earth observations and remote sensing*, 7(6):2094–2107, June 2014.

[22] S. Yilmaz and S. Toklu. A deep learning analysis on question classification task using word2vec representations. *Neural Computing and Applications*, 32(7):2909–2928, January 2020.

[23] D. W. Otter, J. R. Medina, and J. K. Kalita. A survey of the usages of deep learning for natural language processing. *IEEE transactions on neural networks and learning systems*, 32(2):604–624, April 2020.

[24] D. Zhang, H. Xu, Z. Su, and Y. Xu. Chinese comments sentiment classification based on word2vec and svmperf. *Expert Systems with Applications*, 42(4):1857–1863, March 2015.

[25] J. Kim, J. Aum, S. Lee, Y. Jang, E. Park, and D. Choi. Fibvid: Comprehensive fake news diffusion dataset during the covid-19 period. *Telematics and Informatics*, 64(C):101688, November 2021.

[26] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. arXiv:1301.3781. `https://doi.org/10.48550/arXiv.1301.3781`.

[27] F. Sun and H. Chen. Feature extension for chinese short text classification based on lda and word2vec. In *Proc. of the 13th IEEE Conference on Industrial Electronics and Applications (ICIEA'18), Wuhan, China*, pages 1189–1194. IEEE, May 2018.

[28] B. Xue, C. Fu, and Z. Shaobin. A study on sentiment computing and classification of sina weibo with word2vec. In *Proc. of the 2014 IEEE International Congress on Big Data (Big Data'14), Anchorage, USA*, pages 358–363. IEEE, June 2014.

[29] B. Jang, I. Kim, and J. W. Kim. Word2vec convolutional neural networks for classification of news articles and tweets. *PloS one*, 14:e0220976, August 2019.

[30] M. Aydoğan and A. Karci. Improving the accuracy using pre-trained word embeddings on deep neural networks for turkish text classification. *Physica A: Statistical Mechanics and its Applications*, 541:123288, March 2020.

[31] A. U. Rehman, A. K. Malik, B. Raza, and W. Ali. A hybrid cnn-lstm model for improving accuracy of movie reviews sentiment analysis. *Multimedia Tools and Applications*, 78(18):26597–26613, June 2019.

[32] S. V. Georgakopoulos, S. K. Tasoulis, A. G. Vrahatis, and V. P. Plagianakos. Convolutional neural networks for toxic comment classification. In *Proc. of the 10th Hellenic Conference on Artificial Intelligence (SETN'18), Patras, Greece*, pages 1–6. ACM, July 2018.

[33] F. Wei, H. Qin, S. Ye, and H. Zhao. Empirical study of deep learning for text classification in legal document review. In *Proc. of the 2018 IEEE International Conference on Big Data (Big Data'18), Seattle, WA, USA)*, pages 3317–3320. IEEE, December 2018.

[34] K. Pei, Y. Chen, J. Ma, and W. Nie. Short text classification research based on tw-cnn. In *Proc. of the 22nd Pacific Asia Conference on Information Systems (PACIS'08), Yokohama, Japan*, page 41, June 2018.

[35] D. Zhang and D. Wang. Relation classification via recurrent neural network. arXiv:1508.01006, December 2015. `https://doi.org/10.48550/arXiv.1508.01006`.

[36] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998.

[37] S. Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(2):107–116, April 1998.

[38] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, November 1997.

[39] C. Li, G. Zhan, and Z. Li. News text classification based on improved bi-lstm-cnn. In *Proc. of the 9th International conference on information technology in medicine and education (ITME'18), Hangzhou, China*, pages 890–893. IEEE, October 2018.

[40] L. Xiao, G. Wang, and Y. Zuo. Research on patent text classification based on word2vec and lstm. In *Proc. of the 11th International Symposium on Computational Intelligence and Design (ISCID'18), Zhejiang, China*, pages 71–74. IEEE, December 2018.

[41] X. Zhou, X. Wan, and J. Xiao. Attention-based lstm network for cross-lingual sentiment classification. In *Proc. of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP'16), Austin, TX, USA*, pages 247–256. ACL, November 2016.

[42] D. Liang and Y. Zhang. Ac-blstm: asymmetric convolutional bidirectional lstm networks for text classification. arXiv:1611.01884, June 2017. `https://doi.org/10.48550/arXiv.1611.01884`.

[43] A. Rao and N. Spasojevic. Actionable and political text classification using word embeddings and lstm. arXiv:1607.02501, July 2016. `https://doi.org/10.48550/arXiv.1607.02501`.

[44] H. Kim and Y. S. Jeong. Sentiment classification using convolutional neural networks. *Applied Sciences*, 9(11):2347, June 2019.

[45] IDAfrica. 13 youtube facts: The most interesting infographic you'll see today, Ocotober 2015. https://idafricans.com/13-youtube-facts-the-most-interesting-infographic-youll-see-today/ [Online; Accessed on August 10, 2022].

[46] X. Zhang, F. Chen, and R. Huang. A combination of rnn and cnn for attention-based relation classification. *Procedia computer science*, 131:911–917, Janauary 2018.

Author Biography

Jaewon Choi is an associate professor of Business Administration, Global Business School, Soonchunhyang University. His research areas are investigating big data analysis, social network analysis, block chain, personalized intelligent agents in e-commerce and m-commerce. He published papers on Journal of Electronic Commerce Research, International Journal of Electronic Commerce, Cyberpsychology Behavior and Social Networking, and other journals.

Xiuping Zhang is a doctoral student at the Department of Business Administration, Global Business School, Soonchunhyang University, South Korea. Her primary research interests are in the areas of analyzing the impact of new technologies and social networking sites on the customer experience and behavior, electronic commerce and multi-channel retail, online sponsorship and communities, social network analysis, as well as the big data analysis with a particular focus on customer behavior.