Research on attention and reasoning mechanism of autonomous English learning based on machine learning

Shanshan Wang1,*

1 School of General Education, Wuchang University of Technology, Wuhan, Hubei 430223, China

*Corresponding author's Email: 19951013@hbut.edu.cn
*Corresponding author's ORCID: https://orcid.org/0000-0002-1797-2733

Abstract. The wave of the Internet era has swept over our lives and learning, and the Internet is gradually being widely used in all scenes of life and learning. The use of Internet technology can greatly improve efficiency compared to traditional methods. With the popularization of computers, internet technology has entered millions of households, and now more and more school courses are related to the Internet. However, the teaching evaluation of online learning is generally derived from the post-class feedback of teachers and students. In this paper, we analyze the video feeds of students during class and then qualitatively assess students' attention. In this paper, the English grammar error correction model is mainly based on deep learning, and the English grammar error correction evaluation criteria are given. Besides, the SEQ-based English grammar error correction model is proposed, and some improvements are made to the basic SEQ model, and then the algorithmic model is experimented and validated. Based on the above discussion, the feedback mechanism is proposed. Moreover, we also introduce the N-gram model and related implementation techniques, and experiments are conducted and experimental results are given with the corpus.

1. Introduction

In recent years, domestic research about machine learning has developed rapidly, mainly in the extraction and classification of data features [1]. Intelligent applications are realized by algorithmic processing of large amounts of data, such as the use of machine learning for decision making by Alpha Dog, the use of machine learning for driverless cars, the processing of astronomical data using artificial intelligence techniques, and so on. There are also applications in the field of biologies, such as the prediction of DNA sequences, the analysis of the spatial structure of proteins, and the application of financial big data. The prediction of stock futures market can guide investors to better investment. In summary, it can be seen that the research results of machine learning at home and abroad are very considerable. In the 21st century, AI technology is booming, and artificial intelligence is coming into our lives more and more. Machine learning is a very important part of the field of artificial intelligence, and machine learning based on big data technology is becoming more and more important in the field of computer research [2]. A large amount of capital is pouring into this field, which greatly promotes the development of machine learning. Machine learning will gradually feed the field of artificial intelligence and even push the whole human society to undergo great changes [3].

Recently, as research on the interpretability of convolutional neural networks has intensified, researchers have discovered that the potential of convolutional neural networks is more powerful than believed. Researchers have shown that the convolutional units of each layer of a convolutional neural network
can be considered as an independent object detector. However, they do not acquire supervised information about the object location during training [4]. However, although the target localization ability of individual convolutional units is excellent, it is lost when the output of the convolutional neural network is fed into the subsequent fully connected layers of the model for classification. Researchers have proposed methods to minimize the parameters and maintain the high performance target localization of the convolutional neural network by avoiding the use of fully connected layers [5]. The researchers went a step further and pooled the final output of the convolutional neural network globally averaged as shown in Figure 1, before feeding it into the fully connected layer used for classification [6]. They can mark the important regions in the images with the output layer weights by this simple connection structure. Previous work also demonstrated experimentally that this method relying on the localization properties of the convolutional neural network itself can perform preliminary weakly supervised target localization tasks, and can achieve 43.6% localization using the GoogLeNet network with GAP on the ILSVRC2014 dataset accuracy. Moreover, this method also provides a reference for many later weakly supervised target detection methods based on convolutional neural networks [7].

![Figure 1. Data processing under convolutional neural network machine learning](image)

For English learning, learning and practicing grammar is an essential piece of the puzzle. English writing is an effective way to test and improve English grammar, so English learners will do a lot of writing training to improve their English. On the one hand, due to the increasing importance of English, there are more and more learners. While at the same time, the number of English teachers is growing slowly and teaching resources are becoming more and more scarce. Moreover, because it is difficult and time-consuming to correct essays, students have a long period of time to get feedback from teachers after practice, and the longer the feedback period, the less effective it is, which affects the learning effect. In daily teaching, an English teacher usually needs to teach several classes at the same time and has a very heavy teaching load, so it is difficult to spend a lot of energy on reviewing essays. In order to assess the attention of students' attention online, the criteria used to assess the students' visual attention to the screen, mainly in terms of eyes open and eyes closed. This paper presents an analytical study on this issue and verifies that the combined classification of Gabor and SVM has an accuracy of 92.5% and has potential applications. More cross-sectional studies are needed to extend the validity of the research results. In the future, we can analyze students' eye states such as blink frequency by using existing classifiers. We hope to design an effective evaluation system that can be widely used in online learning based on the research in this paper.

2. Target detection of attentional mechanisms

2.1. Attention mechanism
The attentional mechanism is a very common data processing approach in machine learning, which originates from the study of human cognitive ability in cognitive science. In cognitive science, it is believed that there is a bottleneck in human's ability to process information. In order to process the
acquired information in time, human will selectively pay attention to some parts of the information while ignoring other information. For example, human usually reads only a few key words in a sentence, which is why it does not affect the reading when there are several disordered words in a sentence. Therefore, the attention mechanism is essentially a bionic mechanism, a machine simulation of human reading, listening and visual behavior. But theoretically, computers do not need to have an attention mechanism because they have enough memory and computing power to process the acquired information. Researchers want to implement an attention mechanism on computers because if computers cannot simulate the human attention mechanism, they may allow irrelevant information to interfere with the model's computing results, e.g., in the field of natural language processing. For example, if a user enters a typo that causes a semantic ambiguity, it will affect the model's understanding of the text's original meaning. The same statement will have different meanings in different scenarios and different contexts, so in order to make the computer more adaptable to human communication scenarios. So it must learn the ability to pay attention, forget and relate to the context, and this mechanism is the so-called attention mechanism.

The widespread application of the attention mechanism originated from a 2014 paper by the GoogleMind team, which applied the attention mechanism to recurrent neural networks for image classification tasks and achieved excellent results. Based on this, this paper uses a similar attention mechanism on a machine translation task to perform translation and alignment simultaneously with the features of the attention mechanism, and also achieves good results. Their work is also recognized as the first example of applying the attention mechanism to the field of natural language processing. This is also an opportunity for the attention mechanism to be widely used in various natural language understanding tasks based on deep learning techniques. In the field of computer vision, attention mechanisms have been introduced mainly for visual information processing and filtering, which is essentially a method without a strict mathematical definition. Methods such as local image feature extraction, saliency detection, sliding window method, etc., have been widely used in traditional computer vision. And it can be regarded as an attention mechanism in a broad sense, but they all use the computer's powerful computing power to simulate the attention mechanism in a crude way. This so-called "attention mechanism" still requires a large amount of computing resources, and is not a real sense of attention mechanism. With the development of deep learning, attention mechanism is also integrated into deep neural network models, as shown in Figure 2. It is a practical example of researchers applying attention mechanism to image description tasks, from which it is easy to find that attention mechanism can achieve a certain degree of target detection effect without supervised information about the location.

![Figure 2. Schematic diagram of the application of attentional mechanisms in the image description task](image-url)
2.2. Target detection model design for machine learning

Here, we suppose the feature map of the original input image extracted from the convolutional neural network is \( X_F \), where \( n_m \) and \( W \) denote the height and width of the feature map computed by the convolutional neural network, respectively, and \( c \) denotes the feature dimension of each channel in the feature map output by the convolutional neural network. The attention module takes as its input, and then normalizes it in spatial dimension by a 1*1 convolutional layer and outputs a weight map of the attention mechanism as \( A_w \), and then adds the attention mechanism weight map with the feature map and the original feature map to obtain the augmented feature map by spatial normalization and inputs it to the subsequent network module for processing. The weight map of the attention mechanism can be considered as a spatially normalized feature map of the original input image so that the model enhances the relevant regions of the target to be detected and suppresses the irrelevant regions of the target to be detected. Formally, the entire attention module consists of a convolutional layer, a nonlinear activation layer and a spatial normalization layer, represented by the following equation:

\[
\begin{align*}
\hat{A} &= [1 + N(t)] / x_{i,r} \\
E(R) &= \sum A_{s,r} / g \\
P &= E(R) / a
\end{align*}
\]

where \( F \) is a nonlinear activation function. \( W \) and \( c \) are parameters in the 1*1 convolutional layer in the attention module. Combining the above equations, it can be deduced that the feature map \( X_r \) after spatial enhancement is as:

\[
z_{i,j} = F\left(\text{We}^{(l-n_m)r-n_m}x-iWt\right)
\]

It is worth mentioning that conventional convolutional neural networks incorporating attention mechanisms are usually applied to the encoder part of end-to-end deep neural networks simply by using the feature. Also, attention mechanisms can perform key information filtering, and are not concerned with the impact of their output on subsequent decoders. In short, the output of most attention mechanisms is independent of the content in the image, i.e., the attention mechanism outputs the same region of interest regardless of the input image. In this case, the attention mechanism is more concerned with a general distribution of objects, information, and background content in the image.

3. Attentional Reasoning Mechanism of English Independent Learning

3.1. Feature extraction algorithm

The Gabor function is a modification of the Fourier function, which is based on a window Fourier function. The Gabor function can extract relevant data features at different ranges and directions of the data. Besides, the Gabor function is very similar to the response of animal retinal cells to light stimuli. It has excellent properties in extracting spatial information of data, as shown in Figure 3. Gabor function is sensitive to the edges of the image, which can provide good orientation selection and scale selection properties. It is insensitive to changes in luminance, and will not be disturbed by different brightness of the picture. Gabor has excellent spatial localization and orientation selectivity, and can catch the spatial region in each of the partial regions of the image. The Gabor decomposition can be seen as a sniper with high skills. PCA is often used to reduce the dimensionality of a dataset while keeping the most useful features in the dataset. This usually preserves the features we need, which could speed up the processing of the data and reduce the GPU consumption when processing the data. PCA provides an efficient way to reduce the dimensionality of data and is effective for analyzing large amounts of data, such as in the field of face recognition.
3.2. Classification Model for Autonomous English Learning
The human eye pictures in autonomous English learning are extracted according to the location of the pictures. Here the width of the pictures is the size of the human eye pictures, then the grayscale correction of the pictures is performed, and finally, the scaling to 36×36 pixels size is done. The samples are processed, and then the feature data are extracted by the feature extraction algorithm. An image set containing 2000 faces was used for the experimental data. The images were collected from 1000 different people who were learning English independently, of yellow ethnicity and aged between 10 and 50 years old. The image size was 260 × 340 pixels. There were 1000 closed-eye images and 1000 open-eye images. Here, we choose a total of 1000 closed-eye samples and 1000 open-eye samples. We randomly selected 500 closed-eye samples and 500 open-eye samples for English self-learning training, and the remaining 1000 samples for testing. The size of the image is 36×36, and the first row is the closed-eye sample. The second row is the open-eye sample. In the experiment, we compared the above 2 feature extraction algorithms and 6 algorithms with 3 classifier combinations. The results of different algorithms are compared as shown in Figure 4. Through the analysis, we found that the Gabor features outperformed the PCA features in eye feature extraction. All three classifiers showed better prediction accuracy under the same conditions. For example, the classification effect of Gabor+KNN is 2.9%, which is higher than that of PCA+KNN. In addition, by comparison, SVM has better classification properties. For example, for the same Gabor features, SVM outperforms KNN and NB by 25% and 8.3%, respectively. Therefore, we conclude that the Gabor+SVM algorithm can achieve the most efficient classification results for autonomous English learning, with an accuracy of 92.5%.
Besides, KNN is one of the simplest methods in data mining. It is inert learning that delays the local approximation and all computations until after classification, and has a better performance in multi-classification problems. Moreover, KNN can be used not only for classification model problems but also for all kinds of linear regression problems. NBC is a plain Bayesian classification based on Bayes’ theorem. The plain Bayesian classifier is highly scalable and therefore requires parameters whose features and classifiers are linearly related. The idea underlying the plain Bayesian classification is that for the category to be classified, the category is considered to belong. For some types of probabilistic models, in many practical applications, the estimation of the parameters of the plain Bayesian model requires the use of maximum likelihood estimation. Maximum likelihood estimation can be done by evaluating a closed expression without the iterative approximation used by other types of classifiers. SVM is a learning machine that performs binary classification of data in a supervised learning manner and it is used in the classification of patterns, especially in the fields of image recognition and face recognition. It is applicable to high-dimensional small sample problems and performs well.

**4. Conclusion**

Machine learning is based on research in probability theory, statistics, computer vision recognition, data mining, etc. This work summarized these experiences to get experience to make reasonable predictions for the future. Machine learning is based on the above-mentioned pattern of human learning, imitating the pattern of human learning, transmitting a large amount of data to the machine. It can be simulated by the machine to get some relevant conclusions through the idea of induction and summarization. We also find that the idea of machine learning is the simulation of human learning, which can be approximated as a branch of bionics. In order to assess the attention of students in online English independent learning, the assessment criteria utilized the degree of attention of the student's vision to the screen, mainly in terms of eyes open and eyes closed. This paper presents an analytical study on this issue and verifies that the combined classification of Gabor and SVM has an accuracy of 92.5% and has potential applications. We hope to design an effective evaluation system that can be widely used in online learning based on the research in this paper.
References

[1] Huang, Shuguang, and Joseph M. Schimmels. (2018) "Geometric Construction-Based Realization of Spatial Elastic Behaviors in Parallel and Serial Manipulators." IEEE Transactions on Robotics, 34(3):764-780.

[2] Ma, Anqing, et al. (2018) "Construction and Realization of the Estuarine Water Environment Evaluation System Based on SOM." 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), pp. 58-64.

[3] Wang, Guili. (2015) "Construction of College English Autonomous Learning Mode Based on Task-Based Learning under the Network Environment." 2015 International Conference on Social Science and Higher Education, pp. 175-178.

[4] Rui, Zhang. (2015) "Research on University English Autonomous Learning Ability Based on Dynamic Systems Theory." 2015 International Conference on Social Science and Technology Education, pp. 570-578.

[5] Jin, Wu, and Ding Zhirui. (2017) "Research on Mobile Learning Model of College English Based on WeChat Platform." Eurasia Journal of Mathematics, Science and Technology Education, 13(8): 5847-5853.

[6] Hu, Haiqing. (2016) "Teaching Model of College English Using a Computer Network." International Journal of Emerging Technologies in Learning (IJET), 11(8):9-15.

[7] Wang, Xiaowen, and Tianyong Hao. (2017) "Designing Interactive Exercises for Corpus-Based English Learning with Hot Potatoes Software." International Symposium on Emerging Technologies for Education, pp. 485-494.