

Implementation of Cloud Service Platform for English Video Teaching Based on Detailed Content Recognition Algorithm

Chen Xixi

School of Foreign Languages
Leshan Normal University
Leshan, Sichuan, China, 614000

Abstract: The Fusion-SSD network is proposed to extract the semantic concept of the detailed content recognition algorithm. In this network, a multi-branch convolution kernel structure is designed, and the network's ability to retain detailed information is improved by introducing parallel structure convolution kernel branches. Online resources are built on the information-based teaching platform, and teachers organize students to study online before class, the combination of online and offline teaching is implemented in the class, and online learning tasks are assigned to students after class. The whole process of assessment is implemented in all three links. The application of cloud computing services in teaching has gradually become a hot topic of curriculum informatization reform in higher vocational colleges. This paper introduces the current common cloud computing services, and puts forward examples of their use in higher vocational English teaching.

Keywords: Cloud Service Platform, English Video Teaching, Detailed Content Recognition Algorithm

1. INTRODUCTION

In order to promote the development of education informatization in the new era and promote the in-depth integration of information technology and education and teaching, the state has successively issued documents such as the "Ten-Year Development Plan for Education Informatization (2011-2020)" [1] and the "Thirteenth Five-Year Plan for Education Informatization". As an international communication language, English is infiltrating people's life step by step. Whether studying or working, at home or abroad, the frequency that people need to use English is gradually increasing [2]. Therefore, all kinds of English education institutions have emerged as the times require. Education cloud technology has been continuously promoted and applied, especially the distance teaching urban and rural interactive teaching can interact in various forms such as classrooms, teaching research [3], teacher feedback and management, and apply urban teaching resources to improve the teaching level in rural areas. Meet the needs of students of different levels, and then change their learning attitudes [4].

Students learn from each other through group activities in the learning process, understand each other and help each other [5], which is also what higher vocational English teachers need to think about. With the development of information networks, the in-depth advancement of radio and television digitization [6], and the emergence of new media business forms such as mobile TV, mobile TV, online video, and IPTV, the situation of copyright protection and content supervision of digital media has become more and more severe. For example [7], according to the frequency of data access predicts future data access patterns. Only a small number of web proxy cache coverage [8] schemes proposed in recent years have discussed the explicit utilization of semantic information in Web document management. Recommend Bestavros to use more application-level information [9].

Video content detection and recognition is a new problem raised with the explosive development of network multimedia

and various video surveillance applications in recent years. At present [10], the research on video content detection and recognition at home and abroad mainly focuses on two points: 1) Recognition of various features of main objects in videos, especially people [11]. With the development of neural networks, people combine these operators with neural networks to To extract the relevant features in the time domain of the video, Simonyan et al. proposed a two-stream model [12]. The model is a two-branch network structure, and the input of one network is the static RGB image frame of the video. The "cloud interaction" [13] teaching mode is named after "Cloud Interactive English Workplace General Knowledge Exchange", a new form of integrated teaching material for higher vocational public [14] English independently developed by the English teaching team of Hebei Software Vocational and Technical College [15].

The system focuses on the development of the IELTS teaching field [16], and uses an independent streaming media server to provide voice and video services. Users do not need to install any third-party tools in the process of using this system [17]. The features of high performance and high concurrency make the system run more smoothly. You can watch while downloading, which can effectively solve the playback speed of online [18] video and improve user experience. Due to the superiority of streaming media technology, this technology is widely used in video-on-demand and video conferencing in the Internet era. The learning threshold is low [19], and network resources can be accessed anytime and anywhere through the network, integrating existing teaching resources. There are also more mature systems launched, such as video search engines provided by OpenV and Google [20].

In contrast, there are few reports on the method of online real-time content recognition of broadcast video programs, and the related technologies are not yet mature. The author proposes [21] a real-time video content recognition algorithm model. Usually the content cached by the proxy can be organized into a simple data structure. With the help of the idea of the basic

data model in the database system, the main purpose of introducing [22] a hierarchical model structure for Web data is to support different Web data model to enable proxy cache users and administrators to efficiently utilize rich data resources. MEI emphasizes where, focusing on where the movement occurs [23]. It records the accumulated motion energy in the sequence. Let $I(x,y,t)$ be the frame sequence and $D(x,y,t)$ be the binary image sequence, which reflects the motion area, and can be obtained by half a simple frame difference D , then MEI can be obtained by: Multilayer perceptron (Multi layer perception.MLP) is a multi-layer network structure formed by complex connection of a single neuron. Multilayer perceptron is composed of input layer, hidden layer and output layer [24].

2. THE PROPOSED METHODOLOGY

2.1 The Detail Content Recognition Algorithm

The multilayer perceptron does not specify the number of hidden layers, so the appropriate number of hidden layers can be selected according to different needs. The essence of blended teaching is to combine the networked pure technology learning environment with traditional learning, and give full play to face-to-face teaching. And the advantages of online learning. jQuery Easy UI is a collection of jQuery-based user interface plug-ins, which optimizes the ajax interface while providing a series of common UI components for web development, including accordion, combo box, menu, dialog, tabs, tree, validate box, data grid, window, etc.

Firstly, students are divided into groups, and the subject learning activities are carried out in the mode of group cooperative learning. Group division of labor is carried out according to the theme assigned by the teacher, and group members are assigned tasks according to the learning content. Real-time video recognition is more difficult than static video file recognition. It is reflected in the following aspects: because video content recognition needs to accumulate a certain amount of data to draw a conclusion. The data file is a simple disk data file; a series of linked or embedded data file collections are used as a semantic cluster, established in the on top of one or more data files. Color is a visual effect of light through the eyes, brain and our life experience. Human perception of color is not only determined by the physical properties of light, for example, human perception of color is often affected by surrounding colors. Sometimes people also directly refer to the physical properties of substances that produce different colors as colors. Audio signal classification technology is essentially a pattern recognition technology. The basic principle is: first, for a large number of audio signals with known categories, feature extraction and analysis are performed to form reference patterns for various types of audio, and then, for audio signals of unknown categories.

2.2 The English video teaching

Gradient Descent algorithm is a very widely used optimization algorithm in machine learning, and it is also the most commonly used optimization method among many machine learning algorithms. The gradient descent algorithm works by taking the opposite direction of the gradient along the objective function (e) parameter e.e.r. The construction process of the "cloud interaction" mode is shown in Figure 1. The course group does not simply upload the teaching resources to the online teaching platform, but takes the development of the school-based textbook "Cloud Interactive English" as the starting point, and expresses the high-level

goals of the organization or customers through theoretical learning and business needs. Business needs usually come from project investors, customers who buy products, managers of actual users, marketing departments or product planning departments. The business requirement describes why the organization develops a system, that is, the goal that the organization hopes to achieve. These six paths play the recording and broadcasting function and the docking function.

On the one hand, through the study of classroom interactive teaching, sharing teaching resources; Only in order to produce video works of a certain level, the requirements for computer hardware equipment are very high. However, how can the cloud service of video editing provided by the network be used by tens of millions of users at the same time? In the real-time video recognition algorithm model proposed by the author, two important values need to be determined in advance: one is the sample video duration L . If the sample video duration is greater than L , the segment whose front end is L is taken. TF/IDF scheme calculation the weight of each keyword is then indexed.

Suppose k represents a keyword, F represents a document, N_f represents the total number of documents, $k, F()$ search term frequency represents the number of times the keyword k appears in the document FN . In this way, as long as the three-dimensional corresponding to a space-time cube is calculated Integrating the image, the time required to calculate the sum of the pixel values of any cuboid region of any size in the space-time cube is constant, regardless of the size of the calculation region, so that the calculation time will not increase exponentially as the calculation interval becomes larger. The structure of the SSD model is shown in Figure 2-13. The characteristic of the SSD model is to perform location target location regression and category classification on feature maps of different scales.

2.3 The Implementation of English Video Teaching Cloud Service Platform

Because the feature maps of different scales have different receptive fields, different receptive fields mean that they contain different levels of feature information, so targets of different scales can be detected. Teachers upload teaching resources such as course PPT, micro-lectures, and audio to the smart vocational education cloud platform, and students can preview independently and complete audio-visual homework on the vocational education cloud platform. For students' "online" learning situation. Non-functional requirements refer to the features that software products must have in order to meet user business requirements, and are not functional requirements. The non-functional requirements of software products include system performance, reliability, maintainability, scalability, and adaptability to technology and business.

Haoshitong cloud classroom system is safe and stable, video is clear and smooth, real-time data sharing, can support voice question and answer, text question and answer and other question and answer methods; the teacher terminal supports two-way video capture, which can be adjusted freely, remote roll call, video interaction between teachers and students. Video production When teaching, the teacher must first guide the students to write the script. Students can simulate the introduction of exhibitions, booth marketing and other scenarios, or adapt the stories they have read, or directly use the storyline of the original book to write the script.

When a content recognition is started, it will intercept the video segment S with a length of $L+T$ from the current time point forward for content recognition, which contains the newly added video data with a length of T since the last recognition was started. Unless Evaluate component values for more semantic information in a document. Generally, words, words or phrases can be selected as feature values. At present, the representation of text mainly adopts the vector space model (VSM), the basic idea of which is to represent text with vectors.

3. CONCLUSIONS

The blended teaching mode of higher vocational public basic courses. In the course of teaching, the English course group shares online resources, teaching methods and assessment methods, which avoids teachers' independence and realizes the standardized teaching of public English courses in the whole school. This paper also concludes that the calculation amount of the recognition algorithm is inversely proportional to the recognition delay, that is, the better the real-time performance, the greater the amount of computation. The adopted statistical model of color features and texture features is robust to changes in video quality, and the fusion of the two features achieves high video matching accuracy.

4. REFERENCES

- [1] Shang Yuwei, Liu Wei, Su Jian, et al. A whole-process training method and system of machine learning algorithm based on cloud architecture:, 2018.
 - [2] Men Xin. Research on Video Content Recognition and Search Algorithms Based on Deep Learning [D]. Beijing University of Posts and Telecommunications, 2019.
 - [3] Jin Guitao, Shi Yuanbing, Wei Zhong, et al. A Word2vec-based Sensitive Content Recognition Technology [J]. Communication Technology, 2019, 52(11):7.
 - [4] Duan Yucong, Cao Kai. Cross-Boundary, Cross-DIKW Modal, Mesoscale Content Subjective and Objective Semantic Fusion Modeling and Processing Research [J]. China Science and Technology Achievements, 2021(16):5.
 - [5] Gu Xuejing, Wang Xu, Kan Kuo. Research on Hierarchical Correction Method of Leap Motion Gesture Interaction [J]. Computer Engineering and Applications, 2020.
 - [6] Ding Chuanyun, Lan Quanxiang. Spam Content Recognition Based on AC Automata and Bayesian Method [J]. Journal of Heilongjiang Institute of Technology: Comprehensive Edition, 2019.
 - [7] Li Weibin. An English teaching system and its use method: CN112437135A[P]. 2021.
 - [8] Xing Jing, Zhang Peixuan, Zhang Lihua, et al. A cloud platform-based machine vision algorithm training data management system and method: CN111709459A[P]. 2020.
 - [9] Zhang Dongfeng, Zeng Chupeng, Li Junbo. PDF drawing text recognition method, system and equipment: CN111401312A[P]. 2020.
 - [10] Ding Chuanyun, Lan Quanxiang. Spam Content Recognition Based on AC Automata and Bayesian Method [J]. Journal of Heilongjiang Institute of Technology: Comprehensive Edition, 2019, 19(2):4.
 - [11] Zhang Congyue, Yin Ziming, Sun Dayun, et al. Content recognition technology of medical test list based on Tesseract [J]. Beijing Biomedical Engineering, 2019, 38(3):7.
 - [12] Luo Jing. Design of content recognition system based on artificial intelligence technology [J]. Modern TV Technology, 2018(7):4.
 - [13] Zhang Xi. An Intelligent Teaching System Based on Neural Network Algorithm:, 2018.
 - [14] Wan Hongyang, Xu Jie, Zhu Bingyan. Design of spoken English learning system based on cloud computing [J]. Computer and Network, 2019, 45(9):2.
 - [15] Shen Huizhong, Zhang Qian, Wang Bin, et al. A depth image inpainting algorithm based on WMF-ACA [J]. 2021.
 - [16] Gu Xuehui. Violent video content recognition based on information fusion algorithm [J]. Journal of Jinan University: Natural Science Edition, 2019, 33(3):5.
 - [17] Li Chungue, Wang Xinqiang. Dual-target no-reference image block content recognition simulation [J]. Computer Simulation, 2020, 37(11):5.
 - [18] Li Penghui, Jin Xin, Qiu Xipeng. A method and device for document table content recognition based on classification algorithm: CN111209831A[P]. 2020.
 - [19] Fan Xiaojun, Gao Minyu, Hou Rui. A method and system for English text-assisted teaching based on AI vision: CN111610901A[P]. 2020.
 - [20] Cui Changsheng, Song Xuesong. A method for intelligent analysis of communication protocols based on cloud computing: CN106713343B[P]. 2019.
 - [21] Niu Zhendong, Yi Kun, Zhang Chunxia, et al. A multi-granularity content processing method for English paper documents based on pattern recognition: CN112597267A[P]. 2021.
 - [22] Wang Xinyong. An algorithm for automatic classification of video content recognition:, CN111191077A[P]. 2020.
 - [23] Song Shengli, Gao Haichang, Qin Guimin, et al. A Chinese-English Mixed Input Content Recognition Method Based on Keystroke Recording: CN104391589B[P]. 2018.
- You Jingjing, Huang Xianjun, Hou Jin. A method and system for self-service damage determination and collection of vehicle exterior parts based on video: CN112492105A[P]. 2021.