Data acquisition model for online learning activity in distance English teaching based on xAPI

Yuanyuan Wang and Mingqi Wang*

Foreign Language Department,
Qinhuangdao Vocational and Technical College,
Qin Huangdao Hebei 066000, China
Email: henganig@sina.com
Email: yanuigeng@sina.com
*Corresponding author

Abstract: In order to overcome the problems of low accuracy and low recall rate of online learning activity data acquisition, an online learning activity data acquisition model of distance English teaching platform based on xAPI is proposed. According to the visual attribute data of elements, this model divides and judges the online learning activity web page and preliminarily determines the location of noise data. According to the context content rules, the segmentation results of online learning activity web pages are processed in detail, the subblocks are evaluated comprehensively, and the noise data are locked and filtered. The data acquisition model is built by using the modules of user login verification, user mouse track replication, and English online learning activity data acquisition report display. Experimental results show that the model has high recall and precision.

Keywords: distance English teaching; online learning; data acquisition; noise data; user information.


Biographical notes: Yuanyuan Wang received his Master’s degree in English Linguistics and Literature from Yan Shan University in 2012. She is currently an Associate Professor at Qinhuangdao Vocational and Technical College. Her research interests include foreign languages and translation.

Mingqi Wang received his Master’s degree in Foreign Languages and Culture from Yan Shan University in 2011. He is currently a Vice-Professor at Qinhuangdao Vocational and Technical College. His research interests include foreign languages and translation.

1 Introduction

The continuous popularisation of network technology and education informationisation has accelerated the speed of mass learning, and changed people’s thinking and cognitive mode. Online learning activity under distance learning platform have sprung up rapidly,
and gained public recognition at the same time. Online learning is a new mode of education and learning, which will surely speed up the scientific and informational transformation of education. There are many obstacles in the development of online learning, among which the lower completion rate of courses and the higher loss rate of users are the most significant manifestations. In order to improve this situation, the data of online learning activity on the distance learning platform are analysed under the big data environment. The analysis results can provide some guidance suggestions for teachers and managers of distance learning platform (Swart, 2016; Yang, 2019; Du et al., 2018).

Based on the large-scale activity data on the online learning platform, learners can be analysed in an all-round way, including learners’ subordinate categories and learning approaches. For the managers of the platform, the main purpose is to improve the quality of online learning of learners. There are many functions of the current learning platform, but the overall effect of the application is quite different. By analysing the learning data, we can indirectly understand the problems existing in the application process of platform functions, and provide more reliable support for platform management (Li and Li, 2016; Sun, 2018b; Yin and Zhang, 2018). In addition, from the perspective of teaching mode reform and innovation, the analysis of learners’ learning process can customise better learning programs for them. To sum up, it is very necessary to collect data of online learning activity on distance education platform.

Pan et al. (2016) combines the research and development concepts and practices of the two learning platforms as the research basis, collects the data of English online learning activity, analyses the key points and difficulties in English online courses, and dissects the best way of cooperation and innovation between foreign language resources and platform technology. The main purpose is to improve the quality of English online courses construction, but this method can not complete the accuracy of information collection. In Chai and Lei (2018), for the purpose of in-depth study of online learning behaviour, data mining technology is used to analyse online learning activity data. According to Web of Science database, relevant reference data are collected, and the data are summarised and visualised. The general steps of realising online learning activity based on data mining technology are analysed, and at the same time, data mining technology is applied to distance learning platform. The application of online learning activity is clustered, and the related research results are analysed in detail. However, the recall rate of the method for data collection of learning activities is low. Liu et al. (2018) focuses on online teaching data mining, takes students and teachers involved in teaching and learning as objects, combines personalised recommendation method with knowledge in the field of education, and provides direction for intelligent learning based on collected and mined online learning activity data. However, the data acquisition process of this method is more complex. On the premise of big data of education, Shumin (2018) collects, pre-processes and displays data related to students’ resource browsing, participation in activities, homework test scores and so on. On this basis, the relationship between students’ learning behaviour and learning effect is obtained by using big data learning analysis, but the data acquisition coverage of this method is low. Caiyou (2018) collects and analyses data in English teaching by means of e-mining digital classroom, applies data acquisition technology to English teaching classroom, and maximises the visualisation of learning process, which is conducive to the realisation of personalised teaching, but this method has the problem of small amount of data acquisition.
Because the existing methods are not based on the visual attribute data to divide and judge the online learning activity web page, which affects the determination of the location of noise data, it is difficult to lock and filter the noise data. This paper presents a remote English teaching platform based on the xAPI online learning activity data collection model, content rules from context, build data acquisition model of online learning activities, to improve accuracy and recall rate of the remote data collection in teaching English, better improve online learning technology, promote the continuous development of education informatisation.

2 Data acquisition model for online learning activity in distance English teaching based on xAPI

2.1 Noise recognition and filtering of online learning activity web pages

In order to speed up the data acquisition speed and improve the accuracy and recall of online learning activity on the platform of distance English teaching, visual attribute analysis method is introduced to identify and filter the noise in online learning activity web pages. In the process, according to the visual attribute data of elements, the online learning activity web pages are divided and judged, and the location of noise data is preliminarily determined. According to the contextual content rules, the partition results of online learning activity pages are processed in detail, the partition sub-blocks are evaluated comprehensively, and the noise data are locked and filtered.

2.1.1 Noise recognition of web pages

In the process of identifying noise in web pages by combining visual attributes and content rules, the noise information domain with the highest probability of noise data is determined based on the visual attributes data of web pages. Then, the noise data features are analysed and the noise data are filtered in the noise information domain. The following is a detailed process:

The areas where noise data are most likely to occur are identified. Using the visual attributes of web pages to determine the possible location of noise data is mainly based on the following knowledge: the designers of online learning activity web pages usually make planning according to a certain layout in the design of web pages, which makes the structure of web pages more reasonable (Niu, 2018).

According to the above knowledge, we can make use of and study the habits of web designers and users of online learning activity, filter out the noise data in web pages, and lock the edges as the initial location of noise data. Specific operations are as follows: web pages are initially divided into five areas according to visual attribute data to describe the upper, lower, left, right and middle modules of web pages, which are represented $H, F, L, R$ and $C, H, F, Wl$ and $Wr$ represent the threshold in the process of region partitioning, which can effectively adjust the threshold size and change the range of each region. This process uses information such as coordinates of visual attribute data of web pages and can be implemented according to WVP_DOM tree. Web page partition is shown in Figure 1.
Because the size of each online learning activity web page is very different, the threshold is not set in advance, but determined based on the specific size of the web page (Sun, 2018a; Wu, 2018). For example, threshold $H$, if the total height of the web page is $H^*$, and the width is $W^*$, then $H = mH^*$. According to the test, as shown in Figure 2, when $m = 0.1$, it basically covers the top noise and achieves noise filtering. When $m$ is more and more, it will increase the subsequent workload, thus setting $m$ to 0.1, that is, $H$ is $0.1H^*$. In the same way, the final thresholds of $F$, $Wl$ and $Wr$ are $0.5H^*$, $0.4W^*$ and $0.4W^*$, respectively.

Figure 1  Diagram of web page partition

Figure 2  Diagram of $H$ threshold judgement
Figure 3  Noise recognition diagram of online learning activity web page

Start

Enter the HTML document to create the WYP_DOM tree

Determine the noise domain threshold H, F, Wl, Wr based on the page attribute

Process the data block and read in the next data block

Determine whether it is a noise domain based on data block attributes and threshold comparison

Noise domain

N

Y

Calculate the number of punctuation

Calculate hyperlink body text ratio

Both match

N

Y

The data block is noisy and filters it

Last data block

Y

End

N
According to the above content, after determining the area where the noise data appears, the features of the noise data are analysed. As mentioned above, by analysing noise data blocks, some attribute information of noise data blocks can be obtained, such as the number of punctuation symbols Signnum, the number of hyperlinks LinkNum, the length of anchor text LinkTextLength and so on. According to the experience of noise data feature analysis, the features presented by punctuation symbols in semantic blocks can be converted into quantitative forms compared with those presented by hyperlinks. These features have an important impact on determining whether the data block is noise data. Noise data block can be determined based on noise features.

The concrete operation procedure of the above ideas is shown in Figure 3.

2.1.2 Noise filtering

Noise filtering is used for isolation processing of noise data, so as to realise noise data filtering web pages without distortion effect. It is based on WVP_DOM tree.

After defining the purpose, the paper begins to construct a deformable noise filtering method for online learning activity web pages based on WVP_DOM tree’s similarity in distance English teaching platform. According to Figure 3, WVP_DOM tree can organise the whole structure and content of HTML references and the nesting rules based on tags using N-tree structure. The block-level node is the inner node of WVP_DOM tree, and the data information module is the leaf node. If HTML web references are rendered in N-tree mode, it is very easy to find nodes with similarity of some nodes in the reference. The similarity nodes mentioned are usually siblings with similar labels. If the noise data blocks to be filtered also have such similarity nodes, then one of them is filtered out, so that the web page will not be deformed. Based on the above situation, in order to better solve the web page deformation problem, this paper introduces a new solution to determine whether sibling nodes have the same tag or not in the process of retrieving similar structures. Siblings may have information such as label style and script and comment, which will have a negative impact on querying siblings with the same label. Considering that the WVP_DOM tree has filtered such labels, the negative impact of such labels will not be considered (Ye, 2016; Chen et al., 2018; Wang and Mao, 2018). The detailed process of querying similar structures is as follows:

The WVP_DOM tree is traversed and the following judgements are made for each node:

- If (node. noise == 1) && there is no siblings, then the problem is traced back to the parent node and transformed into a similar structure for retrieving the parent node, as shown in Figure 4(a).
- If (node. Noise == 1) && sibling structure has similarity, then the del attribute of node is set to 1 and the node id filtered, as shown in Figure 4(b);
- If (node. Noise == 1) && sibling structure does not have similarity, && all sibling node noise == 1, then go back to the parent node, and the problem is transformed into whether the search parent node has similar structure, as shown in Figure 4(c);
- If (node.noise == 1) && sibling structure does not have similarity, && sibling node has noise! = 1, assuming that the label of the node is a special label, then the del attribute of the node is set to 1 and the node is filtered; otherwise, the node does not have similar nodes to ensure that the page does not deform, and does not delete the node, as shown in Figure 4(d).

Figure 5 is a sketch of deformable web pages after noise filtering.
Figure 4  Similar structure diagram of traversing WVP_DOM tree
2.2 Data acquisition of online learning activity based on xAPI

Under the noise data filtering of the online learning activity web pages mentioned above, the introduction of xAPI for data acquisition can effectively enhance the accuracy and recall of the acquisition results.

The data acquisition tools used in the data acquisition model for online learning activity based on xAPI include user login verification, user mouse trajectory reproduction and data acquisition report display of online learning activity. Details are shown in Figure 6.
In Figure 6, the distance English teaching platform based on xAPI mainly has three functional modules: data acquisition module, e-learning system and learning activity information database. The main function of the data acquisition module is to collect the data of the online learning activities of the client, including user information, page access operation and data submission. Then send the collected information to the learning activity information database; The e-learning system mainly uses the statement processing of converted data information to achieve the extraction of required information, as well as the acquisition of page data of the operation object, and then sends it to the data acquisition module. The main function of the learning activity information database is to receive the data collected by the data collection module and display it in the form of reports.

The following three main modules used in the data acquisition process are analysed separately:

1. **Mouse trajectory tracking**: Mouse tracking can effectively reflect users’ online learning habits and facilitate data acquisition. Because the user’s use of the mouse is not a continuous action, the acquisition of the position information of the mouse only needs to track the mouse trajectory to get the mouse position (Xu et al., 2018; Bai et al., 2018). The detailed process is shown in Figure 7.
In Figure 7, learners log on to the English online learning activity page and begin to use the mouse to operate on the page, such as moving and clicking; according to the JS function, the information data of the client can be gotten and the position of the mouse is recorded; if the learner clicks the button to reproduce the mouse track, the mouse track will be played sequentially.

2 **Representation practice:** The module mainly generates the transformation statements mentioned above. After the user logs in, it collects the basic information of the user in the database. Based on user’s browsing behaviour and page retrieval activity, the specific activity target of user’s operation is determined (Huang and Liu, 2018; Chang, 2017). In order to judge the length of users’ access to online learning activity pages and the selected verbs in the xAPI verb library, the operation is implemented according to the pre-set rules in the Servlet layer. Details are shown in Figure 8.

In Figure 8, when the user logs in to the English online learning activity page, he visits the related activity page in a certain order, and the client obtains the user information and access page information including page attributes and addresses, as well as the length of stay, etc. The Servlet layer receives the collected information, and processes the information based on the settings of xAPI verb library and activity library, which constitutes all the information needed for sentence conversion. The information is submitted to the required database.
Figure 8  Schematic diagram of representing the practice

Data acquisition model for online learning activity

Figure 9  Schematic diagram of online learning analysis
3 Online learning analysis: The construction of this module is based on English learning resources. It mainly analyses the experience and status of the learners. Details are shown in Figure 9.

In Figure 9, when the relevant personnel log in and enter the report view page, the page will send an application to the side with data to retrieve the relevant information and data needed; at this time, the database side will return the information and data needed for the report and display its processing into the form of report.

3 Experiments and discussions

In order to verify the overall effectiveness of the online learning activity data acquisition model of distance English teaching platform based on xAPI, a comparative experiment is conducted. Before the experiment, 1,500 web pages from Tencent, Sohu and other representative websites were selected as experimental sample data, and a group of parameters were randomly selected as initial simulation data to verify the effectiveness of the method in this paper. A data acquisition model of online learning activities on the remote English teaching platform was established on the MATLAB experimental platform. The method in this paper was used as the experimental group, and the methods of Pan et al. (2016), Chai and Lei (2018) and Liu et al. (2018) were used as the control group to realise the analysis of different websites. The comparison indexes of the experiment are the accuracy of data collection and recall rate, among which the accuracy of data collection can directly reflect the reliability of the method and the recall rate can further reflect the performance of the method. Therefore, the above two parameters are selected as the experimental indexes. The higher the accuracy, the better the reliability of the method. Recall rate is the ratio of retrieving relevant data, which is used to measure the recall rate of the retrieval system. The higher the value, the higher the data acquisition rate is.

3.1 Comparison of data accuracy

Firstly, data accuracy detection experiment was carried out to compare the data acquisition accuracy of four different methods under 10 and 20 experiments. If the number of experiments is less than 20, the reliability of the experiment cannot be guaranteed. Too many experiments will cause data deviation. Therefore, it is most reasonable to conduct 20 simulation experiments. Experimental results are shown in Figure 10.

According to the analysis of Figure 10, when the number of experiments is 10 and 20, the data collection accuracy of this method is always higher than that of Pan et al. (2016), Chai and Lei (2018) and Liu et al. (2018) methods. The data collection accuracy of the proposed method is 95%–97%, while that of other reference methods is less than 90%. The higher the accuracy of the method, the stronger the reliability of the proof. This is because the method in this paper deals with the partition results of online learning activity web pages in detail through the context content rules, and comprehensively evaluates the partition sub-blocks to achieve the locking and filtering of noise data, thus improving the accuracy of the method.
Figure 10 Comparison of accuracy rates of different research methods, (a) 10 experiments (b) 20 experiments

3.2 Comparison of recall rate of data acquisition

Then, the data collection recall rate experiments of four different methods were carried out, and the sampling time remained unchanged, which was 10 and 20 times respectively.
The higher the value, the better the precision and the better the performance. Experimental results are shown in Figure 11.

Figure 11  Comparison of recall rates of different research methods, (a) 10 experiments (b) 20 experiments

According to the analysis of Figure 11, the recall rate of online learning activity data acquisition model based on xAPI under different interference items and number of web
Data acquisition model for online learning activity

4 Conclusions

The data collection method of online learning activities in distance English is a hot topic. The accuracy and recall rate of traditional teaching activity data collection methods need to be further improved. Therefore, this paper proposes a data acquisition model of online learning activities in distance English teaching based on xAPI. This model uses the visual attribute data to judge and divide the learning activity web page of the line, according to the context content rules, realises the comprehensive evaluation of the partition sub-block, and completes the locking and filtering of the noise data. xAPI is introduced to build the data collection model of online learning activities and realise the collection of information and data. The experimental results show that the accuracy and recall rate of this method are higher than 95%, which fully proves that this method has good application value. In the following research, the security and system response time of the data acquisition model based on xAPI can be mainly studied to better meet the requirements of online learning activities in distance English teaching.

References


