JIMMA UNIVERSITY

INSTITUTE OF TECHNOLOGY

SCHOOL OF COMPUTIG

Research on Online Learning in Data Analysis

Group members	**IDNO**
1.Michael Andom	RU0726/08
2.MohammedFereje suleman	RU0540/08
3.Mikyas Tadele	RU0517/08
4. Mintesnot Solomon	RU0708/08

ABSTRACT

In this study, on the basis of summarizing the current situation of online learning behavior and related theoretical research. Based on the analysis of related research results, considering the existing problems, the main contents of this paper include the following aspects: (1) Define the connotation of online learning, and introduce the theory of artificial intelligence into the classification of online learning from structural dimension, functional dimension and mode dimension; (2) According to the overall architecture of the analysis model, the analysis model is constructed from left to right and top to down under the big data environment. The online learning data model is constructed from the multi-dimensional and multi-level perspective to determine the source, method and process of data collection. After that, designs the horizontal and longitudinal processes of the online learning analysis model. On this basis, using the big data processing technology on the online learning model in all aspects of the specific algorithms involved in the implementation. In-depth study from the following three aspects: the learning behavior clustering analysis based on K-means algorithm, the individualized course recommendation analysis based on Page Rank algorithm and the correlation analysis of learning effects.

Keywords: big data, online learning platform, online learning behavior, data mining

INTRODUCTION

The rapid development of Internet technology and education information technology has speeded up people's learning and changed the way of thinking and cognitive. Online learning model quickly rise and has been widely recognized. This new way of learning model, will drive the education of information technology reform and innovation.

Online Learning

Online learning takes place in a sequence of consecutive rounds. On each round, the learner is given a question and is required to provide an answer to this question. For example, a learner might receive an encoding of an email message and the question is whether the email is spam or not. To answer the question, the learner uses a prediction mechanism, termed a hypothesis, which is a mapping from the set of questions to the set of admissible answers. After predicting an answer, the learner gets the correct answer to the question. The quality of the learner's answer is assessed by a loss function that measures the discrepancy between the predicted answer and the correct one. The learner's ultimate goal is to minimize the cumulative loss suffered along its run. To achieve this goal, the learner may update the hypothesis after each round so as to be more accurate in later rounds.

As mentioned earlier, the performance of an online learning algorithm is measured by the cumulative loss suffered by the learning along his run on a sequence of question-answer pairs. We also use the term *example* to denote a question-answer pair. The learner tries to deduce information from previous examples so as to improve its predictions on present and future questions. Clearly, learning is hopeless if there is no correlation between past and present examples. Classic statistical theory of sequential prediction therefore enforces strong assumptions on the statistical properties of the input sequence (for example, it must form a stationary stochastic process).

State of the literature

• Research on online learning platforms has been relatively perfect, but the analysis of online learning behavior in large data environments is still in its infancy.

• In theory and practice research also has certain achievements, but there are still some shortcomings, such as the data acquisition platform users online learning behavior lack of pertinence, mostly collect data directly from the database, how to extract the valuable data item is an important basic work of learning behavior analysis. At the same time, the analysis method of online learning behavior analysis model is limited.

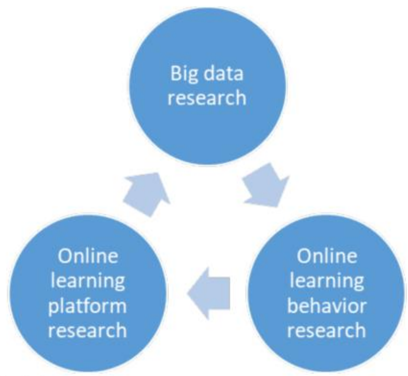

Figure 1. Research status

Characteristics of online learning

Considering the characteristics of online learning behavior within big data environment, therefore, many scholars have explored big data development, online learning platform, online learning behavior, and have obtained rich theoretical and practical results. At present, the study of online learning behavior in big data environment is mainly focused on three aspects, as **Figure 1** shows.

(1) Research on big data. Although the concept of big data was developed early, the development of its technology was in its infancy. So far, the main areas of big data technology include visual analysis, data mining algorithms, semantic engine, data quality and data management. Google's Mapreduce model, for example, focuses on large data sets. In the education area, the University of Purdue uses big data technology to build learning early warning mechanisms by collecting data from students in the course. The main focus of current scholars on big data research is the extraction of value from data sets. To discover knowledge from the data and use it to guide people's

decisions, the data must be analyzed deeply. Social media big data is a hot area of analysis of big data.

(2) Research on online learning platform. Online Learning platform or network Learning platform, network teaching platform, it’s by providing an open, Shared teaching environment, to support the student to carry on the online course Learning, foreign called e - Learning platform. According to the platform for the user object is different, the online learning platform can be divided into two categories, namely for a profit of customization platform and provide free course counseling non-profit platform. Such as Moodle is the world’s most widely used free online learning platform, not only collected from worldwide well-known excellent courses in university, also offers a variety of online communication tools, all kinds of learners for learning. In recent years, the typical representative on the online learning platform is the massive open online course, also known as MOOC.

(3) Research on online learning behavior. A lot of research is involved in the behavior, the research object and the selection of sample has certain limitation, while some studies concerning online learning behavior was conducted under the intervention of the researchers. Kenneth studied the influence of learning behavior and reflective learning in online business courses, Prior has analyzed the impact of online learning behaviors from three aspects: learning attitude, information literacy and self-efficacy, Butcher had studied the different levels of prior knowledge and the relationship between online learning behavior, the results show that a higher level of prior knowledge can lead to learners at a deeper level of learning behavior.

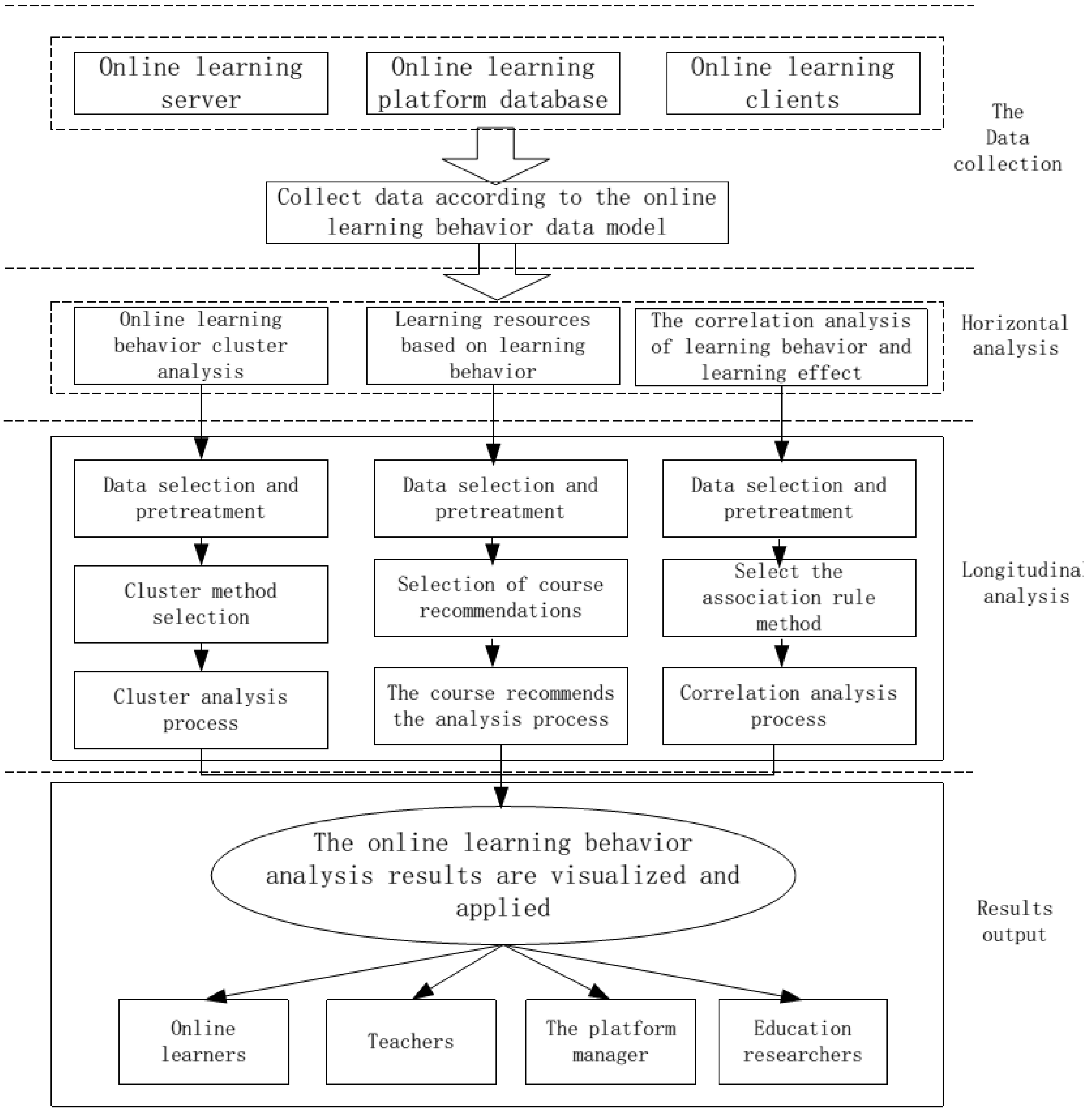

Figure 2. The overall architecture of the online learning behavior analysis model

Implementations

- Vowpal Wabbit: Open-source fast out-of-core online learning system which is notable for supporting a number of machine learning reductions, importance weighting and a selection of different loss functions and optimization algorithms. It uses the hashing trick for bounding the size of the set of features independent of the amount of training data.
- scikit-learn: Provides out-of-core implementations of algorithms for
    - Classification: Perceptron, SGD classifier, Naive bayes classifier.
    - Regression: SGD Regressor, Passive Aggressive regressor.
    - Clustering: Mini-batch k-means.
    - Feature extraction: Mini-batch dictionary learning, Incremental PCA.

Algorithms For online Learning

Introduced a general framework for online convex programming based on the notion of duality. In this will derive various online learning algorithms from the general framework. Algorithms that derive from the framework may vary in one of three ways. First, we can use different complexity functions, $f(\mathbf{w})$, which induces different mappings between the primal and dual variables. we demonstrate the effect of $f(\mathbf{w})$ by deriving the family of quasiadditive algorithms [62] from the framework. We show that our analysis produces the best known mistake bounds for several known quasi-additive algorithms such as the Perceptron, Winnow, and p-norm algorithms. We also use our self-tuning technique from derive a self-tuned variant of the Winnow algorithm for binary classification problems. Second, different algorithms may use different update schemes of the dual variables. We illustrate this fact in by deriving novel online learning algorithms based on a hierarchy of dual ascending methods. The third way in which different algorithms may vary is the form the functions $\{g_t\}$ take.

Quasi-additive Online Classification Algorithms

In this section we analyze the family of quasi-additive online algorithms described in [62, 75, 76] using the algorithmic framework. The family of quasi-additive algorithms includes several known algorithms such as the Perceptron algorithm [95], Balanced-Winnow [62], and the family of *p*-norm algorithms [58]. Quasi-additive algorithms are based on a link function, denoted *L*ink : R*n* → R*n*, and are summarized. We now derive the quasi-additive family of algorithms from our algorithmic framework given. To do so, we choose a complexity function1 $f : S \to$ R so that *L*ink ≡ ∇*f*?. Denote

PARAMETERS: Link function $\mathcal{L}\text{ink} : \mathbb{R}^n \to \mathbb{R}^n$
Learning rate $c > 0$
INITIALIZE: $\boldsymbol{\theta}_1 = \mathbf{0}$
FOR $t = 1, 2, \ldots, T$
Set $\mathbf{w}_t = \mathcal{L}\text{ink}(\boldsymbol{\theta}_t)$
Receive an instance vector $\mathbf{x}_t$
Predict $\hat{y}_t = \text{sign}(\langle \mathbf{w}_t, \mathbf{x}_t \rangle)$
Receive the correct label y_t
IF $\hat{y}_t \neq y_t$
$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \frac{1}{c}\, y_t \mathbf{x}_t$
ELSE
$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t$

	$\mathcal{L}\text{ink}_i(\boldsymbol{\theta})$
Perceptron	θ_i
Winnow	$\exp(\theta_i)$
p-norm	$\frac{\text{sign}(\theta_i)\, \lvert\theta_i\rvert^{p-1}}{\lVert\boldsymbol{\theta}\rVert_p^{p-2}}$

Figure 1 Quasi-additive Online Classification Algorithms

M = *{t* ∈ [*T*] : ˆ*yt* 6= *yt}* to be the set of rounds in which the algorithm makes a prediction mistake. Since the algorithm does not change its vector whenever *t* /∈ *M* we can simply ignore the rounds not in *M*. For a round *t* ∈ *M*, we set *gt*(**w**) = [*γ* − *yth***w***t*, **x***ti*] + , (5.1)

where [*a*] + = max*{*0, *a}* is the hinge function and *γ* > 0 is a parameter. Finally, we use the update scheme given to ascend the dual objective. We now prove the equivalence between the above family of algorithms and the quasi-additive family given in Figure 1. To do so, it suffices to show that for each round *t*, the vector ***θt*** defined in Figure 1 equals to – 1 *c* P*Ti*=1 ***λ****ti*. This claim can be proven using a simple inductive argument. To simplify notation, denote **v***t* = P*Ti*=1 ***λ****ti*. Initially, ***θ***1 = − 1*c***v**1 = 0 and the claim clearly holds. Assume that ***θt*** = − 1 *c***v***t*. If *t* /∈ *M* then ***θ****t*+1 = ***θt*** and **v***t*+1 = **v***t* and the claim holds for the round *t*+1 as well. Otherwise, the function *gt* is differentiable at **w***t* and its gradient at **w***t* is −*yt***x***t*. Thus, **v***t*+1 = **v***t* − *yt***x***t*, which implies that ***θ****t*+1 = ***θt*** + 1*c yt***x***t* = − 1*c*(**v***t* − *yt***x***t*) = − 1*c***v***t*+1. This concludes our inductive argument. To analyze the quasi-additive family of algorithms we note that for *t* ∈ *M* we have *gt*(**w***t*) ≥ *γ*. In addition, the functions *gt*(**w**) are clearly *X*-Lipschitz where *X* = max*t*∈*M k***x***tk ?*. Thus, the regret bound in Corollary 1 implies that *γ* |*M*| ≤ *c f*(**u**) + *X*2 |*M*| 2 *c* + X *t*∈|*M*| *gt*(**u**) . (5.2)

The above inequality can yield a mistake bound if we appropriately choose the parameters c and γ. We now provide concrete analyses for specific complexity functions f. For each choice of f we derive the specific form the update takes and briefly discuss the implications of the resulting mistake bounds.

Perceptron Algorithm

The Perceptron algorithm can be derived from Figure 1 by setting *L*ink to be the identity function. This is equivalent to defining $f(\mathbf{w})$ = 12 $k\mathbf{w}k$ 2 over the domain S = Rn. To see this, we note that the conjugate of f is $f?(\boldsymbol{\theta})$ = 12 $k\ \boldsymbol{\theta}k$ 2 and thus $\nabla f?(\boldsymbol{\theta}) = \boldsymbol{\theta}$. The update $\mathbf{w}t+1 = L$ink($\boldsymbol{\theta}t+1$) thus amounts to, $\mathbf{w}t+1 = \mathbf{w}t$ + 1$c\ yt\ \mathbf{x}t$, which is a *scaled* version of the well known Perceptron update. To analyze the Perceptron algorithm we note that the function $f(\mathbf{w})$ = 12 $k\mathbf{w}k$ 2 is strongly convex over S = Rn with respect to the Euclidean norm $k \cdot k$ 2. Since the Euclidean norm is dual to itself we obtain that each function $gt(\mathbf{w})$ defined in Eq. is X-Lipschitz where X = max$t \in M\ k\mathbf{x}tk$ 2. Setting $\gamma = 1$ in we obtain that $|M| \leq c\ k\mathbf{u}k$ 2 2+ X2 $|M|$ 2 c + X $t \in|M|$ [1 − $yth\mathbf{u}$, $\mathbf{x}ti$] + . (5.3) We next show that since we are free to choose the constant c, which acts here as a simple scaling, the above yields the tightest mistake bound that is known for the Perceptron. Note that on round t, the hypothesis of the Perceptron can be rewritten as, $\mathbf{w}t$ = 1c X $i\in M$:$i<t\ yi\ \mathbf{x}i$. The above form implies that the predictions of the Perceptron algorithm do not depend on the actual value of c so long as $c > 0$. Therefore, we can choose c to be the minimizer of the bound given in, namely, $c = X\ k\mathbf{u}k$ p$|M|$. Plugging this value back into Eq.yields $|M| \leq X\ k\mathbf{u}k$ p$|M|$ + X $t\in|M|$ [1 − $yth\mathbf{u}$, $\mathbf{x}ti$] . Rearranging the above and using Lemma we obtain the following:

Corollary 5 *Let* ($\mathbf{x}$ 1, y 1), . . . , ($\mathbf{x}m$, ym) *be a sequence of examples and assume that this sequence is presented to the Perceptron algorithm. Let* $M = \{t \in [T]$: ˆyt 6= $yt\}$ *be the set of rounds on which the Perceptron predicts an incorrect label and define* X = max$t \in M\ k\mathbf{x}tk$ 2. *Then, for all* $\mathbf{u} \in$ Rn *we have*

$$|M| \leq \text{X}\ t \in M\ [1 - yth\mathbf{u},\ \mathbf{x}ti] + + X\ k\mathbf{u}k\ \text{stX}{\in}M\ [1 - yth\mathbf{u},\ \mathbf{x}ti] + + X2\ k\mathbf{u}k\ 2.$$

Note that this bound is identical to the best known mistake bound for the Perceptron algorithm. However, our proof technique is vastly different. Furthermore, the new technique also enables us to derive mistake and regret bounds for new algorithms such as the ones discussed in.

CONCLUSION AND RECOMMENDATION

Based on the analysis of big data environment based on the main problems of online learning, on the basis of behavioral science and artificial intelligence theory of the concept and classification of online learning behavior is defined, the analysis of the large data online learning behavior under the environment of associated factors, build online learning behavior analysis model, design from behavioral clustering, personalized recommendation and association rule mining process of longitudinal analysis of online learning behavior. The results of this study are as follows: Firstly, from the evolution of artificial intelligence theory and the theory system of relations between the three dimensions, and found it can satisfy the demands of online learning behavior classification dimension choice, from the structure, function and mode of three dimension reveals the behavior of the classification and analysis process, which includes behavior of clustering analysis, recommend analysis and correlation analysis, thus forming a complete and comprehensive.

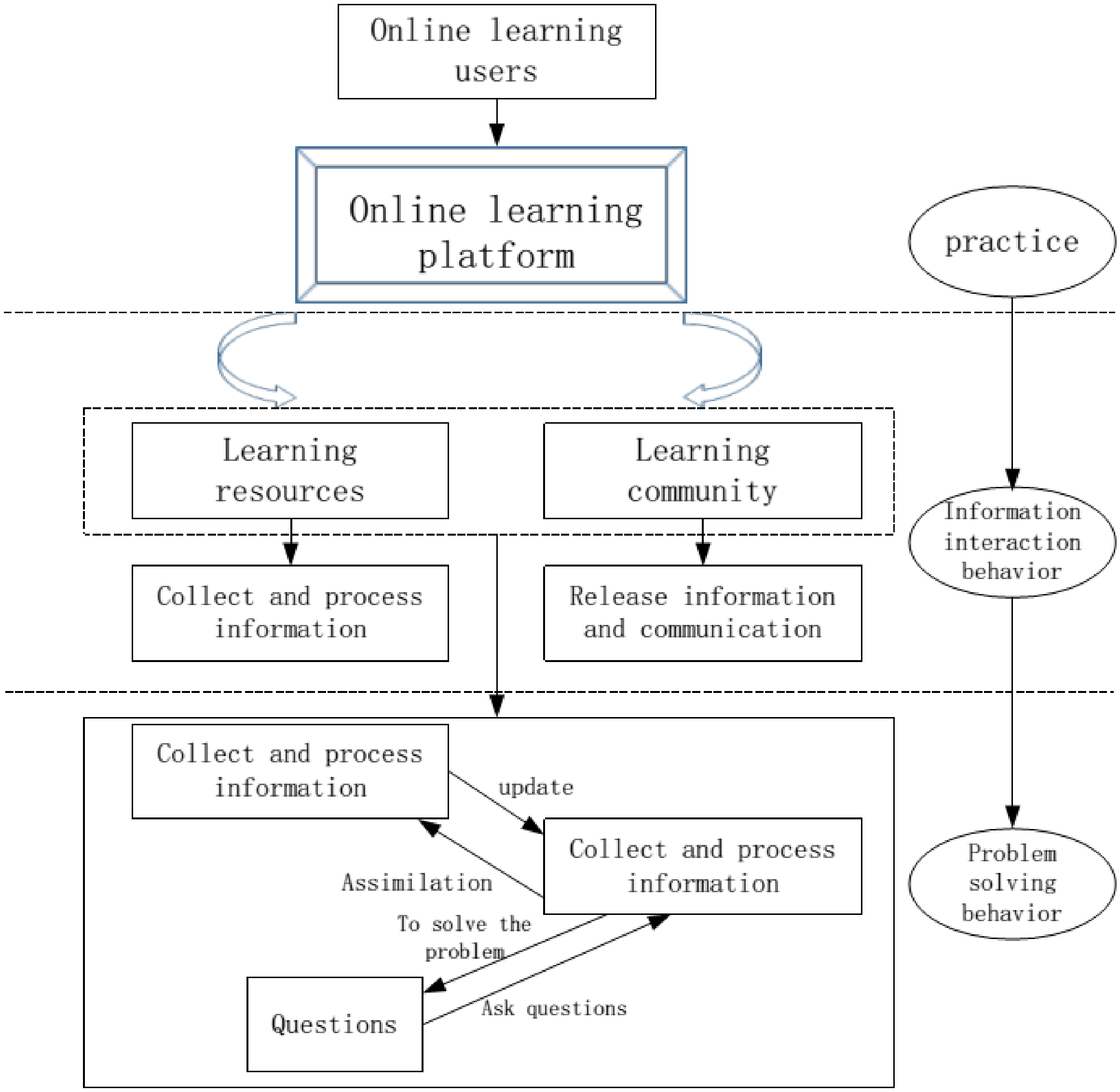

Figure 3. The process of online learning behavior