

An Achievement Prediction Method on the Video Learning Time Based on EKF

Jun Xiao and Hongliang Gu

Abstract—In China, the prediction of the examination achievement is important in the field of online learning, for both the learners and the service providers. On the other hand, the learners' online learning time, e.g. the time of watching the teaching video, can be acquired and recorded easily on most learning platforms. In this paper, we try to establish an EKF model on the exam/test achievement and the learning time, and use the model to predict the achievement in the next exam. The experiment on a two-semester network education course proves, the method can achieve the prediction precision, which is not inferior to the classic methods, and meanwhile own some advantages such as, the capability of noise resistance, uniform convergence, and controllable complexity etc.

Index Terms—Achievement prediction, extended Kalman filter, online learning.

I. INTRODUCTION

Today, the online learning [1], including the mobile learning [2], is becoming more and more popular, due to its advantages over the traditional learning (namely face-to-face learning/classroom learning) mode, such as learning without location limit, materials (e.g. video record) able to be learned (e.g. watched) repetitively etc. The online learning owns the many features different from the traditional learning. One of these features is that, the learners' operation and interaction in the learning platforms (e.g. website), such as watching/download a certain clip of video, testing in a certain exercise and so on, are easier to be acquired and recorded as data in online learning than in traditional learning. Those data is always massive, however, in fact, most of which come into sleep after being acquired, -- not utilized any more, and even not recorded.

On the other hand, the prediction of achievement, e.g. the average score, and the passing/fail rate etc, is very interesting to both the learners and the platform providers (e.g. the teachers, the administration departments, and the organizer etc). For example, if predicting a certain group of students will likely fail in the coming examinations, the teachers or administrator will urge and even compel the students to enhance learning. Nevertheless, the achievement's prediction mostly needs to depend on the learning data record, which is the sleeping data called above in most cases. That is, to a

great extent, the achievement's prediction is the "sleeping" data's mining or reusing of "rubbish" data.

The network education [3]-[5] is a kind of online learning in the field of high diploma (above bachelor degree) education by means of the network media, such as Internet, which are always held/organized by the open universities or network education colleges in respective provinces.

How to realize the achievement's prediction on the basis of the learning data record? The crux is the prediction method, the performance of which determines its popularity. The higher accuracy a prediction method renders, the more popular it is in network education. The more fine-grained a prediction is, more favorite it is. The less realization complexity a prediction method owns, the more intriguing it is.

II. RELATED WORK

Some people have started to study the method of achievement's prediction.

Gerben *et al.* [6] tried to use the Weka classifiers to predict the students' drop-out concerned with the final exam results. Feng *et al.* [7] explored to predict the student's knowledge according to the exercise error record. Li Ping *et al.* [8] studied to use the method of one dimensional linear regression to predict the performance of "University Computer Foundation" course. Hu Shuai *et al.* [9] used the PCA-BPNN method to predict the students' writing performance. Ma Jieming *et al.* [10] researched the effect of the relevance vector machine for performance prediction. Tie Jue *et al.* [11] used the Grey Markov model to predict the sport achievement.

Other people focus on using those methods above of data mining directly to predict the achievement in the MOOCs (Massive Open Online Courses) platform [12], [13], a mode of online learning.

The prediction methods based on classifier, though having outstanding tolerance of uncertainty and noise, rarely consider the accumulation of historical data, due to its belonging to a FIR (Finite Impulse Response) system in essence. Hence, those methods' prediction accuracy will not be promoted evidently with the increasing of the historical data set, e.g. the more and more score record of previous test, which usually will boost the prediction accuracy.

The prediction methods based on regression analysis, though being widely used, have some difficulties in applying. One of those difficulties is hard to find an exact regression model for the achievement trend. Linear model, quadratic model or exponential one or none of them is appropriate? Another problem is that those methods do not take into account the uncertainty and noise.

Manuscript received January 3, 2017; revised April 12, 2017. This work was supported in part by the Science and Technology Commission of Shanghai Municipality research project "Shanghai Engineering Research Centre of Open Distance Education (13DZ2252200)".

Jun Xiao is with Shanghai Open University, Shanghai, China (e-mail: xiaoj@shtvu.edu.cn).

Hongliang Gu was with Shanghai Jiao Tong University, Shanghai, China (e-mail: hlgu@sjtu.edu.cn).

The measures based on neural networks methods have a conspicuous advantage of self-study ability, which can automatically react to the future parameters according to the historical data as a training set. However, the advantage induces the risk at the same time. Namely, the accuracy relies so heavily on the training set that, once the training data has a little deviation, the outcome will deviate very much.

The other approaches based on Markovian, or vector machine, though having a good tolerance with uncertainty and noise, face some problems in applying, such as the difficulty in establishing a precise model, the excessive computing complexity, and so on.

III. OUR METHOD

A. The Model of Our Method

EKF (Extended Kalman Filter) [14] is an IIR (Infinite Impulse Response) filter working in a mode of time iteration, with a controllable complexity, which renders higher accuracy with the increasing of iteration's time. And EKF can endure a zero-average normal-distributing noise in the system. Attributing to those advantages, the filter is widely used in a variety of fields, including location's prediction, and navigation etc.

Being somewhat dissatisfied with the effect of some traditional methods, we try to adopt the EKF method to predict the achievement. In the method, we establish the EKF model, on the basis of the historical examination (including the simulation tests) data. The general model is as follows:

In an online learning platform, we can record the cumulative time which a certain student or a certain group students (e.g. typically a class) spend in watching the teaching video. In the case of group, the time is the sum of cumulative watching time of all members in the group. It is worth explaining that the cumulative time is not about the video but on the watching. That is, if a certain clip of video is watched twice (repetitively), the cumulative time is the twice of the video clip's time. In our method, we take the cumulative watching time as an observation variable of the EKF model, which is notated as t . Though there is some error in acquiring the cumulative watching time, which is called the measurement noise in a professional term, we can reasonably consider the error conforming to a zero-mean Gaussian distribution. So we notate the model's observation equation as follows:

$$Z_k = t_k + V_k \quad (1)$$

Here, Z_k is the k -th observation variable, t_k is the cumulative time in the k -th measurement, and V_k is the k -th measurement noise.

We consider a test/examination score, notated s , in the case of a group, which is an average score, has the following relation to the cumulative learning/watching time s , the knowledge mastery degree from the learning video, notated as m , and the difficulty degree of the test/examination, notated as d .

$$s = t * m * d \quad (2)$$

Here, we also consider that the knowledge mastery degree from the learning video of a certain or group of student(s) conforms to a Gaussian distribution in essence, and so are the difficulty degree of all tests/examinations likewise. So we notate them as $m \sim N(M, \alpha^2)$, and $d \sim N(D, \beta^2)$. As mentioned above, the measurement noise of learning time is also a Gaussian distribution, notated as $\tau \sim N(0, \gamma^2)$.

We define the state vector $X = [s \ t \ m \ d]^T$. According to the equation (2), we get the following relation of the state vector in each iteration calculation:

$$\begin{cases} s_k = t_{k-1} * m_{k-1} * d_{k-1} \\ t_k = t_{k-1} + \Delta t_{k-1} \\ m_k = m_{k-1} + \omega_{k-1} \\ d_k = d_{k-1} + \sigma_{k-1} \end{cases} \quad (3)$$

Here, Δt_{k-1} is the increased learning time between the k -th and $k-1$ -th iteration. s_k is the value of s in the k -th iteration respectively. Similarly, the notations t_k , m_k , m_{k-1} , d_k and d_{k-1} represent the corresponding meaning. The notations ω_{k-1} and σ_{k-1} represent the deviation from the previous iteration of m and d , respectively

According to the equation (3), we can the state equation of our EKF model:

$$X_k = f(X_{k-1}) + \Gamma W_{k-1} \quad (4)$$

$$\text{Here, } f(X_{k-1}) = \begin{pmatrix} t_{k-1} * m_{k-1} * d_{k-1} \\ t_{k-1} + \Delta t_{k-1} \\ m_{k-1} \\ d_{k-1} \end{pmatrix} \text{ and } \Gamma = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (5)$$

$$W_{k-1} = \begin{pmatrix} \omega_{k-1} \\ \sigma_{k-1} \end{pmatrix}^T \quad (6)$$

We define Φ as the Jacobian of the function f , and calculate it out as follows:

$$\Phi = \begin{pmatrix} 0 & m \cdot d & t \cdot d & t \cdot m \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (7)$$

It is proved that, the distribution of the variables in the system noise matrix W , is independent of one another. That is, their covariance is all zero, which satisfies the pre-condition of EKF. According to the distribution features, we can get the following relation: $E(W_k) = 0$ and $E(W_k W_j^T) = Q_k \delta_{kj}$, where δ_{kj} is a Kronecker function. Hence, we get:

$$Q_k = \begin{pmatrix} \alpha^2 & 0 \\ 0 & \beta^2 \end{pmatrix} \quad (8)$$

We rewrite the observation equation (1) into the standard form of EKF as follows:

$$Z_k = H_k \cdot X_k + V_k \quad (9)$$

Here, $H_k = [0 \ 1 \ 0 \ 0]$. According to the distribution features, we can get the following relation: $E(V_k) = 0$ and $E(V_k V_j^T) = R_k \delta_{kj}$, where δ_{kj} is a Kronecker function. Hence, we get

$$R_k = R = \gamma^2 \quad (10)$$

Having owned the system equations above that an EKF model needs, we can list out the steps of iteration calculation as follows, which are divided into 2 stages.

1) Predict (time update) stage.

$$X_k^{(-)} = \Phi X_{k-1}^{(+)} \quad (11)$$

Here, $X_k^{(-)}$ represents the predicted state, and Φ is given in the equation (7), where the parameters t , m and d , are set to the values of corrected state in the previous iteration $X_{k-1}^{(+)}$.

$$P_k^{(-)} = \Phi P_{k-1}^{(+)} \Phi^T + \Gamma Q_{k-1} \Gamma^T \quad (12)$$

Here, $P_k^{(-)}$ represents the prediction error covariance matrix, and the Φ , Γ , Q_{k-1} is explained in the equations(7), (5) and (8).

2) Correct (measurement update) stage.

$$K_k = P_k^{(-)} H_k^T (H_k P_k^{(-)} H_k^T + r)^{-1} \quad (13)$$

Here, K_k is called Kalman gain factor, and H_k is shown in the equation (9).

$$X_k^{(+)} = X_k^{(-)} + K_k [m_k - h(X_k^{(-)})] \quad (14)$$

Here, $X_k^{(+)}$ represents the corrected state, as well as the model's output, the component s of which is the prediction result, namely the achievement in the future test.

$$P_k^{(+)} = P_k^{(-)} - K_k H_k P_k^{(-)} \quad (15)$$

Here, $P_k^{(-)}$ represents the correcting error covariance matrix, which is an output of the equation (12) in the next iteration.

B. The Initial Parameters of Model

The equations (11)-(15) are the whole calculation steps of an iteration. The initial parameters can be given as the following equation shows.

$$\begin{cases} X_0^{(+)} = E[X_0] \\ P_0^{(+)} = \text{var}[X_0] \end{cases} \quad (16)$$

In most cases, neither the expectation nor the covariance matrix of the initial state vector X_0 is known. In fact the estimation of the initial expectation and covariance is of empiricism, to some extent. The teachers having the enough experience will probably estimate them more precisely than those having less experience. However, the worry for the inexperienced persons about rendering a large deviation or error in using the model, is groundless, for due to the stability of EKF, though there will probably emerge some deviation in the beginning iterations, the error covariance will, with the increase of the iteration times, uniformly converge to a stable minimum.

IV. EXPERIMENT AND TESTIFICATION

Here, we test the method in a network education course, Computer Network [15]. We use the method to predict the class' average score in the final exam.

The course has about hundreds of students in each semester/year, most of who study by the online means, including watching the teaching video, fulfilling the tests etc. When a student plays any clip of video in the learning website, the system will record the time data, including the beginning time, the stopping time and the duration etc. Now, we only need to account the sum of duration of the whole class during the given time range.

The course has 7 chapters. The teaching video is divided and organized according to the chapter structure. In each chapter, there is an online exercise test to test the students' grasping level on the chapter's knowledge. And when a student does not fulfill the test, he/she will be not allowed to learn the next chapter, including leaning the video and taking the test. When a student fulfils a test, his/her score will be recorded. Thus, the average score of the whole class in each test will be inferred out easily. The general reviewing video is laid before the final exam, after the last chapter test. We tried our best to make all tests own the same demands as the final examination, such as the same testing time (75min), the same score system (percentage grading system), and the similar question kind and amount, etc. which make the tests and examination conforming to an identical distribution as far as possibly.

When the students fulfill a chapter's test, an iteration calculation will be conducted wholly once, where the test's average score is the component s of the state vector, and the total of video learning duration is the component t . The last iteration, namely after the reviewing video's learning time is acquired and before the final exam is taken, is used for predicting the class' average score in the final exam. Naturally, the actual outcome of the final exam, as a comparison, is used to examine the method's precision.

We applied the method in the course's data for 2 semesters, namely twice, respectively in 2014, 2015. That means that, the same learning materials, including the tests, the final exam paper, and the video etc, are used in the two semesters, except for the different students. The experiment outcome is as follows:

The total time of all 35 clips of teaching videos is over 28 hours. In the Table I, we use the percentage of total learning time, instead of the original learning time, to represent the

cumulative time. This process is for realizing the time's normalization. The percentage is defined as:

$$p_t = \frac{t}{T} * N * 100 \quad (17)$$

Here, t is the cumulative learning time of the students, T is the total time of the teaching video, and N is the student number.

In the Table I, the m and d are the last value of the corresponding components of the state vector. In the experiment, the d is set to 1 in the initial state vector. The reason for this setting is that, we want to use the exam as a comparison base to evaluate the other exams' difficulty degree. Because the same exams are used for the two semesters, naturally their difficulty degree has an identical value. Actually in the experiment, since there is neither other course nor other exams, the role for the d to be a comparison base is not played.

TABLE I: EXPERIMENT OUTCOME

Channels	p_t	m	d	average score		precision	student number
				predicted	actual		
2014	109.31	0.0049	1	72.49	53.30	0.64	222
2015	111.02	0.0051	1	73.02	56.60	0.71	206

Though the iteration time is few, only 8*2 steps totally in the experiment, which depresses the EKF's advantage, more iteration bringing more precise, and our model performs a remarkable precision at last, a precision of 64%~71%. The twice predicted scores are always higher than the actual ones. The cause we guess is that, the (previous) tests' difficulty degree may be not as we imagine, equal to, but less than that of the final exam in fact.

Another interesting phenomenon is that, the prediction's precision of the later semester, 2015 is superior to the former, 2014. The cause we guess is concerned with the initial value of the parameter m . In first round calculation, the initial value of m is wholly an empiric value, which most likely has a remarkable deviation from the actual. As a contrast, in the 2nd calculation (2nd semester), the m 's initial value is set to the output of the first round, which is nearer to the reality than the 1st calculation, due to the feature of uniform convergence of EKF.

V. CONCLUSION AND FUTURE WORK

It can be seen above, though the testing sample in the experiment is limited, the method achieves a notable precision, which is not inferior to those classic methods [16]. Though the iteration time is not large, our method, which converges to a rational value well, renders a compelling resistance against the noise. There are reasons to believe that, with the iteration time's increasing, e.g. in the China network education national examination courses with many simulation tests, the model will achieve a better performance, a better prediction precision.

On the other hand, in the method there is some space for improvement yet. For example, as the equations above, e.g. the equation (2), shows, the model does not take into account

the student's knowledge base, which make the student get a certain part of score without learning any online teaching material. Another issue is about the inconsistency between the exercise tests and the final exam. The recording data shows that most of students achieve a good mark in the exercise tests while performing poorly in the final exam. This phenomenon damage the method's precision in prediction very much. So we look forward to mending those issues in the future work, so as to improve the method.

REFERENCES

- [1] E. H. Perry, "Online learning," *New Directions for Teaching and Learning*, vol. 129, pp.95-104, 2012.
- [2] L. F. Motiwalla, "Mobile learning: A framework and evaluation," *Computers & Education*, vol.49, No.3, pp.581-596, 2007.
- [3] L.-L. Wang and B.-Z. Sun, "The new development trend of network education in the background of Internet+," *Distance Education in China*, vol. 2015, no. 12, pp. 12-17, 2015.
- [4] X.-Y. Pan *et al.*, "Increasing interactivity in blended classrooms through a cutting-edge mobile learning system," *British Journal of Educational Technology*, vol. 39, No 6, pp. 1073-1086, 2008.
- [5] R.-M. Shen *et al.*, "Mobile learning in a large blended computer science classroom: System function, pedagogies, and their impact on learning," *IEEE Trans. on Education*, vol. 52, no. 4, pp. 538-546, 2009.
- [6] G. W. *et al.*, "Predicting students drop out: A case study," in *Proc. Educational Data Mining 2009*, pp. 41-50
- [7] M. Feng *et al.*, "Looking for sources of error in predicting student's knowledge," in *Proc AAAI'05 Workshop on Educational Data Mining*, 2005, pp. 1-9.
- [8] P. Li *et al.*, "The application of one dimensional linear regression in performance prediction," *Computer Knowledge and Technology*, vol. 12, no. 4, pp. 32-36, 2016.
- [9] H. Jiang *et al.*, "Research on the prediction model of students' writing performance based on PCA-BPNN," *Foreign Electronic Measurement Technology*, vol. 2015, no.12, pp. 35-38, 2015.
- [10] J.-M. Ma *et al.*, "Construction of SPOC performance prediction model based on relevance vector machine," *Journal of Suzhou University of Science and Technology (Engineering and Technology)*, vol. 29, no. 1, pp. 77-80, 2016.
- [11] J. Tie *et al.*, "The prediction of Grey Markov model based on Chinese women walking performance," *Mathematics in Practice and Theory*, vol. 2015, no. 3, pp. 187-194, 2015.
- [12] C.-K. He *et al.*, "Learning behavior analysis and prediction of edX platform education data," *China Distance Education*, 2016, no. 6, pp. 27-32, 2016.
- [13] Q.-L. Hao *et al.*, "Research on MOOC learning behavior analysis and performance prediction method," *Electronic Technology and Software Engineering*, no. 7, pp. 167-168, 2016.
- [14] G. Welch *et al.*, "An introduction to the Kalman filter," in *Proc SIGGRAPH 2001*.
- [15] Onlinesjtu. [Online]. Available: <http://course.onlinesjtu.com/course/view.php?id=41>
- [16] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications*, no. 33, pp. 135-146.



Jun Xiao was born in Dec. 1974. He received the P.H. degree in education technology of HuaDong Normal University, China, in 2005. His major field of study includes data mining, education technology etc. He works in Shanghai Open University since 2007. Dr. Xiao is the director of the Open Digital Education Lab.



Hongliang Gu was born in June, 1977. He received the P.H. degree in computer science and technology of Tsinghua University, China, in 2006. His major field of study includes data mining, e-learning etc. He works in Shanghai Jiao Tong University since 2006. Dr. Gu is an associated professor of Shanghai Jiao Tong University.