TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 05/11 pp412–422
DOI: 10.26599/TST.2018.9010110
Volume 24, Number 4, August 2019

# Course Drop-out Prediction on MOOC Platform via Clustering and Tensor Completion

Jinzhi Liao, Jiuyang Tang, and Xiang Zhao*

**Abstract:** As a supplement to traditional education, online courses offer people, regardless of their age, gender, or profession, the chance to access state-of-the-art knowledge. Nonetheless, despite the large number of students who choose to begin online courses, it is easy to observe that quite a few of them drop out in the middle, and information on this is vital for course organizers to improve their curriculum outlines. In this work, in order to make a precise prediction of the drop-out rate, we propose a combined method MOOP, which consists of a global tensor and local tensor to express all available feature aspects. Specifically, the global tensor structure is proposed to model the data of the online courses, while a local tensor is clustered to capture the inner connection of courses. Consequently, drop-out prediction is achieved by adopting a high-accuracy low-rank tensor completion method, equipped with a pigeon-inspired algorithm to optimize the parameters. The proposed method is empirically evaluated on real-world Massive Open Online Courses (MOOC) data, and is demonstrated to offer remarkable superiority over alternatives in terms of efficiency and accuracy.

**Key words:** MOOC platform; drop-out prediction; tensor completion; clustering

## 1 Introduction

With the proliferation of the Internet, Massive Open Online Courses (MOOC), which integrate picture, voice, and flash animations into online courses, significantly influence traditional education while relating to it as a competitor and supplement. Current MOOC platforms, such as Coursera and EdX, offer the opportunity for every person, regardless of their age, gender, or previous educational background, to access state-of-the-art courses and communicate with top lecturers. While hundreds of courses are registered on these platforms on a daily basis, an ever-larger number of students[1] choose to participate.

Nevertheless, what ought not to be neglected is the large number of participants who drop out of on-line educational platforms, since it is common for students to select courses on impulse, only to give up midway through. Instead of focusing on student behavior at course selection, it is advisable for educators to pay more attention to the drop-out behavior, which provides more information about the quality of courses and the preferences of students. In contrast to traditional modes of education, on-line courses such as MOOC are unable to supervise students' learning process, which, to a certain degree, influences the outcome of these courses. Once the system is able to capture students' behavior after selecting a course, educators will formulate better schemes to model the long-term interest of participants and, in turn, the overall quality of courses. Therefore, the ability to predict drop-outs is of urgent significance.

An ideal drop-out prediction model can learn from the history of participant behavior and make more precise predictions. To illustrate, assume a student has already selected "python", "java", and "world economic forum", but has finished only the first two. Then, when the student is planning to select "macro-economics" and

• Jinzhi Liao, Jiuyang Tang, and Xiang Zhao are with National University of Defense Technology, Changsha 410072, China. E-mail: jinzhiliao19@163.com; 1935601196@qq.com; xiangzhao@nudt.edu.cn.
∗ To whom correspondence should be addressed.
Manuscript received: 2018-05-15; accepted: 2018-06-26

"TensorFlow", the model is supposed to provide a drop-out prediction for these two courses. What should be accounted for is the inner connection of the courses. In other words, "python", "java", and "TensorFlow" all belong to the machine learning framework, while "macro-economics" and "world economic forum" can be categorized as economics. Consequently, the prediction model ought to yield a high drop-out probability for the newly selected course "macro-economics", and a lower likelihood of dropping out of "TensorFlow". Figure 1 depicts the aforementioned scenario. Therefore, for the aim of constructing an efficacious drop-out prediction model, the focus is to harness the similarities between students and courses.

In this research, targeted at predicting the drop-out rate of students enrolling in on-line courses, we propose a combinational model of a global tensor and local tensor to describe the inner correlations between students and courses. In contrast to conventional methods, we estimate missing values in a combinational manner, taking into account (1) the global connections between students and all courses, and (2) local connections between students in similar course groups. To enhance the accuracy of clustering, a new similarity calculation method is devised in such a way that it works closely with High-accuracy Low-Rank Tensor Completion (HaLRTC) to provide an accurate drop-out prediction. With the goal of predicting MOOC drop-out, we put forward a method named Mooc drOp-Out Prediction (MOOP).

To summarize, this article makes four key contributions.

- We propose to process drop-out prediction by merging a global tensor and local tensor to express all available feature aspects. To the best of our knowledge, this represents one of the first attempts at this approach.
- A global tensor structure is proposed to model the MOOC data, while a local tensor is clustered to represent course connections. A new similarity estimation method is introduced to enhance the explanatory power of the cluster.
- Drop-out prediction is achieved by adopting a high-accuracy low-rank tensor completion method, equipped with a pigeon-inspired algorithm to optimize the parameters.
- We experimentally evaluate the proposed method on real-world MOOC data, and thereby demonstrate remarkable superiority over alternatives in terms of efficiency and accuracy.

The rest of this paper is organized as follows. Section 2 presents an overview of previous work relating to various prediction methods in the domain of MOOC drop-out. After outlining the necessary background knowledge, the proposed MOOP method is introduced in Section 3, including the specific model and algorithms. Experimental studies are reported in Section 4, followed by a concluding Section 5.

## 2 Related Work

MOOC, as a new approach to education, revolutionize traditional education methods and attract attention both in academic and industrial settings. Although the courses are of huge benefit and convenience, one critical problem that should not be neglected is their high drop-out rates[2–4]. When course organizers arrange their course schedules, the drop-out rate should be given equal, if not greater, attention when compared with the selection rate, since the act of selecting

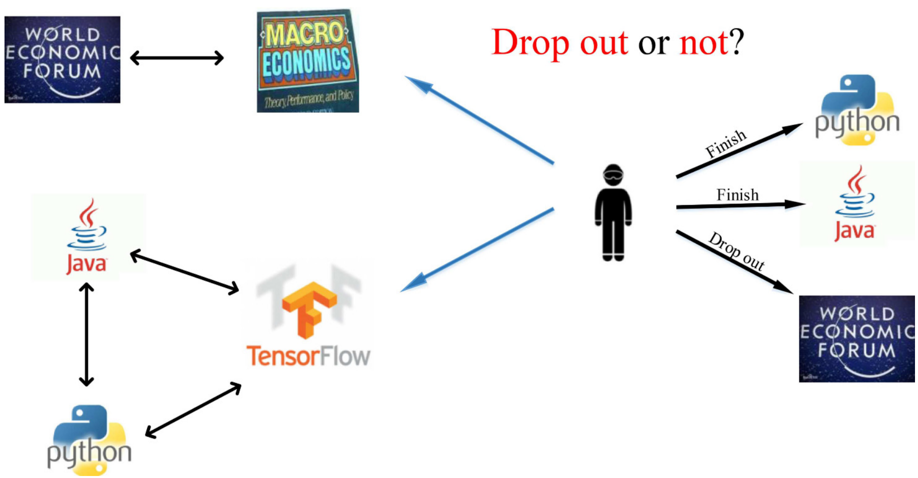



**Fig. 1 Sketch of course drop-out problem.**

does not guarantee long-term persistence with study. Consequently, a large number of methods have been devised to deal with the task of predicting MOOC drop-outs.

Traditionally, statistical techniques have been widely utilized to explore the correlations between the drop-out rate and other factors, with the aim of working out the determinant factors of success in schooling. Some previous studies discovered that university entry scores were a good index to forecast student performance[5], since it is obvious that the higher the entry score, the better equipped the student is for study. However, Thomas et al.[6] found that entry score actually had a poor correlation to performance in some specific domains, such as mechanics courses, and that these scores were not easily accessible, since many academies did not arrange other examinations to test the students' professional skills. The common weakness of these factors lies in the uncertainties inherent in different situations. Specifically, the differences between the age might determine degree of education and the institution[7], and gender plays an important role in engineering and physical sciences, where the proportion of female students is very low. Motivation has been generally shown to correlate positively with student success and retention[8]. The drop-out rate is also considered an aspect of the analytic framework; it is used, for example, to predict the performance of participants in online classes[9] and to identify predictors of academic persistence in distance education[10]. Although the addition of drop-out rates expands the statistical methods, the intrinsic disadvantages of tackling high-dimensional data restrict the development of successful methods.

Based on its remarkable performance in prediction and classification of context analysis, Natural Language Processing (NLP) is also promoted as a method for the analysis of MOOC data. There are many methods based on NLP aimed at investigating elements of MOOC that are unrelated to student success. Elouazizi[11] used the linguistic aspects of point-of-view as an indicator of cognitive presence to construct an exploratory framework for examining students' learning-based inquiry in the MOOC context. Focusing on finding the leaders in MOOC discussion forums, Moon et al.[12] proposed a method to measure language accommodation which represents the students' choices of words given a specific theme. Several NLP methods are adopted by Crossley et al.[13] to compare the performance of different languages in forum posts to evaluate which one is predictive of MOOC completion. Wen et al.[14] tested three courses using a sentiment analysis approach to verify the significant correlation between the sentiments expressed in the course forum posts and the number of students who drop out of the course. Moreover, in Ref. [15], two approaches are devised to quantify the engagement, which is validated on three courses with different topics. In particular, the motivation and engagement attitudes contained in course forums are identified to achieve this task.

Based on its success in other domains, researchers have gradually began to apply neural networks to predict MOOC drop-outs. An artificial neural network was initially put forward to predict the drop-out rate, and achieved a high level of accuracy and efficiency[16]. In order to further improve the accuracy, a multi-layer network was proposed, upon which a MOOC drop-out predictor was constructed by utilizing the collected data[17]. Additionally, drop-out prediction was viewed as a sequence classification problem and a temporal model was utilized. Among the methods, long short-term memory networks yield the best results. Sinha et al.[18] paid attention to students' behavior when they interacted with MOOC video lectures, to capture behavioral patterns in student activity. They then utilized these patterns to evaluate students' information processing index. In a following work[19], they constructed a graph to capture the sequence of active and passive learning activities, and used graph metrics as features for predicting attrition.

With the increasing popularity of machine learning techniques, many other attempts have been made to handle the problem. Kloft et al.[20] proposed a machine learning method based on support vector machines for predicting drop-outs, where the MOOC data of the current and previous weeks were learned to estimate the missing data in the following week. Nagrecha et al.[21] combined feature engineering, data preprocessing, and metrics as parameters in the context of interpretability to construct a decision tree to accomplish drop-out prediction. Hidden Markov models were employed by Balakrishnan and Coetzee[22] to help predict student retention as well as to infer general patterns of behavior between those students that complete the course, and those that drop out at different points in time. Logistic regression, one of the most conventional machine learning methods, is also widely utilized[23–26].

It is undoubtedly difficult for any given solution to overmatch the others in all scenarios, since different drop-out prediction tasks require different models to capture the core of problem, and the fittest models ought to deal with diverse realistic demands. In this paper, in order to fully exploit the latent patterns of data, we propose the MOOP model, which is validated to achieve promising outcomes.

## 3 Methodology

In this section, we first elaborate the tensor structure to model the MOOC data, then introduce a new similarity method to cluster course data for constructing the local tensor, and follow this with details of a pigeon-inspired parameter optimization procedure. The overall process is depicted as Algorithm 1.

### 3.1 Tensor structure for prediction

We first throw light on the basics of tensors, and then present the tensor structure equipped with a fast low-rank tensor completion for drop-out prediction.

**Tensor basics**. A tensor is a high-dimensional data representation, whose expression is vector (1-mode) and matrix (2-mode). An $n$-mode tensor can be defined as $X \in \mathbf{R}^{I_1 \times I_2 \times \cdots \times I_n}$, where $I_n$ denotes the quantity of mode $n$, and its elements are denoted as $x_{(I_1,\ldots,I_k)}$, where $1 \leqslant k \leqslant n$. The matriculating operator, which unfolds a tensor into a matrix, is defined as unfold $(X, n) = X_{(n)}$, in which the tensor element $(I_1, I_2, \ldots, I_n)$ is mapped to the matrix element $(I_n, J)$, where

$$J = \prod_{m=1,\ m \neq n}^{k-1} I_m \tag{1}$$

The reverse of the matriculation is defined as fold $(X_{(n)}, n) = X$ in a similar way.

The inner product of two same-size tensors $A, B \in \mathbf{R}^{(I_1 \times I_2 \times \cdots \times I_N)}$ is defined as the sum of the products of their entries:

$$(A, B) = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} a_{(i_1,i_2,\ldots,i_n)} b_{(i_1,i_2,\ldots,i_n)} \tag{2}$$

For any $1 \leqslant n \leqslant N$, the product of a matrix $M \in \mathbf{R}^{J \times I_n}$ with a tensor $A \in \mathbf{R}^{(I_1 \times I_2 \times \cdots \times I_N)}$ is expressed as $A \times_n M$, and transformed into the product of two matrixes.

$$Y = A \times_n M \Leftrightarrow Y_{(n)} = M A_{(n)} \tag{3}$$

Denote $\|X\|_{\mathrm{F}} = \sqrt{(X, X)}$ as the Frobenius norm of a tensor. It is clear that $\|X\|_{\mathrm{F}} = \|X_{(k)}\|$.

**Global tensor**. Suppose that there are $m$ courses and $n$ students in the MOOC data. In order to employ the excellent explanatory power of the tensor structure, we construct the data into 3-dimension tensor $\mathcal{T}$, as Fig. 2 shows.

The abscissa axis and vertical axis represent the students, which means the scale of each matrix in the tensor is $n \times n$, and the other dimension represents the $m$ course. Thus the size of $\mathcal{T}$ is $n \times n \times m$, and we randomly select a slice of size $k$. There are different value interpretations for $\mathcal{X}_{iik}$: (1) The value $\mathcal{X}_{iik} = 0$ represents "student $i$ does not select the course $k$"; (2) the value $\mathcal{X}_{iik} = 1$ represents "student $i$ does select the course $k$, and fails to finish it"; (3) the value $\mathcal{X}_{iik} = 2$ represents "student $i$ does select the course $k$, and finishes it"; (4) the value $\mathcal{X}_{ijk} = 0$ represents "student $i$ and student $j$ do not select the same course

**Algorithm 1 MOOP**

**Input:** global tensor $\overline{\mathcal{T}}$, completion tensor $\mathcal{T}$, balance parameter $\alpha$, weight parameter $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$, and $\gamma_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{ik})$, $i \in [1, k]$, selecting feature $F_1$, finishing feature $F_2$, global rating feature $F_3$, and criterion precision $P$;

**Output:** prediction $\beta, \gamma_i$.

initialize $\alpha = 0.5, \beta, \gamma_i$
local tensor $\mathcal{L}_k$ = J-Sim($\overline{\mathcal{X}}$, $F_1$, $F_2$, $F_3$)
**for** $n = 1 : k$ **do**
  train data $\mathcal{X}$ = random($\overline{\mathcal{T}}$)
  test data $\overline{\mathcal{X}} = \overline{\mathcal{T}} - \mathcal{X}$
  **while** Function($\overline{\mathcal{X}}, \mathcal{X}$) reach $P$ **do**
    $\beta_n, \gamma_n = \mathrm{PIO}(\mathcal{X}, \mathcal{L}_k, a, \beta_n, \gamma_n)$
    $\mathcal{X} = \mathrm{HaLRTC}(\mathcal{X}, \mathcal{L}_k, a, \beta_n, \gamma_n)$
    $\mathcal{L}_k = \mathrm{refresh}(\mathcal{L}_k, \mathcal{X})$
    **return** $\beta_n, \gamma_n$
  **end while**
**end for**
**return** $\beta, \gamma_i$;

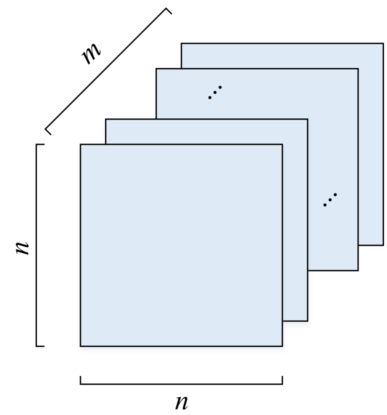



**Fig. 2 Sketch of global tensor.**

$k$"; (5) the value $\mathcal{X}_{ijk} = 1$ represents "student $i$ and student $j$ do select the same course $k$, and they both fail to finish it"; and (6) the value $\mathcal{X}_{ijk} = 2$ represents "student $i$ and student $j$ select the same course $k$, and they both finish it". As we can see, the tensor can not only express the drop-out performance of all students in all courses, but also reveal the relationship between students in the specific course. Therefore, we put all data into one tensor to construct a global tensor as a step in generating a MOOP for drop-out prediction.

### 3.2 Similarity model for local tensor

We throw light on the local tensor, and propose our new similarity calculation method, J-Sim, to cluster course dimension and construct the local tensor.

**Local tensor**. The idea of constructing a local tensor comes from the fact that there are evident differences among courses in different categories, such as humanities and social sciences, which renders the simple combination of student's behaviour across all courses inappropriate. For example, a student majoring in humanities might pay more attention in a course such as literature, and naturally shares little similarity with another student taking a course in physical sciences. These uncorrelated student pairs increase computation cost and decrease accuracy. To enhance the performance of the model, we construct a local tensor as a supplement to the global tensor so as to predict drop-out rate.

In order to model the local tensor, firstly, we introduce course vectors, which comprise three types of features:

- Selecting feature ($f_1$). For courses, this feature is assigned with the value 0 or 1 for each student, representing whether the student has chosen the course or not;
- Finishing feature ($f_2$). For courses, this feature is assigned with the value 0 or 1 for each student, representing whether the student has finished the course or not;
- Global rating feature ($f_3$). For courses, this feature records all the finished courses in the selected courses.

Secondly, we expect to cluster course dimension based on the features above to acquire several local tensors. In a local tensor, there are only correlative courses, and students' behaviours will be evaluated in one subject framework, as Fig. 3 shows. There is no doubt that with the combination of local tensors the pertinence of MOOP sharply increases.

**Similarity estimation method**. The cluster performance is largely influenced by the results of similarity estimation, which is elaborated as follows.

Suppose that there are two students $X$ and $Y$, and $X_{i,j} = (x_1, x_2)$ denotes the relation between $X$ with course $i, j$. $Y_{i,j} = (y_1, y_2)$ can be interpreted similarly as above. In the traditional method, the similarity between two vectors is usually calculated via the cosine function, which can be expressed as below:

$$\cos\theta = \frac{x_1x_2 + y_1y_2}{\sqrt{x_1^2 + y_1^2} \times \sqrt{x_2^2 + y_2^2}} \tag{4}$$

The advantage of the cosine function lies in the fact that it has a high sensitivity to common behaviours. In other words, the more mutual courses $X$ and $Y$ select in common, the more similar they might be.

Nevertheless, as can be observed from Eq. (4), we may easily draw the conclusion that when $X$ expands to $aX$, the similarity does not change.

$$\cos\theta = \frac{ax_1x_2 + ay_1y_2}{\sqrt{(ax)_1^2 + (ay)_1^2} \times \sqrt{x_2^2 + y_2^2}} \tag{5}$$

where $a$ denotes the arbitrary constant.

When it comes to MOOC, the problem can be

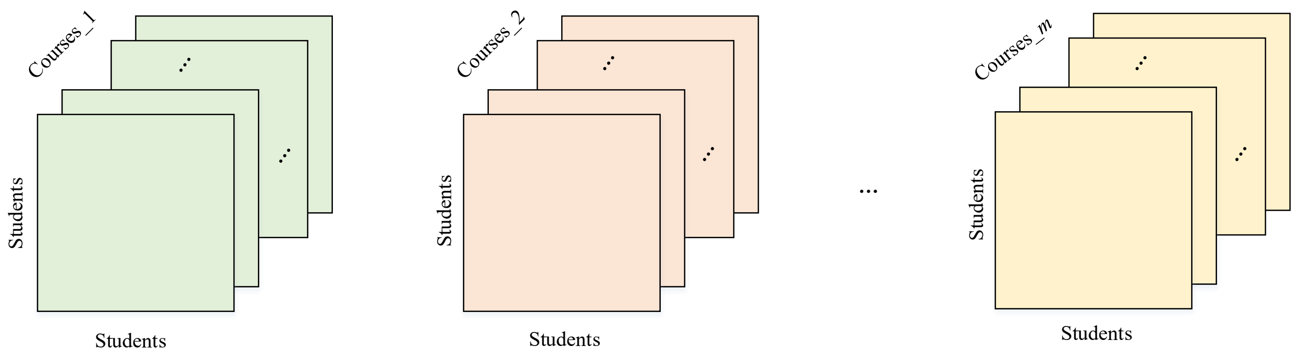



**Fig. 3 Sketch of local tensor.**

interpreted by saying that if student $X$ selects the same courses as student $Y$, where $X$ gives up on all of the courses while $Y$ finishes them, the similarity of these two students still reaches 100%. In fact, although the interests of the two students might be similar, the outcome of the learning process is disparate. Besides, the cosine function cannot distinguish the different vectors having the same angle, which might result in considering two students with totally different tastes as similar. Thus, we have to construct another unique similarity estimation method to modify the cosine function.

Inspired by the promising properties of hyperbolic functions, we propose the similarity estimation method to cluster dimensions.

$$\mathrm{sim}(X, Y) = \frac{2 \times \exp\left(-\alpha\left(\dfrac{1}{\cos\theta} - 1\right)\right)}{1 + \exp\left(-\alpha\left(\dfrac{1}{\cos\theta} - 1\right)\right)} \tag{6}$$

where $\cos\theta$ denotes cosine similarity between $X$ and $Y$, and $\alpha$ represents balance factor. The effectiveness of our similarity equation, namely J-Sim, is evaluated in Section 4.

### 3.3 Combination method

The merge of global tensor and local tensor equipped with a tensor completion method, as well as a parameter optimization algorithm applied to each combination block, are illustrated in this subsection.

**Merge of global and local tensors**. Considering just a global tensor or local tensor in isolation will distort the inner connections between students and courses. Combining a global tensor and local tensor is advantageous in three ways: (1) A global tensor introduces the overall features into the model, which can reveal how the whole course framework influences students; (2) the local tensor specifies the problem of different course groups possibly having different effects on students; (3) the merge of two tensors leads to a comprehensive analysis of the intrinsic relations between courses and students.

Suppose that we get $k$ local tensors after the former clustering. Therefore, we utilize $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$ to describe the connections between the global tensor and each local tensor, and $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_k)$ represents the corresponding parameters of each local tensor. To further explore the influence of students' behaviours from other local tensors, we expand $\gamma$ into $\gamma_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{ik})$, $i \in [1, k]$. In other words, when we concentrate on estimating the drop-out rate of courses in a certain local tensor, not the global tensor's but other local tensors' influence should be taken into account, and it is obvious the parameter $\gamma_i$ must be different in different local tensors. The vivid description is illustrated in Fig. 4.

**Tensor completion method**. In order to accurately predict and accomplish the above merge, we adopt a tensor completion method to estimate MOOC drop-outs.

Conventional tensor completion methods, such as Canonical Polyadic (CP) decomposition[27] and Tucker decomposition[28], adopt a heuristic algorithm, which lowers the dimension of existing data to simplify the calculation. Particularly in CP decomposition, a tensor $A \in \mathbf{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is represented with a suitably large $r$ as a linear combination of $r$ rank-1 tensors (vectors); that is,

$$A = \sum_{i=1}^{r} \lambda_i \alpha_i^1 \otimes \alpha_i^2 \otimes \cdots \otimes \alpha_i^d \tag{7}$$

$$\min_{X, \alpha_1, \ldots, \alpha_n} \lambda_i \parallel X - \alpha_i^1 \otimes \cdots \otimes \alpha_i^n \parallel_{\mathrm{F}}^2 \tag{8}$$

$$\text{s.t. } X_\Omega = D_\Omega \tag{9}$$

where $\Omega$ denotes the real number filed of tensor.

In Tucker decomposition, a tensor $A \in \mathbf{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is decomposed into a set of matrices $U_{(m)} \in \mathbf{R}^{I_m \times J_m}$ $(1 \leqslant m \leqslant d)$, and one small core tensor $G \in \mathbf{R}^{J_1 \times J_2 \times \cdots \times J_d}$; that is,

$$A = G \times_1 U_{(1)} \times_2 U_{(2)} \times \cdots \times_d U_{(d)} \tag{10}$$

$$\min_{X, G, U_{(1)} \ldots U_{(n)}} \frac{1}{2} \parallel X - G \times_1 U_{(1)} \times \cdots \times_n U_{(n)} \parallel_{\mathrm{F}}^2 \tag{11}$$

$$\text{s.t. } X_\Omega = D_\Omega \tag{12}$$

Nonetheless, these methods require the transformation of data structures and each decomposition partially distorts, hence giving rise to gradual error accumulation.

In contrast to conventional methods, this paper focuses on global and local tensor completion, which produces high accuracy and demands low computational efforts from the algorithm. Thus, we incorporate an HaLRTC algorithm[29] for this purpose, which has been proven to outperform other tensor completion methods.

**Parameter optimization procedure**. We mainly utilize the Pigeon-Inspired Optimization (PIO) procedure to optimize the parameters of MOOP. As introduced in Ref. [30], PIO, which is a population-based swarm intelligence algorithm, imitates pigeons

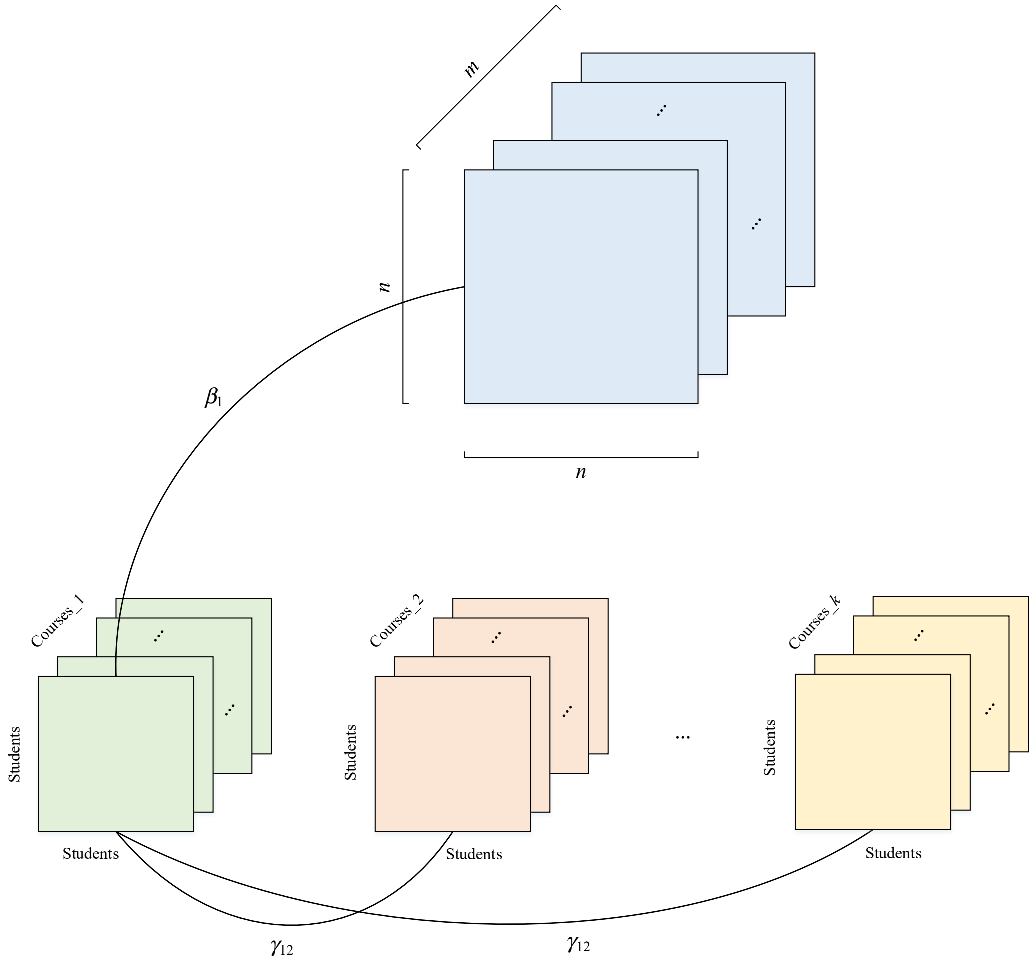



**Fig. 4 Sketch of connecting global and local tensors.**

navigation homing behavior. In the algorithm, Duan and Qiao[30] adopted two operators to describe two stages of the homing phenomenon.

In the first stage, pigeons can briefly picture the topographic map in their head by means of magnetic sense. They take the height of the sun as a compass to modify their flight path. When approaching the destination, the dependence on the sun decreases.

**Landmark operator**. In the second stage, when approaching the destination, the pigeons pay more attention to the landmark. When spotting the familiar building, they will fly straight to their goal. Otherwise, they will follow leaders that are familiar with the landmark.

PIO sets an initial location $X_i = [x_{i1}, x_{i2}, \ldots, x_{in}]$ and velocity $V_i = [v_{i1}, v_{i2}, \ldots, v_{in}]$ for the pigeons. Then, the new location and velocity of each pigeon are updated.

$$V_i^t = V_i^{t-1}\mathrm{e}^{-R\times t} + \mathrm{rand}(X_{\mathrm{gbest}} - X_i^{t-1}) \tag{13}$$

$$X_i^t = X_i^{t-1} + V_i^t \tag{14}$$

where $R \in [0,1]$ denotes the map compass operator, rand represents random number values in $[0,1]$, $t$ denotes the current iteration, and $X_{\mathrm{gbest}}$ is the global optimum in $t-1$ iterations. The first stage operation will repeat until $T_1$ iterations.

In the second stage, with the landmark operator utilized, pigeons compare the operator with the destination. If matching well, the pigeons fly straight to the goal. After each iteration, the half of the pigeons which are furthest from the destination can be weeded out. $X_{\mathrm{center}}$, which is the central location of the remainder, will be set as the new landmark. The second stage operation will repeat until $T_2$ iterations. The combined system is defined as follows:

$$X_{\mathrm{center}}^{t-1} = \frac{\sum X_i^{t-1} \cdot F(X_i^{t-1})}{N_p^{t-1} \sum F(X_i^{t-1})} \tag{15}$$

$$N_p^t = \frac{N_p^{t-1}}{2} \tag{16}$$

$$X_i = X_i^{t-1} + \mathrm{rand}(X_{\mathrm{center}}^{t-1} - X_i^{t-1}) \tag{17}$$

where $F(\cdot)$ is the quality of the individual pigeon individual, and defined as

$$F(X_i^{t-1}) = \begin{cases} \dfrac{1}{\text{fitness}(X_i^{t-1}) + \varepsilon}, & \text{for maximization;} \\ \text{fitness}(X_i^{t-1}), & \text{for minimization} \end{cases} \tag{18}$$

where $\varepsilon$ denotes the random noise.

With PIO applied to optimize the parameters of MOOP, each global tensor and local tensor pair has more explanatory and typical power for prediction.

## 4 Experiments and Results

In this section, the experimental results are reported, followed by an in-depth analysis.

### 4.1 Experiment settings

We utilized the most widely used public MOOC students-courses data (https://kddcup2015.com) for empirical evaluation.

**Data sources**. Enrolment data are leveraged for evaluation, containing information on students, courses, and the completion states of specific courses. Specifically, the number of registered students is 79 186, and the number of courses is 39. We count how many students are selecting each given course, and how many courses are registered by each student, and detailed statistics of the data are depicted in Fig. 5. As we can see, a great number of registered students (74 821) have selected less than 3 courses, which might represent the contingency of their selecting behavior and thus sharpen the explanatory ability of MOOC data. Therefore, we filter for students who select less than 3 courses in the following experiment, and Fig. 5c expresses the outcome.

**Evaluation indices**. Since our target is to predict the dropout of students, we adopt F1-score as the evaluation metric.

$$\text{MAPE} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FN} + \text{FP}} \tag{19}$$

TP denotes that the prediction is positive and the true value is positive; FP denotes that the prediction is positive and the true value is negative; and FN denotes that the prediction is negative and the true value is negative. We randomly set several existing values as null to construct training data and testing data, and the selected data was transformed into a selecting list (original value is 1) and a finishing list (original value is 2). Then, we utilized MOOP to estimate the missing data, and evaluated the prediction with the two lists.

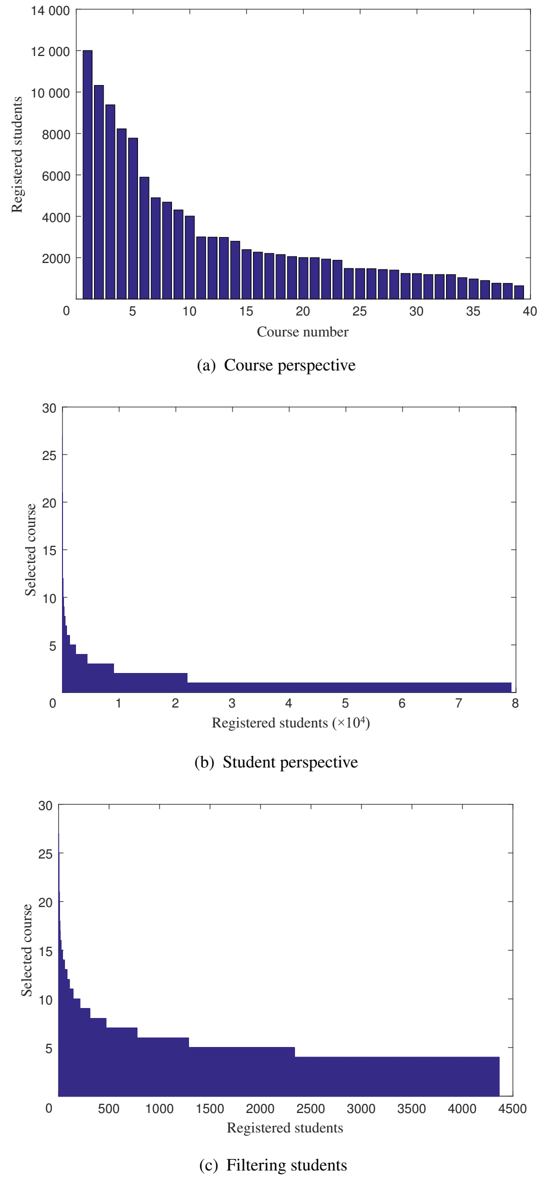



**Fig. 5 Data statistics. (a) The amount of registering students about each course; (b) the amount of courses each student registering; and (c) detailed statistics of the data.**

### 4.2 Experiment results

We firstly showed J-Sim variation trend in $\alpha = [0.1, 8]$, which verified the superiority of the similarity calculation method in Figs. 6 and 7. The outcome reveals that J-Sim has a more flexible variation trend than the cosine function and is better at distinguishing the course similarities for clustering than the conventional cosine similarity method. Note that the lighter the colour is, the less discriminative the MOOC data are, and we choose $\alpha = 0.5$.

We further compare MOOP with other state-of-the-art methods. Matrix Factorization (MF) is a traditional

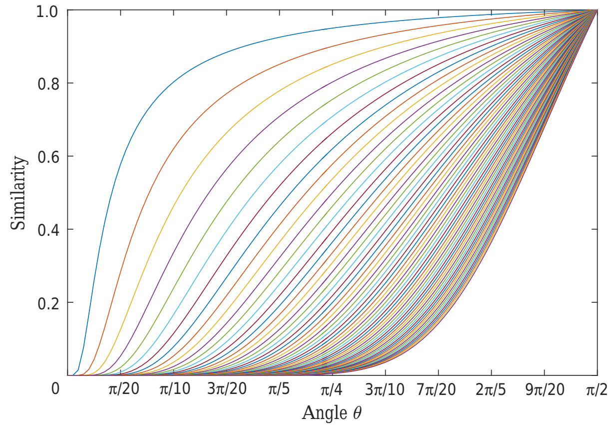
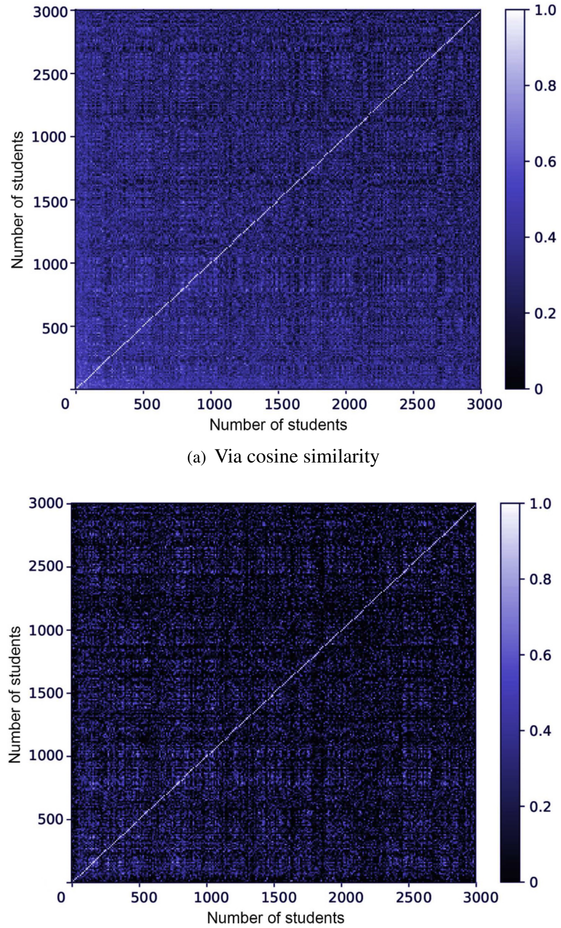



**Fig. 6 Variation trend of J-Sim.**




(a) Via cosine similarity

(b) Via J-Sim

**Fig. 7 Visualization of clutering result.**

method widely used in collaborative filtering and can solve the problem to an extent. The Tucker decomposition method, as a conventional tensor completion method, is also utilized for comparison. A Support Vector Machine (SVM), which can efficiently perform non-linear classification, is also considered as a competitor. Furthermore, in order to validate the superiority of the merged tensors, the prediction method based only on a global tensor, g-MOOP, is implemented for comparison as well. In other words, there are 5 different methods are put into assessment.

As Table 1 shows, MOOP outperforms other methods. The reason that all results in the selecting list are higher than the drop-out list is that we suppose the students who finish courses also select the same courses, and the drop-out list results are considered as an index.

Overall speaking, MF has the worst performance. Since a matrix can only simultaneously process two dimensions of data, namely students and courses, some features of the data, such as the interactions within students or courses, can be lost. Furthermore, the method cannot deeply explore the connection between two dimensions because of limitations of the algorithm. Consequently, it merely achieves 69.5% on selecting and 65.6% on drop-out. Other methods, such as Tucker and g-MOOP, only contain one more dimension, and gain an improvement of 11.8% and 15.3% on selecting, 12.5% and 16.2% on drop-out. On one hand, this reveals the superiority of the tensor structure; on the other hand, it shows that the improvement is limited since it fails to deeply explore the interactions.

The diversity between Tucker and g-MOOP might lay in the core of the tensor completion algorithm. As mentioned above, Tucker transforms the original tensor into a core tensor with a corresponding matrix in each dimension, while HaLRTC, used in g-MOOP, focuses on retaining the original tensor as far as the algorithm can reach. Thus, the latter sacrifices efficiency to gain accuracy, and gains an improvement of 3% over Tucker.

Although SVM has an advantage in features analysis, it is overmatched by MOOP by 6.1% on selecting and 10.7% on drop-out, which can be attributed to the limitations in data size.

### 4.3 Discussion

Based on the superiority of MOOP, educators can accurately orientate students who might give up on the course halfway through. The results will contribute to the following improvements.

**Table 1 Evaluation of different methods about prediction.**

| F1-score | Method | | | | |
|---|---|---|---|---|---|
| | MF | Tucker | g-MOOP | SVM | MOOP |
| Selecting | 0.695 | 0.777 | 0.801 | 0.853 | **0.905** |
| Drop-out | 0.656 | 0.738 | 0.762 | 0.794 | **0.879** |

• Educators can provide some special educational pattern for students with a high drop-out rate. Once these students login to the MOOC platform, the system would recommend those special courses to them, which reflects the idea of teaching students in accordance of their aptitude.

• When it comes to the situation where some analysis is taking place about given course, the behavior of students with a high drop-out rate might decrease the precision. Since some of these students select the course at random, especially the number selecting less than 3 courses, this could lead to errors into the evaluation of a course.

• With more information acquired, educators could analyze the reasons why these students have such a high drop-out rate. The outcome will drive reforms of the education system to some extent.

These measures will eventually enhance the ability of students, and also promote the quality of education.

## 5 Conclusion

In this paper, a MOOC prediction method based on the combination of a global tensor and local tensor, solved with HaLRTC and optimized with PIO, is proposed to tackle the drop-out prediction task. Concretely, local tensors are clustered by J-Sim, which divides all courses into several course groups. When estimating the missing drop-out data in a specific course, the global tensor and other local tensors are import optimization parameters to complete the local tensor that the selected course belongs to. Experiment results not only validate the superiority of our proposed method, but also imply the prospect of application to real-time MOOC drop-out prediction scenarios.

With regard to future work, there are several latent problems which are worthy of exploring. Firstly, the superior experimental results of MOOP can mostly be ascribed to limitations in the scale of MOOC data. When it comes to a situation where more features of data are provided, the problem of how to maintain the superiority arises. Secondly, the tensor completion method, HaLRTC, used in this work sacrifices efficiency to enhance accuracy, which might lead to high time consumption. We plan to propose a new algorithm to handle completing the tensor task, which will balance efficiency and accuracy. Thirdly, additional valuation metrics and datasets could be adopted to verify the performance of MOOP.

## Acknowledgment

This work was partially supported by the National Project of Educational Science Planning (No. ECA160409), the Hunan Provincial Project of Educational Science Planning (No. XJK016QXX001), and the National Natural Science Foundation of China (Nos. 71690233 and 71331008).

## References

[1] C. Alario-Hoyos, M. Pérez-Sanagustín, C. D. Kloos, and P. J. Muñoz-Merino, Recommendations for the design and deployment of MOOCs: Insights about the MOOC digital education of the future deployed in MiríadaX, in *Proc. $2^{nd}$ Int. Conf. Technological Ecosystems for Enhancing Multiculturality*, Salamanca, Spain, 2014, pp. 403–408.

[2] H. Khalil and M. Ebner, MOOCs completion rates and possible methods to improve retention—A literature review, in *Proc. World Conf. Educational Multimedia, Hypermedia and Telecommunications*, Chesapeake, VA, USA, 2014, pp. 1236–1244.

[3] D. F. O. Onah, J. Sinclair, and R. Boyatt, Dropout rates of massive open online courses: Behavioural patterns, in *Proc. $6^{th}$ Int. Conf. Education and New Learning Technologies*, Barcelona, Spain, 2014, pp. 5825–5834.

[4] J. Z. Qiu, J. Tang, T. X. Liu, J. Gong, C. H. Zhang, Q. Zhang, and Y. F. Xue, Modeling and predicting learning behavior in MOOCs, in *Proc. $9^{th}$ ACM Int. Conf. Web Search and Data Mining*, San Francisco, CA, USA, 2016, pp. 93–102.

[5] G. L. Zhang, T. J. Anderson, M. W. Ohland, and B. R. Thorndyke, Identifying factors influencing engineering student graduation: A longitudinal and cross-institutional study, *J. Eng. Educ.*, vol. 93, no. 4, pp. 313–320, 2004.

[6] G. Thomas, A. Henderson, and T. L. Goldfinch, The influence of university entry scores on student performance in engineering mechanics, in *Proc. Australasian Association for Engineering Education Conf.*, Adelaide, Australia, 2009, pp. 980–985.

[7] M. Sheard, Hardiness commitment, gender, and age differentiate university academic performance, *Br. J. Educ. Psychol.*, vol. 79, no. 1, pp. 189–204, 2009.

[8] B. F. French, J. C. Immekus, and W. C. Oakes, An examination of indicators of engineering students' success and persistence, *J. Eng. Educ.*, vol. 94, no. 4, pp. 419–425, 2005.

[9] A. Y. Wang and M. H. Newlin, Predictors of web-student performance: The role of self-efficacy and reasons for taking an on-line class, *Comput. Hum. Behav.*, vol. 18, no. 2, pp. 151–163, 2002.

[10] A. Parker, A study of variables that predict dropout from distance education, *Int. J. Educ. Technol.*, vol. 1, no. 2, pp. 1–10, 1999.

[11] N. Elouazizi, Point-of-view mining and cognitive presence in MOOCs: A (computational) linguistics perspective, in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing*, Doha, Qatar, 2014, pp. 32–37.

[12] S. Moon, S. Potdar, and L. Martin, Identifying student leaders from MOOC discussion forums through language influence, in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing*, Doha, Qatar, 2014, pp. 15–20.

[13] S. Crossley, D. S. McNamara, R. Baker, Y. Wang, L. Paquette, T. Barnes, and Y. Bergner, Language to completion: Success in an educational data mining massive open online class, in *Proc. Int. Conf. Educational Data Mining*, Madrid, Spain, 2015.

[14] M. M. Wen, D. Y. Yang, and C. P. Rosé, Sentiment analysis in MOOC discussion forums: What does it tell us? in *Proc. of Educational Data Mining*, London, UK, 2014.

[15] M. M. Wen, D. Y. Yang, and C. P. Rosé, Linguistic reflections of student engagement in massive open online courses, in *Proc.* $8^{th}$ *Int. AAAI Conf. Web and Social Media*, Ann Arbor, MI, USA, 2014.

[16] D. Chaplot, E. Rhim, and J. Kim, Predicting student attrition in MOOCs using sentiment analysis and neural networks, in *Proc.* $17^{th}$ *Int. Conf. Artificial Intelligence in Education*, Madrid, Spain, 2015.

[17] J. Whitehill, K. Mohan, D. Seaton, Y. Rosen, and D. Tingley, Delving deeper into MOOC student dropout prediction, arXiv preprint arXiv:1702.06404, 2017.

[18] T. Sinha, P. Jermann, N. Li, and P. Dillenbourg, Your click decides your fate: Inferring information processing and attrition behavior from MOOC video clickstream interactions, in *Proc. 2014 Conf. on Empirical Methods in Natural Language Processing*, Doha, Qatar, 2014.

[19] T. Sinha, N. Li, P. Jermann, and P. Dillenbourg, Capturing "attrition intensifying" structural traits from didactic interaction sequences of MOOC learners, in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing*, Doha, Qatar, 2014, pp. 42–49.

[20] M. Kloft, F. Stiehler, Z. L. Zheng, and N. Pinkwart, Predicting MOOC dropout over weeks using machine learning methods, in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing*, Doha, Qatar, 2014, pp. 60–65.

[21] S. Nagrecha, J. Z. Dillon, and N. V. Chawla, MOOC dropout prediction: Lessons learned from making pipelines interpretable, in *Proc.* $26^{th}$ *Int. Conf. World Wide Web Companion*, Perth, Australia, 2017, pp. 351–359.

[22] G. K. Balakrishnan and D. Coetzee, Predicting Student Retention in Massive Open Online Courses Using Hidden Markov Models, University of California, Berkeley, CA, USA, Tech. Rep. No. UCB/EEC5-2013-109, 2013.

[23] R. F. Kizilcec and S. Halawa, Attrition and achievement gaps in online learning, in *Proc.* $2^{nd}$ *ACM Conf. Learning*, Vancouver, Canada, 2015, pp. 57–66.

[24] C. Robinson, M. Yeomans, J. Reich, C. Hulleman, and H. Gehlbach, Forecasting student achievement in MOOCs with natural language processing, in *Proc.* $6^{th}$ *Int. Conf. Learning Analytics & Knowledge*, Edinburgh, UK, 2016, pp. 383–387.

[25] C. Taylor, K. Veeramachaneni, and U. M. O'Reilly, Likely to stop? Predicting stopout in massive open online courses, arXiv preprint arXiv:1408.3382, 2014.

[26] C. Ye and G. Biswas, Early prediction of student dropout and performance in MOOCSs using higher granularity temporal information, *J. Learn. Anal.*, vol. 1, no. 3, pp. 169–172, 2014.

[27] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, Scalable tensor factorizations for incomplete data, *Chemom. Intell. Lab. Syst.*, vol. 106, no. 1, pp. 41–56, 2011.

[28] T. G. Kolda and B. W. Bader, Tensor decompositions and applications, *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[29] J. Liu, P. Musialski, P. Wonka, and J. P. Ye, Tensor completion for estimating missing values in visual data, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, 2013.

[30] H. B. Duan and P. X. Qiao, Pigeon-inspired optimization: A new swarm intelligence optimizer for air robot path planning, *Int. J. Intell. Comput. Cybern.*, vol. 7, no. 1, pp. 24–37, 2014.

**Jinzhi Liao** received the BS degree from National University of Defense Technology (NUDT), China, in 2016. He is currently working toward the MS degree at NUDT. His research interests include knowledge graph and tensor completions.

**Jiuyang Tang** received the PhD degree from National University of Defense Technology (NUDT), China, in 2006. He is currently a professor with NUDT. His research interests include knowledge graph and text analytics.

**Xiang Zhao** received the PhD degree from University of New South Wales, Australia, in 2014. He is currently an assistant professor at National University of Defense Technology, China. He has published several refereed papers in international conferences and journals, such as SIGMOD, SIGIR, *VLDB Journal*, and *IEEE TKDE*. His research interests include graph data management and knowledge graph construction.