Automatic Recommendation Technology for Learning Resources with Convolutional Neural Network

Xiaoxuan Shen, Baolin Yi*, Zhaoli Zhang
National Engineering Research Center for E-Learning,
Central China Normal University, Wuhan, Hubei, China
*E-mail: epower@mail.ccnu.edu.cn,

Jiangbo Shu, and Hai Liu, *Member, IEEE*
National Engineering Research Center for E-Learning,
Central China Normal University, Wuhan, Hubei, China
hailiu0204@gmail.com

***Abstract*—Automatic learning resources recommendation has become an increasingly relevant problem: it allows students to discover new learning resources that matches their tastes, and enables e-learning system to target their learning resources to the right students. In this paper, we propose an automatic learning resources recommendation algorithm based on convolutional neural network (CNN). The CNN can be used to predict the latent factors from the text information. To train the CNN, its input and output should be solved firstly. For its input, the language model is employed. For its output, we propose the latent factor model, which is regularized by L_1-norm. Furthermore, the split Bregman iteration method is introduced to solve the model. The major novelty of the proposed recommendation algorithm is that a new CNN is constructed to make personalized recommendations. Experimental results on public database in terms of quantitative assessment show significant improvements over conventional methods. Especially, it can also work well when the existing recommendation algorithms suffer from the cold-start problem.**

Keywords—Resources recommendation; Convolutional neural network; L1 norm; split Bregman iteration method.

I. Introduction

Recommendation algorithms have become extremely popular in recent years, which has been widely used in the field of movies [1, 2], music [3], news [4, 5]and so on. The task of recommendation algorithms in e-learning systems is to give student a personalized and suitable service. For students, they can study online and achieve all kinds of learning resources. And for teachers, they can upload teaching materials and make their teaching materials online. In the existing e-learning systems, they often suffer from the information overload problem. It's difficult for students to find resources that match their tastes. To address this problem, a high-quality recommendation algorithm is needed to make a better resource service [6-11].

In the past decades, a multitude of recommendation algorithms have been developed. They can be divided into two groups: history data-based recommendation method (HDBR) and content-based recommendation method (CBR). The HDBR methods have been widely researched for recommendation system. In those methods, they only rely on the user's history data without requiring the resources detail. Collaborative Filtering (CF) [12, 13] is the one of the most distinguished approaches. In contrast, the CBR methods [14-17] create a profile to characterize each user and item. Then the system recommends items base on their profiles. But the profile of learning resources is arduous to build.

Recently, deep learning (DL) has become the most popular tool for the big data analysis [18] and artificial intelligence [19-21]. Deep learning simulates the hierarchical structure of human brain, processing data from lower level to higher level, and gradually composing increasing semantic concepts. By using deep learning algorithm, artificial intelligence has big breakthrough in many areas, such as face recognition [19], image processing [20, 22-24], and speech recognition [21]. Inspire by those, we attempted a new recommendation framework by using deep learning algorithm.

In this paper, we used the text information in learning resources (like the course introduction or the classroom content in MOOC platform, the abstract or full content of the learning resources) to predict the latent factors of the learning resources. In the following, the latent factors are utilized to predict the rating scores between students and learning resources. The proposed method can process the content from the learning resources directly without tagging by human. From the result of our experiments, it indicates that convolutional neural network is a promising model for us to bridge the semantic gap between text information and the vectors of latent factors.

The outline of the paper is organized as follows. In the next section, we introduce the proposed recommendation system based on convolutional neural network. To construct CNN, the latent factor model is proposed for its output and language model for its input. The detail processing is presented in Section 3. Some results and discussions are given in Section 4, followed by a brief conclusion section.

II. . Outline of the Proposed Recommendation Algorithm

A. Training process

In Fig. 1, a new CNN was constructed, which is shown as Level 2. To train the CNN, its input and output should be solved firstly. For the input, language model is employed. For

978-1-5090-1651-8/16 $31.00 © 2016 IEEE
DOI 10.1109/ISET.2016.12

the output, we proposed the latent factor model, which is regularized by L_1-norm.

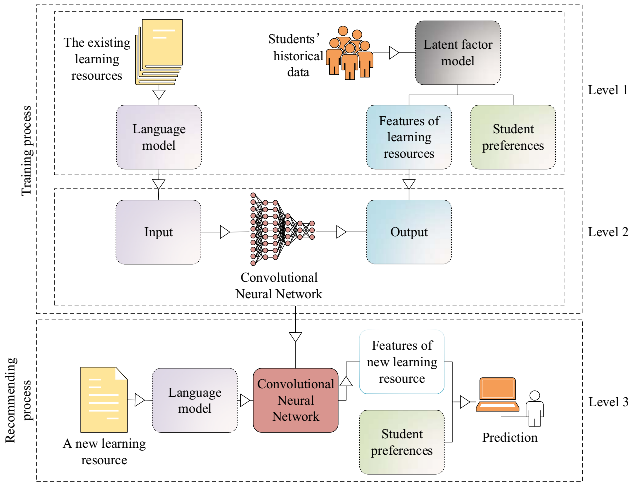

Fig. 1. Architecture of our recommendation algorithm. There are three levels in our recommendation algorithm. Level 1 denotes how the input and output of CNN is produced from the historical data. Level 2 presents that the CNN parameters is trained by its input and output. Both level 1 and level 2 are the training process. Level 3 finished a new learning resource recommendation process.

The training data for LFM is the historical rating scores between the students and the learning resources. The rating scores can be explicit that marked by students or implicit that conjectured from students' behaviors. The input data for language model is the existing learning resources' text information.

B. Recommendation process

In the recommendation process as Level 3, it obtains the text information regarding the input learning resource, maybe the content or brief introduction concerning it. Then CNN turn the input text information into the features of the learning resource, finally combining with the student's preferences the rating can be predicted.

The recommendation algorithm can provide a service to predict the rating score between a student and a learning resource. And it can work well for new learning resources. The predicted rating score indicates that whether the student needs the learning resource or not. The proposed recommendation algorithm can be utilized as a new recommendation system. And also it can be applied to enhance an existing recommendation system.

III. Proposed Method

In this section, we will explain how to build a convolutional neural network model. The CNN can be used to predict the latent factors from the text information. To construct the CNN model, its output and input need to be set firstly. The CNN output is solved by latent factor model from the historical rating scores data. The CNN input is achieved by language model according to the text information.

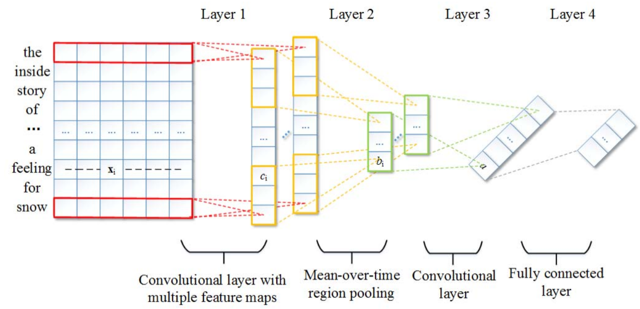

Fig. 2. Construction of convolutional neural network. Layer 1 is a convolutional layer with multiple feature maps. Layer 2 is a mean-over-time pooling layer. Layer 3 is an over-time convolutional layer. Layer 4 is a full connected layer then is the output.

A. Construction of the CNN model

The CNN model, shown in Fig. 2, is a four layers' convolutional neural network. Let $x_i \in \Re^k$ be the k-dimensional word representation vector corresponding to the i-th word in the article. An article of length n is represented as $x = [x_1, x_2, ..., x_n], x \in \Re^{nk}$. A convolution operation involves a filter $w \in \Re^k$, which is applied to a word representation vector to produce a new feature, shown in Layer1.

$$c_i = f(w \cdot x_i + b) \tag{1}$$

here $b \in \Re$ is a bias term and f is a non-linear function such as the sigmoid function. This filter is applied to each word in the sentence to produce a feature map $c = [c_1, c_2, ..., c_n], c \in \Re^n$. We then apply a mean-overtime region pooling operation over the feature map, Layer 2 achieve the pooling operation in λ regions

$$b_i = \max\{c_{(i-1)\times(n/\lambda)+1}, ..., c_{i\times(n/\lambda)}\}, i \in [1, \lambda] \tag{2}$$

Layer 3 is a convolutional layer. A convolution operation involves a filter $w \in \Re^\lambda$, which is applied on $b = [b_1, b_2, ..., b_\lambda]$ to produce a feature value

$$a = f(\mathbf{w} \cdot \mathbf{b} + b) \tag{3}$$

We have described the process that extracts one feature from one filter. The model uses multiple filters to obtain multiple features. These features are passed to a fully connected layer whose output is the predicted latent factors, shown in Layer 4.

Latent factor vectors are real-valued, so the most straightforward objective is to minimize the mean squared error (MSE) of the predictions. Let y_i' be the latent factor vector for article i and y_i is the output of the CNN model. The objective functions are then (w, b represent all the parameters in the CNN model):

$$\arg\min_{w,b} \sum_i \|y_i' - y_i\|^2 \tag{4}$$

B. CNN output solved by latent factor model

To achieve the information of students' preferences and learning resources features, we tried the latent factor model (LFM) to get those. Foremost the historical rating matrix is a

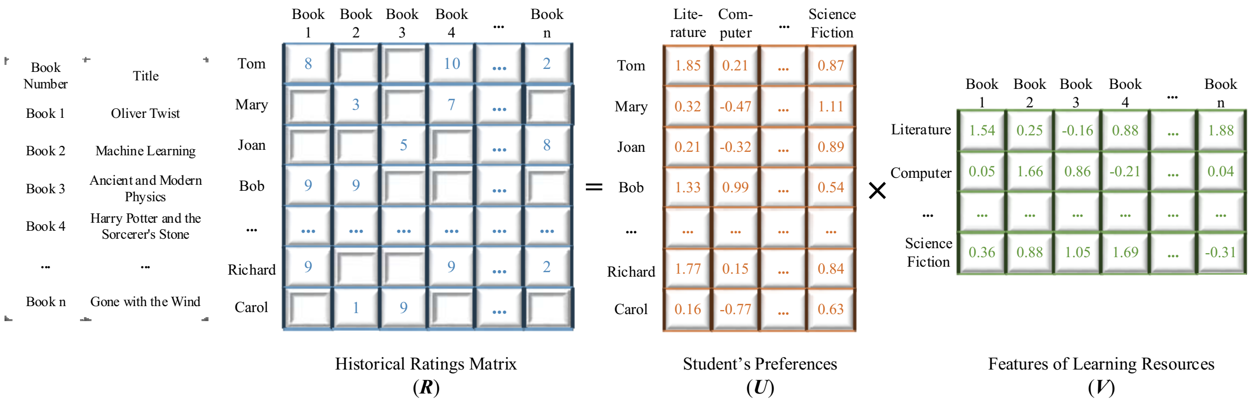

Fig. 3. Process of the Latent factor model (LFM). Matrix R denotes the historical rating scores matrix made by all the users, the vacancy means the user did not give corresponding rating. Matrix U denotes the relationship between the students and the latent factors. Matrix V denotes the relationship between the learning resources and the latent factors.

sparse matrix, so we only calculate the nonempty parts in matrix R which is showed in Fig. 3 in LFM. Second, LFM is an alternative approach that tries to explain the ratings by characterizing both items and users on a set number of factors inferred from the ratings patterns. Just like the matrixes $\boldsymbol{U}$ and $\boldsymbol{V}$ shows in Fig. 3. Third, the latent factors can indicate the index of students' taste or need which is showed in Fig. 3. To constrain the solution space, the L_2-norm regularization is often used in the traditional LFM. However, it usually results in the over-smoothing problem. In our algorithm, the LFM results represent the features of students and learning resources. Thus, it is more reasonable for us to use the sparse prior to regularize the result.

Accord to the analysis above, we proposed a modified matrix factorization method, try to use the sparse prior (L_1-norm based regularization) to normalize the solution, model is presented below

$$J(U,V)=\sum_{ij}\left(U_{i*}\cdot V_{*j}-r_{ij}\right)^2+\lambda_1\|U\|_1+\lambda_2\|V\|_1 \tag{5}$$

where the first term is the fidelity term, and rest terms are the regularization terms. The matrix $\boldsymbol{U}$ denotes the relationship between the students and the latent factors, matrix $\boldsymbol{V}$ means present the relationship between the learning resources and the latent factors. The symbol $\boldsymbol{r}_{ij}$ means the rating score that made by i-th student to the j-th learning resource (only calculate the nonempty part in the historical rating data). The λ_1 and λ_2 are the regularization parameters, which can balance the fidelity term and regularization terms. We call this method as latent factor model with L_1-norm regularization (LFM-L_1R). The L_1-norm based model in Eq. (2) is a non-convex function. To minimize it, the split Bregman iteration method [25] is introduced.

C. CNN input computing by language model

Topic Model. A topic model captures this intuition in a mathematical framework, which allows examining a set of documents and discovering, based on the statistics of the words in each, what the topics might be and what each document's balance of topics is. Then we can represent a word as the probability that belongs to the topics. In this paper, the Latent Dirichlet Allocation (LDA) method [26] is used to train the topic model.

IV. Experiments and Discussion

A. Setup Experiments

To verify the effective of the proposed recommendation algorithm, we have done some experiments on the BookCrossing dataset [27]. The dataset is collected by Cai-Nicolas Ziegler from the Book-Crossing virtual book community in 2004. According to the BookCrossing dataset, we supply the brief introductions of the books from Amazon.

TABLE I. Introduction of Dataset.

Items	Values
Number of Books	10591
Number of Users	26387
Numbers of Rating score	407573
Numbers of book introductions	10393
Average words of the book introduction	133.7
Minimum words of the book introduction	50
Maximum words of the book introduction	2791

After cleaning the null values and the false codes in the dataset, we got 10393 samples. Then 8500 samples was chosen randomly to train the model, then the rest 1893 samples to test the model.

In the dataset, the rating scores marked by users are between 0 and 10. The higher the score is, the more favorite it represents. According to the analysis for the database, we find that most score values are 0 or between 7 and 10. In order to simplify this problem, we defined that this user loves the book if his/her rating score is higher than 6. And otherwise means that he/she hates the book.

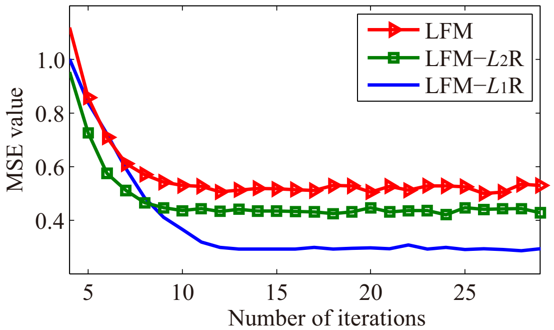

Fig. 4. MSE versus the iteration number of the three methods (LFM, LFM-L_2R, LFM-L_1R) for the BookCrossing dataset [27]. All the comparison methods here are with the same language model, namely, Latent Dirichlet Allocation (LDA) model [26].

In the first place, three latent factor models are trained with 40 latent factors (latent factor model, latent factor model with L_2-norm Regularization and latent factor model with L_1-based Regularization). Furthermore, we trained the topic model and distributed representation of words with 100 topics by LDA. Most words in our dataset are specialized, and highly related to a specific book, thus, it's inappropriate for us to use pre-trained word vectors. Finally, we built the above-mentioned CNN model for experiments. The Mini-batch-SGD (Mini-batch Stochastic Gradient Descent) was employed to optimize the CNN model. And the contrast experiment was set with linear regression (LR) model and multilayer perceptron (MLP) with one-hot bag-of-words representation language model as input.

B. *Results*

In this paper, two indexes are used to assess the model. First is the mean square error (MSE). It can evaluate the fitting degree of the model. Second is precision, it means the precision concerning the ratings we predicted in the test samples. The prediction we made is whether users would like the book or not by the brief introductions of the books. There are 71732 ratings concerning 1893 samples need to be predicted. Upper bound means the best result we could get, calculated directly by latent factors.

TABLE II. RESULTS OF THE EXPERIMENTS IN DIFFERENT MODELS

Algorithm	MSE values	Precision	Upper bound
Random	8.3610	0.51299	0.9972
LR	1.5679	0.66153	0.9953
MLP	0.54638	0.69194	0.9972
CNN (LDA+LFM)	0.51353	0.75567	0.9997
CNN (LDA+LFM-L_2R)	0.42513	0.76007	0.9972
CNN (LDA+LFM-L_1R)	**0.28981**	**0.76158**	0.9987

In the results in TABLE II, it shows that CNN model had better performance than other models. So result indicates that CNN is a superb model for us to bridge the huge semantic gap between text information and the vectors of latent factors.

In these three CNN models, CNN with sparse representation latent factors (LFM-L_1R) have better performance than others. And we could come to the same conclusion in the training process (Fig.4). Therefore, the sparse representation latent factors (LFM-L_1R) is better than the smooth representation (LFM-L_2R).

In TABLE III, we demonstrate recommendation results (most similar books) according to user' query. The second column is predicted by the latent factor model, therefore it based on users' historical data. And the third column is predicted by our CNN model, so it based on the brief introduction of the book. For example, to the query "Harry Potter and the Chamber of Secrets" the most similar books predicted by latent factors are other books from Harry Potter series. It means the folks who read "Harry Potter and the Chamber of Secrets" have a fair chance to read the rest books in Harry Potter series and gave similar ratings. The most similar books predicted by CNN are all the fictions with teenagers and fantasy. And the most similar books with "Dark Apprentice (Star Wars: The Jedi Academy Trilogy)" predicted by CNN are all science fictions about the space and there are 2 books in Star Wars series.

TABLE III. RESULTS REGARDING THE PREDICTION OF THE SIMILAR BOOKS.

Query	Most similar books (latent factors)	Most similar books (predicted by CNN)
Harry Potter and the Chamber of Secrets	1. Harry Potter and the Sorcerer's Stone 2. Harry Potter and the Prisoner of Azkaban 3. The Blue Sword 4. Harry Potter and the Goblet of Fire	1. Diary of a Mad Bride (Summer Display Opportunity) 2. The Sixteen Pleasures: A Novel 3. Mister Posterior and the Genius Child 4. Men and the Girls
Dark Apprentice (Star Wars: The Jedi Academy Trilogy)	1. The Courtship of Princess Leia (Star Wars) 2. Blood Orchid 3. Against All Enemies: Inside America's War on Terror 4. Jane Eyre (Signet Classics)	1. Temple 2. Invasion: The Soldiers of Fear (Star Trek: The Next Generation) 3. Wraith Squadron (Star Wars: X-Wing Series) 4. Into Thin Air : A Personal Account of the Mt. Everest Disaster

V. CONCLUSION

In this paper, we have proposed a content-based learning resources recommendation algorithm based on convolutional neural network. The CNN can be used to predict the latent factors from the text information. To train the CNN, its input and output should be solved firstly. For the input, language model is employed. For the output, we proposed the latent factor model, which is regularized by L_1-norm. Furthermore, the split Bregman iteration method is introduced to solve the model. The major novelty of the proposed recommendation algorithm is that a new CNN is constructed to make personalized recommendations and achieved a superior result. The proposed algorithm is verified on a public dataset. Results indicate that our recommendation algorithm for recommending new and unpopular learning resources is feasible. And we believe that the convolutional neural network-based model will play a crucial role in the e-learning systems or intelligent tutoring systems in the future. Although the application considered here is learning resources recommendation, the method is more generally applicable to news recommendation, and so on.

ACKNOWLEDGMENT

The authors thank the editor and anonymous reviewers for their valuable suggestions. This research was partially funded by the National Natural Science Foundation of China under Grant (No. 61505064), the National Social Science Fund of China (14BGL131), the Project of the Program for National Key Technology Research and Development Program (2013BAH18F01, 2014BAH22F01, 2015BAK07B03), the Self-Determined Research Funds of CCNU from the Colleges' Basic Research and Operation of MOE (CCNU15A05009, CCNU15A05059), and the Project of the Program for National Key Technology Research and Development Program (2013BAH72B01, 2013BAH18F02, 2015BAH33F02).

REFERENCES

[1] R. M. Bell and Y. Koren, "Scalable collaborative filtering with jointly derived neighborhood interpolation weights," in *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, 2007, pp. 43-52.

[2] M. N. Moreno, S. Segrera, V. F. López, M. D. Muñoz, and Á. L. Sánchez, "Web mining based framework for solving usual problems in recommender systems. A case study for movies′ recommendation," *Neurocomputing,* 2015.

[3] B. McFee, L. Barrington, and G. Lanckriet, "Learning content similarity for music recommendation," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 20, pp. 2207-2218, 2012.

[4] G. De Francisci Morales, A. Gionis, and C. Lucchese, "From chatter to headlines: harnessing the real-time web for personalized news recommendation," in *Proceedings of the fifth ACM international conference on Web search and data mining*, 2012, pp. 153-162.

[5] H. Liu, L. Yan, Y. Chang, H. Fang, and T. Zhang, "Spectral deconvolution and feature extraction with robust adaptive Tikhonov regularization," *IEEE Transactions on Instrumentation and Measurement,* vol. 62, pp. 315-327, 2013.

[6] H. Liu, S. Liu, Z. Zhang, J. Sun, and J. Shu, "Adaptive total variation-based spectral deconvolution with the split Bregman method," *Applied Optics,* vol. 53, pp. 8240-8248, 2014.

[7] K. Verbert, X. Ochoa, M. Derntl, M. Wolpers, A. Pardo, and E. Duval, "Semi-automatic assembly of learning resources," *Computers & Education,* vol. 59, pp. 1257–1272, 2012.

[8] C. Romero, S. Ventura, A. Zafra, and P. D. Bra, "Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems," *Computers & Education,* vol. 53, pp. 828–840, 2009.

[9] H. Liu, Z. Zhang, S. Liu, T. Liu, L. Yan, and T. Zhang, "Richardson-Lucy blind deconvolution of spectroscopic data with wavelet regularization," *Applied Optics,* vol. 54, pp. 1770-1775, 2015.

[10] A. Klašnja-Milićević, B. Vesin, M. Ivanović, and Z. Budimac, "E-Learning personalization based on hybrid recommendation strategy and learning style identification," *Computers & Education,* vol. 56, pp. 885-899, 2011.

[11] K. Verbert, N. Manouselis, X. Ochoa, M. Wolpers, H. Drachsler, I. Bosnic*, et al.*, "Context-aware recommender systems for learning: a survey and future challenges," *Learning Technologies, IEEE Transactions on,* vol. 5, pp. 318-335, 2012.

[12] Y. Koren and R. Bell, "Advances in Collaborative Filtering," in *Recommender Systems Handbook*, ed: Springer, 2015, pp. 77-118.

[13] A. Salah, N. Rogovschi, and M. Nadif, "A dynamic collaborative filtering system via a weighted clustering approach," *Neurocomputing,* vol. 175, pp. 206-215, 2016.

[14] M. Balabanović and Y. Shoham, "Fab: content-based, collaborative recommendation," *Communications of the ACM,* vol. 40, pp. 66-72, 1997.

[15] P. Lops, M. De Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender systems handbook*, ed: Springer, 2011, pp. 73-105.

[16] Y. Blanco-Fernandez, J. J. Pazos-Arias, A. Gil-Solla, M. Ramos-Cabrer, and M. Lopez-Nores, "Providing entertainment by content-based filtering and semantic reasoning in intelligent recommender systems," *Consumer Electronics, IEEE Transactions on,* vol. 54, pp. 727-735, 2008.

[17] R. Krestel and P. Fankhauser, "Personalized topic-based tag recommendation," *Neurocomputing,* vol. 76, pp. 61-70, 2012.

[18] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil, "Learning semantic representations using convolutional neural networks for web search," in *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, 2014, pp. 373-374.

[19] C. Ding and D. Tao, "Robust Face Recognition via Multimodal Deep Face Representation," *Multimedia, IEEE Transactions on,* vol. 17, pp. 2049-2058, 2015.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[21] A. Graves, A. R. Mohamed, and G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks," in *Acoustics, Speech and Signal Processing, IEEE International Conference on*, 2013, pp. 6645 - 6649.

[22] H. Liu, Z. Zhang, S. Liu, L. Yan, T. Liu, and T. Zhang, "Joint Baseline-Correction and Denoising for Raman Spectra," *Applied Spectroscopy,* vol. 69, pp. 1013-1022, 2015.

[23] H. Liu, S. Liu, T. Huang, Z. Zhang, Y. Hu, and T. Zhang, "Infrared spectrum blind deconvolution algorithm via learned dictionaries and sparse representation," *Applied Optics,* vol. 55, pp. 2813-2818, 2016.

[24] H. Liu, Z. Zhang, S. Liu, J. Shu, and Z. Liu, "Blind infrared spectroscopic data restoration with the similarity of multi-scales regularization," presented at the IEEE Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Hong Kong, December 17-20, 2015.

[25] T. Goldstein and S. Osher, "The Split Bregman Method for L1-Regularized Problems," *SIAM Journal on Imaging Sciences,* vol. 2, pp. 323-343, 2009.

[26] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research,* vol. 3, pp. 993-1022, 2003.

[27] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification," in *Proceedings of the 14th international conference on World Wide Web*, 2005, pp. 22-32.