Advances and Applications in Mathematical Sciences
Volume 21, Issue 6, April 2022, Pages 3293-3302

SOCIAL MEDIA POPULARITY PREDICTION BY TTM CLUSTERED BAYESIAN FACTORIZATION

NAVDEEP BOHRA and VISHAL BHATNAGAR

Research Scholar, USICT, GGSIPU
Dwarka -110078, India
and

Maharaja Surajmal Institute of Technology
Janakpuri New Delhi -110058, India

NSUT, East Campus (Formerly
Ambedkar Institue of Advance Communication
Technology and Research)
New Delhi -110031, India
E-mail: navdeepbohra@gmail.com

Abstract

Social media has converted into a marketplace due to the presence of almost everyone on it. By classifying people based on their prior interactions, it is possible to forecast which postings will attract the most attention from customers. Using tensor-based Bayesian factorization, this research proposes a unique method for removing outliers and forecasting popularity. In order to cluster user popularity, the tensor trace maximisation approach has been presented. This two-level technique is also investigated for several additional topological features. An increase in forecast popularity accuracy of 18.65% is seen using the presented novel popularity prediction scheme.

1. Introduction

The user's visibility on social media has grown recently. The social media sharing of ideas and products is one of the key drivers of this expansion. Social media has changed marketing and is now the dominant means for

2020 Mathematics Subject Classification: 68T09.
Keywords: Tensor factorization, Popularity prediction, Bayesian modeling, Clustering.
Received September 28, 2021; Accepted January 15, 2022

brands to attract customers. The knowledge was circulated via Emails in the early days of the internet, but its unstructured dispersal made analysis difficult.

Feature-based and generative-based approaches are the two most used forecast popularity methods. The features-based method forecasts future popularity by building a classification model with user, content, temporal, and structural factors. Various feature-based algorithms have efficiently predicted social media content popularity [1-5]. To grasp the popularity associated information, generative techniques describe and model less information [10]. But generative techniques have lower predictive power. Using the Deep Hawakes [10] technique, which combines deep learning with generative approach processes, can alleviate this issue. It aided in interpreting information cascades and was highly predictive.

The online social network allows for tracking individual perspectives on product distribution. Twitter Managers, for example, identify the tweets most likely to become viral. Online content forecast popularity is divided into two groups: (i) Popularity among users and (ii) population. The popularity system may be lacking certain user data that is sensitive to their emotions. The popularity prediction system lacks adaptability due to a lack of user behavior and interest information [6, 13]. A group-level popularity mechanism was presented [6]. For example, on Twitter or Facebook, some people create groups based on their interests and localities. The user's reactions to the material are quite constant. The group-level popularity improved prediction quality and scheme learning cost. It is more compact than the user-level popularity forecast [6].

A time series forecasting approach was presented to predict social media content popularity [7, 14]. It uses previous data to anticipate the future. Short time forecasting is simple to use and successful in predicting popularity. [8] uses trend prediction to anticipate the popularity of crime-related messages on Twitter. The social networking platform has a lot of online news information. The IDSS (Intelligent Decision Support System) analyses the Mashable online news dataset [9]. The IDSS algorithm predicts the popularity of online news material based on the previously published information. The IDSS technique predicts popularity more accurately than the present approach.

The number of times visitors visit an article's URL and share it on social media determines its popularity. [11] Proposes a feature-based technique to predicting online news item popularity. There is a source of news for each article. A multidimensional feature generation-based technique accurately predicts news item popularity [11].

A. Research Gap and Contribution

The sentiments derived popularity or rating of any social media post should be accurately predicted and such noises should also be identified as outliers and filtered from the prediction algorithm. Following these conventions, this paper has contributed as following to predict the popularity of the social media content: 3

- The data is presented as a tensor and clustered into similar groups based on user's interest with the help of the tensor trace maximization (TTM) scheme.
- The outliers are removed by normalizing the clustered data and calculating the distance of each information point to another considered as weights.
- Bayesian tensor factorization is used for the prediction of the post.

Following the motivation, the paper is categorized further in the proposed work discussion (section II) and the result's discussion (section III). The work is concluded in section (IV), followed by the references.

2. Proposed Work Flow and Popularity Prediction

This work forecast the popularity of social media, which helps to make the right advertisement targeting decisions for companies. Each user's connection to another may be shown as a node in a multilayer graph $G = (V, E)$ where each node represents a user as $V \in V_{i=1, 2, \ldots, n}$ and each user is connected to another through an edge E where $E \subseteq V \times V$. User history and network connections are depicted in figure 1 by this graph.

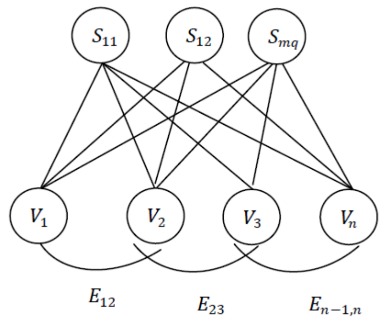

Figure 1. User connection with another user as vertex of the graph and connected via edge weights.

Although users may alter their interests within the time span $t_1 - t_2$, we have set the goal of a grouping of all similar-interested users in the user clustering. Within their network, users are exposed to a variety of content. As a result, the clustering must also consider the group of users as well. The edge weights shared by two users v_j and v'_j is

$$W^*_{v_j v'_j} = \max(A'_{jj}, A_{j'j}) \tag{1}$$

Where A is the adjacency matrix which represents the connection strength between two users based on their interests. When grouping users into distinct categories, the dispersion of users must be considered in comparison to one another. A set of clusters can be $C = \{C_1, C_2, \ldots, C_l\}$. The tensor trace maximization (TTM) is suggested to cluster the content into distinctive groups. The TTM minimizes the weights as

$$\arg\min \sum W^*_{u \cdot v} \mid u \in C^*_i, , v \in C^*_j \tag{2}$$

In the first phase of TTM, graph reduction is achieved by breaking up a graph into k subgraphs by constructing two sets A and B, $(K = 2$ is represented).

$$Ncut(A, B) = cut(A, B) \cdot \left(\frac{1}{assoc(A, V)} + \frac{1}{assoc(B, V)}\right) \tag{3}$$

Where, $cut(A, B) = \sum_{V \in A_i, B_j} W_{ij}$; describes the weighted links between different bunches A and B. The users affiliated with only group A is computed as $assoc(A, V) = \sum_{V \in A_i, V_j} W_{ij}$ and similarly for group B $assoc(B, V) = \sum_{V \in B_i, V_j} W_{ij}$. Minimizing $Ncut(A, B)$ is to achieve the optimal divisions y is given as

$$y = \arg\min_{y \subset R^n, y^T} \frac{(y^T (Dig - W)y)}{(y^T Digy)} \tag{4}$$

Here, $Dig = R^{n \times n}$ is the diagonal elements of the diagonal matrix $Dig_{ii} = \sum_i W_{ij}$. The vertices can be partitioned following every cell entry in y. From Equation 4, the non-diagonal members set to zero, we may create a Laplacian matrix (L). The Fiedler vector serves as the foundation for unnormalized Laplacian matrix articulation in this case. The normalized cuts procedure is made possible by the diagonal matrix's architecture.

In order to proceed, we must first obtain the clusters. A K-means method is applied to the rows of Y, and this includes the Laplacian matrix's top eigenvectors. There is a lack of uniformity in the Laplacian matrix's eigenvectors, which might lead to poor clustering results if they are not well selected. The first k prominent eigenvectors are used in the particular case. For $i = 1 : n$, let $y_i \in R^k$ be a row of Y as a column vector. Hence,

$$Y = \begin{bmatrix} - \; y_1^T \; - \\ \vdots \\ - \; y_n^T \; - \end{bmatrix} \in R^{n \times k} \tag{6}$$

The cluster is defined as k. A cluster indicator matrix $X \in R^{n \times k}$ can be used to determine the divisions of the row of Y. Each column $j = 1, \ldots, k$ of X represents a cluster. Each row $i = 1, \ldots, n$ is the membership of y_i. So, $X_{ij} = 1/\sqrt{f_j}$ if and only if the data point in y_i is the j^{th} cluster $(f_j = \| X^j \|_0 ; X^{(j)})$ is the j^{th} column of X and X_j^0 are the non-zero elements of $X^{(j)}$. The repetitive nature of k-means is evident. Using this strategy, the clustering results are created using all of the information included in each of the many eigenvectors. There are various examples in which the number of clusters is examined and cited in our earlier article [12].

To remove the outliers, top k similar contents in each cluster are extracted and normalized to ease the computation. Top k similar contents are selected based on Euclidean distance. The higher the similarity, the lesser is the distance. If X denotes the clustered tensor and x denotes the contents in each cluster, then normalized tensor in each cluster can be calculated as:

$$\underline{X_{itj}} = \frac{x_{itj}}{\sum x_{itj}} \tag{7}$$

Outliers have been removed, which depicts the normalized content for population popularity at timestamps $[1, t_1]$ and $[1, q]$. If the event at t_q and all other historical contents are separated by an average distance of t_q, then outliers will not appear in the results.

B. Bayesian probabilistic Tensor Factorization for Popularity Prediction.

BPTF tensor factorization scheme introduces a new set of latent features at different time frames. BPTF scheme has been proved very efficient for the sparse data in the past. The seedline of BPTF is probabilistic matrix factorization (PTF) which is extended to tensor factorization for prediction. It can be stated that each tensor can be factorized as cross product of three matrices as:

$$\mathcal{R} = \sum_{d=1}^{D} U_d^{\circ} V_d^{\circ} T_d \tag{8}$$

Where U_d, V_d, T_d are d^{th} rows vector of matrices U, V and T. U represents the users, V represents the Behance projects and T is the time for which ratings are recorded. This is also a representation of CANDECOMP/PARAFAC decomposition. The objective function for the factorization can be represented as in equation 9. This objective function is non-convex and to improve its local minima exploring capability, Adam optimization is used in place of stochastic gradient search in BPTF. 6

$$\sum_{K=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}^{k}(R_{ij}^{k} - \langle U_i, V_j, T_k\rangle)^2 + \sum_{i=1}^{N}\frac{\lambda_U\| U_i \|^2}{2} + \sum_{j=1}^{M}\frac{\lambda_V\| V_j \|^2}{2}$$

$$+ \sum_{k=1}^{K}\frac{\lambda_{dT}\| T_k - T_{k-1} \|^2}{2} + \frac{\lambda_0\| T - \mu_T \|^2}{2} \tag{9}$$

Fully Bayesian treatment is used to tune equation 9. The hyper parameters $\Theta_U = \{\mu_U, \lambda_U\}, \Theta_V = \{\mu_V, \lambda_V\}, \Theta_T = \{\mu_T, \lambda_T\}$ are tuned by the maximum prior probability of the Bayesian approach. The predictive distribution by optimizing and tuning the hyper parameters $\Theta_U, \Theta_V, \Theta_T$ are calculated as:

$$p(\hat{R}_{ij}^{k} \mid R) = \int p(\hat{R}_{ij}^{k} \mid U_i, V_j, T_k, \alpha)p(U, V, T, \alpha, \Theta_U, \Theta_V, \Theta_T \mid R) \tag{10}$$

The hyper parameter's prior distribution is calculated by Wishart's distribution (w) as:

$$p(\alpha) = w(\alpha \mid w_0, \widetilde{v_0})$$

$$p(\Theta_U) = p(\mu_U \mid \lambda_U)p(\lambda_U) = \mathcal{N}(\mu_0, (\beta_0\lambda_U)^{-1})w(\lambda_U \mid w_0, v_0)$$

$$p(\Theta_V) = p(\mu_V \mid \lambda_V)p(\lambda_V) = \mathcal{N}(\mu_0, (\beta_0\lambda_V)^{-1})w(\lambda_V \mid w_0, v_0)$$

$$p(\Theta_T) = p(\mu_T \mid \lambda_T)p(\lambda_T) = \mathcal{N}(\mu_0, (\beta_0\lambda_T)^{-1})w(\lambda_T \mid w_0, v_0)$$

3. Results and Discussion

Data from the Behance network is used to evaluate the suggested algorithm for predicting social media content before it is posted [6]. 85092 people, 1326 projects, and 60-hour timestamps were the sources of this dataset, which was compiled in June 2014. We have tried the suggested approach using TTM and Bayesian tensor factorization for the clustering of tensor and forecasting popularity. Most of the results are contrasted with those of M. X. Hoang's earlier study [6]. The predicted popularity is represented as the area graph in figure 2. Each color in the figure represents a group and training/testing data is distinguished by a red line, plotted at

$t = 30$ days. The left area of this red line is used for training and the right side area is plotted for the training data. The suggested technique has an upside over the prior work that it can also forecast future changes in the testing data. TTM and BPTF successfully predicted the dramatic rise in popularity of the green-colored group in figure 2(a) after 30 days, along with other groups, as seen in the ground truth plot. In contrast, figure 2(b) shows that the current state of the art cannot accurately anticipate future changes for 5th to 10th clusters. For $t_1 = 30$ out of 60-time stamps, these are tested. The proposed strategy improves popularity prediction error performance by 18.65 percent.

The proposed prediction approach is also tested without normalized data and compared with the reference method too for the same unnormalized data. The relative error comparison is recorded in table 1. 7

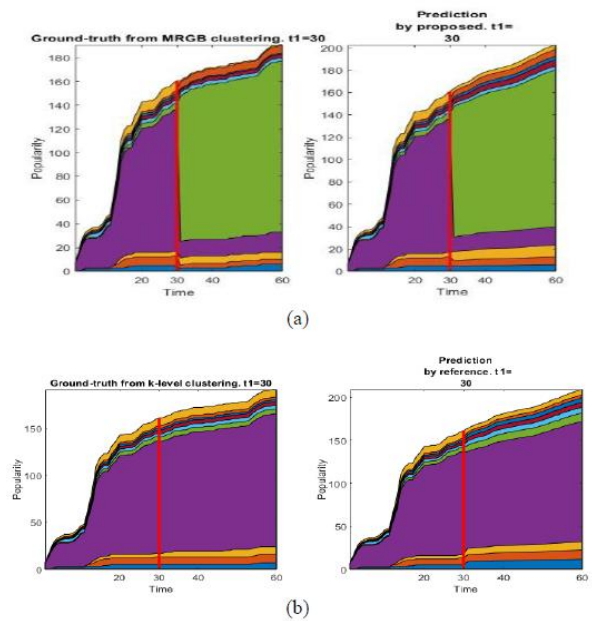

Figure 2. TTM+BPTF popularity prediction for the Behance data (b) state-of-the-art [6].

Table 1. comparison in terms of relative error with the state-of-the-art work.

	Variants	Relative Error
A	Proposed TTL+BPTF	44.796
B	M.X. Hoang [6]	48.606
C	Without norm with proposed	46. 6755
D	Without norm with reference	57.3584

For a forecasting time window of $(q - t_1)$, we tested the relative inaccuracy in prediction. The larger the training data is, the better the prediction accuracy will be since outliers are less likely to occur with smaller intervals. Since there is less training data available for longer time periods, it exhibits an increase in inaccuracy. On the other hand, the suggested strategy has a lower relative error for each prediction interval.

4. Conclusion

The popularity of a social media user group may be predicted using TTM tensor data clustering and Bayesian factorization in this paper. The Behance data is used for the experiment. In the noise filtering which is generated due to a change in user's interest due to uncontrolled circumstances, the TTM clustering with top k similar group contents is used. At the forecasting stage, the Bayesian probabilistic model is used to factorize the tensor data into three factors and trained using a probabilistic model to achieve an accuracy upto 18.65% than the reference work in [6].

References

[1] S. De, A. Maity, V. Goel, S. Shitole and A. Bhattacharya, Predicting the popularity of instagram posts for a lifestyle magazine using deep learning, 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA), Mumbai (2017), 174-177.

[2] Prateek Agrawal, Deepak Chaudhary, Vishu Madaan, Anatoliy Zabrovskiy, Radu Prodan, Dragi Kimovski, Christian Timmerer, Automated Bank Cheque Verification Using Image Processing and Deep Learning Methods, Multimedia tools and applications (MTAP) 80(1) 5319-5350. https://doi.org/10.1007/s11042-020-09818-1

[3] Marouane Birjali, Abderrahim Beni-Hssane, Mohammed Erritali, Analyzing social media through big data using info sphere biginsights and apache flume, Procedia Computer Science 113 (2017), 280-285.

[4] Neeti Sangwan and Vishal Bhatnagar, Video Popularity Prediction Using Stacked Bilstm Layers, Malaysian Journal of Computer Science 34(3) (2021), 42-254.

[5] Kota Yamaguchi, Tamara L. Berg and Luis E. Ortiz, Chic or Social: Visual Popularity Analysis in Online Fashion Networks, ACM Multimedia (2014), 773-776.

[6] Minh X. Hoang, Xuan-Hong Dang, Xiang Wu, Zhenyu Yan and Ambuj K. Singh, GPOP: Scalable Group-level Popularity Prediction for Online Content in Social Networks, Proceedings of the 26th International Conference on World Wide Web (2017), 725-733.

[7] Ying Hu, Changjun Hu, Shushen Fu, Peng Shi and Bowen Ning, Predicting the Popularity of Viral Topics Based on Time Series Forecasting 210 (2016), 55-65.

[8] S. Aghababaei and M. Makrehchi, Mining Social Media Content for Crime Prediction, IEEE/WIC/ACM International Conference on Web Intelligence (WI), Omaha, NE (2016), 526-531.

[9] K. Fernandes, P. Vinagre and P. Cortez, A Proactive Intelligent Decision Support System for Predicting the Popularity of Online News Progress in Artificial Intelligence, Lecture Notes in Computer Science 9273, Springer (2015), 535-546.

[10] Cao, Qi and Shen, Huawei and Keting, Cen and Ouyang, Wentao and Cheng, Xueqi. Deep Hawkes: Bridging the Gap between Prediction and Understanding of Information Cascades (2017), 1149-1158.

[11] Bandari, Roja, Sitaram Asur and Bernardo A. Huberman, The Pulse of News in Social Media: Forecasting Popularity, CoRR abs/1202.0332 (2012), 1-8.

[12] Bohra, Navdeep and Vishal Bhatnagar, Group level social media popularity prediction by MRGB and Adam optimization, Journal of Combinatorial Optimization 41(2) (2021), 328-347.

[13] J. Bindra, S. Ahlawat and M. Javed, Modified COVID-19 Indian and international dataset for automatic prediction of risk in an individual using machine learning models using a mobile APP, International Journal of Intelligent Engineering Informatics 9(2) (2021), 142-160.

[14] Charu Gupta, Prateek Agrawal, Rohan Ahuja, Kunal Vats, Chirag Pahuja, Tanuj Ahuja, Pragmatic Analysis of Classification Techniques based on Hyperparameter Tuning for Sentiment Analysis, International Semantic Intelligence Conference (ISIC'21), Delhi, (2021), 453-459.