Personalized Mobile App Recommendation by Learning User's Interest from Social Media

Miss Pratiksha Sawant[1], Mrs.Dr.Nisha A. Auti[2]
[1,2]*JSPM NTC Pune*

***Abstract*— The knowledge and interests of users are valuable resources in social media, but they are frequently neglected. The vast majority of users on social media are not tagged, which means that their interests and areas of expertise are effectively hidden from useful applications such as personalised recommendation, community detection, and expert mining. For example, social media platforms provide only a partial view of what users are known for (by aggregating the knowledge of crowd tagging), but this information is still available. By learning the interest's association between applications and posts, we are able to generate personalised app recommendations for our users. To do this, we employ a one-of-a-kind generative model known as IMCF+, which converts user interest from rich post information to sparse app usage. We conducted an analysis of the performance of this method, which predicts the top ten apps with an accuracy of 82.5% despite requiring only 10%of the data for training purposes. In addition, this purpose technique outperforms the other six state-of-the-artalgorithmsby4.7percentand10percentrespectively in the high sparsity situation and the user cold-start scenario, which demonstrates the efficacy of our technology. All of these findings indicate that our method is capable of reliably extracting user interests from posts in order to contribute to the solution of the problem of providing personalised app recommendations.**

Index Terms: **Social Media, User Profile, deep learning, Privacy, matrix factorization, App recommendation.**

INTRODUCTION

Mobile applications (apps) have made Internet services extremely accessible and convenient. The majority of people who use the Internet today do so through their mobile applications. The people who create apps and the companies that provide network services will benefit from this. In order to provide services of a higher quality and more individualized nature, it is essential to have a solid understanding of the ways in which apps are used in a variety of temporal and spatial contexts. There are a variety of organizations collecting data on app usage, but growing concern has been expressed regarding the privacy implications of mining or sharing such statistics. It's possible that different users will use different collections of apps due to the unique interests and preferences they each possess. In this research, we present a model for transfer learning that we call Interest-aware Matrix Co-Factorization Plus. This model is intended to solve the problem of making tailored app recommendations. This is a generative model that compensates for the absence of app usage data by transferring user app usage and post knowledge to other users. We perform regular language processing on user posts and extract relevant terms to reflect the characteristics that are part of the user's representation in the social domain. Next, we make use of matrix co-factorization to transfer the newly acquired interest into the app domain. This allows the latent characteristics of users to be shared across domains. To ensure that the statistical connection between app latent vector and word latent vector is consistent with their semantic correlation, we quantify the interest relevance of applications and words and factor it into our model. This ensures that the statistical connection between app latent vector and word latent vector is consistent with their semantic correlation. To summarise, this work makes three significant contributions, which are as follows: We are the first to achieve personalised app evaluation by transferring information about user posts from social media platforms. This enables us to learn about user interests while compensating for the lack of data regarding app usage. The second thing that it does is that it finds and learns the underlying latent or interest correlations that exist between these domains. This allows it to transfer the knowledge domain of textual content found on social media into the domain of application usage. And finally, in the high sparsity scenario and the user cold-start problem, it outperforms the other state-of-the-art approaches by 4.7 percent and 10 percent, respectively. This demonstrates its effectiveness in overcoming the challenge of providing individualized app recommendations.

A. Motivation

Within the scope of this study, we investigate one category of the publicly accessible data that users post on social networking sites like Facebook and Sina Weibo, and we investigate the ways in which we can use this data to better understand the preferences of app users. In general, a user's interests and preferences can be inferred from the apps they use and the posts they publish on social media. For example, frequent use of music applications or frequent publications of posts on social media that are related to music are both indicators of music-related interests. It is extremely difficult to determine the association between app usage and sent posts as well as transfer user interest from one domain to another. This is because app usage and sent posts are two distinct representations of user interest. There has been no prior research conducted on this topic, despite the fact that making individualized app suggestions based on information obtained from social networks is a useful and potentially fruitful application. As a consequence of this, with the assistance of our pertinent dataset, we investigate this important problem, which is referred to as tailored app suggestion by gaining an understanding of the user's interests based on social media. It is necessary to first carry out an empirical investigation of the relationship between app usage and posts before we can begin the process of simplifying the design of the methodology for app suggestion.

II. HISTORY & BACKGROUND

Smartphones and other mobile devices have made it possible for users to access a variety of web services virtually anytime and from almost any location. Internet users of today would rather use the mobile applications that are available on their devices than visit conventional websites. In order to provide services of a higher quality and more specifically catered to the needs of individual users, it is now essential for app developers and network service providers to have an understanding of how apps are used in a variety of different temporal and spatial contexts. While a variety of organisations (including advertising companies, internet service providers, and network operators), among others, are collecting data on app usage, growing privacy concerns regarding the mining or sharing of such datasets have emerged. However, many organisations (i.e., app developers, cellular makers, and cellular operators) find it difficult to get individual app usage data (e.g., the list of utilised apps) over a large population due to the high sensitivity and privacy concerns associated with the information. Additionally, it has a low level of detail, which makes it difficult to attract the attention and preferences of users. Because we can't get a complete picture of user interest from the data that's restricted to app usage, a reasonable question to ask is whether or not we can learn about user interest by analysing other data that's open to the public and readily available. In related works, the singularity of human behaviour across various platforms of online services, such as mobility patterns, web browsing histories, and mobile device sensor fingerprinting, has been investigated. These are examples of the kinds of things that have been studied. When it comes to app usage behaviour, researchers rarely have access to the data necessary to quantify the uniqueness of the behaviours exhibited by individual users of the app. Even though there has been a significant amount of research done on app usage modelling, the privacy perspective, in particular the upper bound of privacy leaking over enormous user populations, is not very well understood. This research is the first step toward gaining a better understanding of the consumers' individual app usage habits, which will pave the way for the next generation of privacy protection systems and personalised internet services for mobile users. New algorithms that were developed to deal with these practical variables in deanonymization and location-privacy-preserving techniques have demonstrated promising performance, which confirms our findings. These new algorithms were developed to deal with these practical variables in both of these techniques. In this work, we explored the potential of deep learning by analysing the challenge of user identity linkage. We proposed an end-to-end deep learning system as a means of linking separate accounts based on various types of mobility data. Authorization for a data user to conduct a granular search to be carried out by the owner of the data. The primary concept here is that a data owner will encrypt an index keyword in accordance with a predetermined access policy. If and only if the attributes of a data user are such that they satisfy the requirements of the access policy, then the data user will be able to conduct searches using the encrypted index keyword. The first system to predict Location-level app usage from Points of Interest (POI) using large-scale mobile data accessing records, and the system outperforms three state-of-the-art methods in terms of top-N prediction accuracy and total app usage distribution estimation. A protected personal health record that protects users' privacy and offers granular access control and effective revocation. By encrypting patient health records (PHRs), patients can achieve more granular control over who has access to their information by associating an expressive access tree structure with the ciphertext. Authors who make use of anonymous key issuing protocol are also successful in protecting their readers' privacy. The author profiles on Google Scholar are crawled and analysed in order to conduct this investigation. By utilising a distributed crawling strategy, we were able to collect 812.98K author profiles from Google Scholar. This strategy covers the vast majority of author profiles that are publicly available, if not all of them. In this line of research, the collaborative variational auto-encoder has been proposed as a method that can jointly model the development of item content while simultaneously extracting the implicit associations that exist between users and things. It is a Bayesian probabilistic generative model that connects collaborative and content information through the use of stochastic deep learning models and graphical models, which ultimately leads to reliable recommendation performance. They begin by recognising the significant roles that various types of ties play in social relationships. Following this, they present a novel social recommendation model, which is a non-trivial extension of probabilistic matrix factorization. This model incorporates the personalised preference of strong and weak ties into social recommendation, and it does so by using probabilistic matrix factorization. UTop+ is a generalised method for learning high-quality user thematic profiles that combines numerous implicit and explicit footprints. It was first introduced by. In the system that we have proposed, the ability to provide individualised recommendations for mobile applications through the transfer of user post data from various social media platforms is covered in this article. When we compared the performance of our method to that of five other state-of-the-art algorithms, we found that our purpose method performed significantly better. According to the findings of the research, app prediction is affected by both user and app characteristics. Our study represents a significant advancement in the process of transferring user data from social media to learn personal app preferences. This opens the door to higher quality customised mobile app recommendations and services for mobile consumers.

III.LITERATURE SURVEY

Using public tweets, the researcher demonstrates in this paper [1] that it is possible to learn about the user's app and his or her own interests in order to compensate for the limited amount of usage data that is available. We create a new generative model that we call IMCF to predict each user's interest in an app in order to make the connection between people's immediate interests and their use of apps.

In this paper [2], the authors propose a deep learning connection end-to-end architecture that they call DPLink. Its purpose is to assist with heterogeneous mobility data that has been obtained from various services.

This paper [3] presents the findings of the first study to examine the use of smart phones across an entire population. During the course of our investigation of a dataset originating from the city of Shanghai, we compiled more than 6 million distinct devices, which over the course of a week generated more than 10,000 distinct applications that were tagged with GPS coordinates. The framework makes use of transfer learning to make predictions regarding which applications will be popular, and it also makes use of data from POI to make predictions regarding the overall distribution of applications based on location.

It has been hypothesised that utilising social networking sites that are cross-linked with e-commerce accounts could be of assistance in mapping people's social characteristics onto another product feature for the purpose of recommendation.

In addition, we argue in favour of using recurrent neural networks for the purpose of inferring consumer and product representations from data that is gathered online.

In this paper, we propose a graph embedding scheme that we will refer to as GE [4]. Which, from a conceptual standpoint, integrates the sequential, regional, temporal, and semiotic impacts, but defines them all within a single lower-dimensional coordinate system? The authors create a novel time-decay approach that makes use of POIs that are embedded in the latent space in order to learn and return the user's current desires in real time in order to capture real-time interests. This allows the authors to capture real-time interests.

In this paper [5], the authors propose an innovative method for the improved analysis of social relationships that is predicated on a probabilistic factorization model. The purpose of this brand-new study, which was recently published, is for the author to demonstrate that people who use smart phones do not all belong to the same type of demographic. After analysing the behaviour of 106,762 Android users regarding the apps they use, we uncovered two months' worth of data and identified 38 distinct user personas based on the capabilities of the apps.

An author proposes[8] a method for predicting human mobility by using various data, which is created from online sources, on the behaviours of his website's users. This method makes use of machine learning to accomplish this. Our strategy is based on the principle of selecting a group of behavioural characteristics (such as location, travel habits, and mobility attributes) and incorporating them into a classification scheme. This is the central tenet of our approach.

The central idea behind this [9] method is to select a number of mobility characteristics (such as location, travel pattern, and mobility indicators) that have an impact, and then incorporate those characteristics into a classification model for the use of apps. Our in-house developed high-speed traffic monitoring system for real-world network traffic is used in the evaluation of this approach (TMS). In our experiment, this method of prediction achieves a precision of 90.3 percent, which verifies the obvious connection between human mobility and application behaviour?

Cold starting is one of the most[10] important challenges that such processes face. Cold starting occurs when there is no prior experience present and a new item is introduced into the system, which means that there are no successful suggestions that can be made. In point of fact, cold starts are a very popular issue: hundreds of new pieces are released on a regular basis across various modern web platforms. We propose that you learn Local Collective Embeddings as a means of resolving this issue. Local Collective Embeddings is a factorization matrix that makes use of property and previous user expectations in order to apply the various structures of group embedding. We have developed an instructional algorithm that is based on a number of easily implementable update laws.

IV. PROPOSED SYSTEM

Because of its widespread application and usage, a significant amount of research has been conducted in this area. In this section, a few of the strategies that have been utilised in the past in order to accomplish the same goal as before are discussed. Techniques for multi-keyword searches and group sharing systems are two factors that significantly set these works apart from one another. The dataset does not include applications that make no network requests or applications that make network requests primarily through the WiFi network. In conclusion, all of our data is collected from a solitary point of origin for network services. Techniques that protect users' location privacy can make use of spatiotemporal mismatches to better protect users' location privacy while also increasing the utility of the trajectory information. This can be done while maintaining or even improving the utility of the information. The multi-modal embedding module needs to be expanded in order to process the raw textual information that is contained in the check-in data. The global co-authorship network is comprised of authors from a wide range of disciplines. As a consequence of this, an approach that draws on multiple disciplines is required in order to investigate this extensive network.

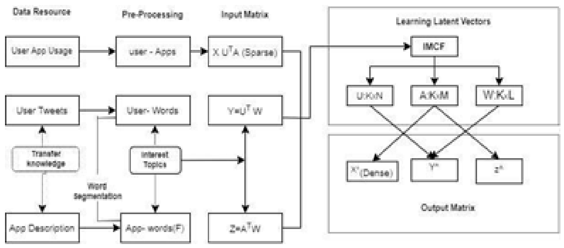

Fig. 1. Proposed System Architecture

A. Algorithms

In IMCF+ model, we evaluate the performance of six state- of-the-art baselines as follows:

- Interest-aware Matrix Co-Factorization (IMCF):It is a degraded version of our IMCF+ that ignores the addition app-word matrix Z, i.e., in our model, setting 6= 0, 0 =
- The baseline is used to demonstrate the value of app functional descriptions, as it assures that the statistical correlation based on latent feature vectors corresponds to the actual interest correlation between apps and post terms.
- Collaborative Topic Regression (CTR): The intention of recommending articles to readers in its original form. On the one hand, collaborative filtering is used to learn latent characteristics of users and articles from the user- article matrix; on the other hand, LDA is used to learn another part of article latent features from the article's word information. Because users have external post word information in our scenario, we learn two components of their latent features and integrate them to generate the final user latent vectors.
- Neural Collaborative Filtering (NCF):In this method re- places the inner product of matrix factorization with a neural architecture, making it the most recent Neural network-based Collaborative Filtering method. It inputs the user-app interactions and uses a multi-layer perceptron to learn the user-app interaction function to super- charge NCF modelling with non-linearity.
- Non-negative Matrix Factorization (NMF):It is a non- negative matrix factorization method with the condition that all matrices have no negative elements, which is a common method in recommendation systems. The user- app matrix is the only one used by NMF.
- Single Matrix Factorization (SMF):It only looks at the user-app matrix and ignores the rest of the data. To complete this common collaborative filtering work, we use a popular low-rank factorization algorithm. SMF, unlike NMF, does not require that all matrix elements be non-negative. In the scenario where we set = 0 and = 0, SMF is indeed identical to our technique.
- K Nearest Neighborhood (KNN): It takes into account both user app usage and post data. KNN calculates the nearest K neighbour users for each user based on the user-post matrix Y, then predicts app usage in the testing set based on these neighbours' app usage characteristics. Assume the post similarity between user Ui and his K neighbors.

V. RESULTS AND DISCUSSION

The Experiments are done by personal computer with configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and JDK 1.8. The application uses web application tool for design code in Eclipse and execute on Tomcat server.

Positive (P) : Observation is positive. Negative (N) : Observation is not positive.

True Positive (TP) : Observation is positive, and is predicted to be positive.

False Negative (FN) : Observation is positive, but is predicted negative.

True Negative (TN) : Observation is negative, and is predicted to be negative.

False Positive (FP) : Observation is negative, but is predicted positive.

Accuracy = TP + TN / (TP + FP + TN + FN)

Precision = TP / (TP + FP) Recall = TP / (TP + FN)

F1-Measure = 2 * Precision * Recall / (Precision + Recall).

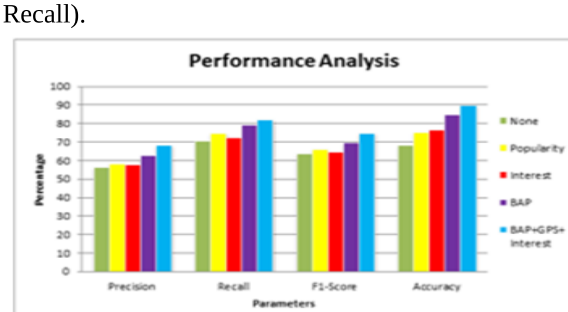

Fig. 2. Accuracy Graph

	Existing System	Proposed System
Accuracy	79.32	89.77

VI.CONCLUSION

In this paper, we demonstrated the capability of providing individualized recommendations for mobile applications by transferring user post data from various social media plat- forms. When we compared the performance of our method to that of five other state-of-the-art algorithms, we found that our purpose method performed significantly better. According to the findings of the research, app prediction is affected by both user and app characteristics. Our study represents a significant advancement in the process of transferring user data from social media to learn personal app preferences. This paves the way for improved mobile app recommendations and services that are tailored specifically for mobile consumers.

REFERENCES

[1] Z.Tu,R.Li,Y.Li,G. Wang, D. Wu, P. Hui, L. Su, and D. Jin, “Your apps give you away: distinguishing mobile users by their app usage fingerprints,” Proc. ACM UbiComp, vol. 2, no. 3, p. 138, 2018.

[2] H. Wang, C. Gao, Y. Li, G. Wang, D. Jin, and J. Sun, “Deanonymiza- tion of mobility trajectories: Dissecting the gaps between theory and practice,” in The 25th Annual Network Distributed System Security Symposium (NDSS18), 2018.

[3] J. Feng, M. Zhang, H. Wang, Z. Yang, C. Zhang, Y. Li, and D. Jin, “Dplink: User identity linkage via deep neural network from heterogeneous mobility data,” in The World Wide Web Conference. ACM, 2019, pp. 459–469.

[4] H. Wang, C. Gao, Y. Li, Z.-L. Zhang, and D. Jin, “From fingerprint to footprint: Revealing physical world privacy leakage by cyberspace cookie logs,” in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM, 2017, pp. 1209–1218

[5] D. Yu, Y. Li, F. Xu, P. Zhang, and V. Kostakos, “Smartphone app usage prediction using points of interest,” Proc. ACM UbiComp, vol. 1, no. 4, p. 174, 2018

[6] L. Xiao, Z. Min, Z. Yongfeng, L. Yiqun, and M. Shaoping, “Learning and transferring social and item visibilities for personalized recommendation,” in Proc. ACM CIKM. ACM, 2017, pp. 337–346

[7] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, “Neural collaborative filtering,” in Proc. ACM WWW. International World Wide Web Conferences Steering Committee, 2017, pp. 173–182

[8] X. Li and J. She, “Collaborative variational autoencoder for recom- mender systems,” in Proc. ACM KDD. ACM, 2017, pp. 305–314

[9] X. Wang, S. C. Hoi, M. Ester, J. Bu, and C. Chen, “Learning personal- ized preference of strong and weak ties for social recommendation,” in Proc. ACM WSDM, 2017, pp. 1601–1610.

[10] C. Cao, H. Ge, H. Lu, X. Hu, and J. Caverlee, “What are you known for?: Learning user topical profiles with implicit and explicit footprints,” in Proc. ACM SIGIR, 2017, pp. 743–752.