

HYPERLINK INDUCED TOPIC SEARCH MODEL TOGETHER WITH AUTOMATIC FEATURE REVIEW FOR SMARTPHONE APPLICATIONS

John Vaseekaran^{1*}, Dr. N. Srinivasan²

¹*Research Scholar, Sathyabama Institute of Science and Technology, Chennai – 600119.*

²*Adjunct Professor, BITS-PILANI, Chennai Centre, Chennai.*

Running Title: HITS based Feature review for smartphone applications

*Corresponding Author: vaseekaranjohn31@gmail.com

Abstract

The demand for smartphone apps has grown with the rising interest of artificial intelligence. Thanks to a vast number of applicant service applications, choosing the smartphone apps you want to use has been very complex for consumers. It is therefore essential that the customer interface is improved and that individual suggestions are made. Conventional recommendation approaches can in some cases be effective but have some drawbacks, which generally lead to unreliable recommendations. This study provides a basis for recommending smartphone applications, which is built on the algorithm of Hyperlink Induced Topic Search (HITS) in conjunction with association rule mining in this context. The approach combines the scores of authority and hub into the applications by means of downloads and ratings and not only takes into account the role of smartphone apps in alliance rules but also the trustworthiness aspect of consumers. Studies with industry data sets from the Samsung framework reveal that the proposed approach increases the recommendation precision greatly relative to conventional approaches.

Keywords:hyperlink, factorization, ranking algorithm, mobile application, association rules

Introduction

Smart phones have increased remarkably in the last decade with the advent of smartphones Internet. Global iOS and Google Play downloads have reached new heights in 2017. While the rapid growth of smartphone applications has brought enormous economic benefits and competition, it is still really intense in the smartphone apps industry. The smartphone apps industry has also focused on helping consumers identify appropriate applications easily. In order to consider users' behaviour, both domestic and international applications industries use advice technologies [1].

An intelligent machine that plays an important role in helping the consumer choose appropriate applications for customized suggestions is a recommendation system. The most classic suggestion algorithm is now collaborative filtering. It contains collaborative memory filters and model-based collaborative filters. Model collaborative filtration primarily involves collaborative user-based filtering and collaborative filtering based on objects. The key concept is to forecast missed values of common consumers or utilities. The selection of the effective similarity estimation tool is an important aspect of collective filtering.

The Pearson rank correlation and cartesian coordinates similarity are two widely used equations. When the service composition matrix is relatively dense, the collective filtering process is very efficient. Customers typically want only just several applications, leading to a very narrow matrix. In this situation it is impossible to identify comparable customers or similar utilities for apps of conventional collaborative filters so that the

forecast predictions are less reliable. In addition, certain consumers can maliciously collect such resources when testing mobile apps and these outliers influence the outcome of the experiment [2].

Matrix factorization should be used for dividing the customer score and the programs into two parts and foresee a missed point by choosing the bigger value and reduce dimension. The simple procedure for factoring matrices is shown. Training is therefore dynamic, the suggested outcomes are not very comprehensible even where a lot of data is available, the recommended results are not reliable. In the field of data mining, the association rules algorithm is significant and is now commonly used in recommended areas. The aim is to find in the repository secret knowledge and fascinating trends. Frequent object sets are collected in the request log by calculating the item sets.

The relation between item sets is exploited by the pruning and relations between the item sets and the future value is detected in the transaction database. Most accesses of the transaction database are required in the rules assembly process and, by contrast, the value of various item sets is identical, the significance of item sets may vary and the removal of a large range of organizations is more difficult. New challenges and enhancements to current recommendations technology are brought on by the development of edge computation. The traditional methods of recommendation cannot be used for smartphone application suggestions.

First, the smartphone app recommendation is frequenter than the conventional product recommendation. Faster upgrades to apps and those modifications can be of concern to consumers. Mobile telephones have lower cost and consumers can use a variety of applications at high volume as a software product. Secondly, ethical sustainability affects the advice of the smartphone app. Subscribers may have various characteristics under different circumstances with the use of the same smartphone application. Finally, the smartphone app is an enduring consumer and its use will change the download behaviour of consumers in future. The customer can download this new software and test it if he or she notices the suggested app that is more app appealing than the downloaded one. Therefore, the smartphone app recommendation is customized.

This paper suggests a solution to the problems in the traditional methods of suggestion method of estimation of HITS-based weighted association laws. The trust in applications and the user's experience is calculated in terms of the authority of the apps and the user interface ratings. The power and the core are paired with the principles of association for mining relationship among various applications. Eventually, the consumer is advised to use apps. The biggest issue with the content-based image search is the so-called semantized discrepancy as content-based searches are linked to low-level characteristics, whereas people search for high-

level definitions. Automatic image annotation methods have been created to solve this problem: processes through which computers automatically add metadata to pictures in the form of captions or keywords [3].

The most famous one is definitely on learning by example paradigm. For model testing, a limited collection of manually noted training images learn the association between image characteristics and high-level text words and enable an automated description of other visually invisible images. Strong teaching examples are obviously vital here, such as representative and correct pairs of photographs and associated marks. Social networking, and Instagram in particular, provide a rich image-tag couple source. Automatic or semi-automatic mining of the right ones is highly valuable for training examples. However, in several instances, hashtags accompanying images of social media do not contribute to the content of the post, but serve many other functions such as the emotional expression of the user, increased clicks and discovery capabilities for users and the beginning of new conversation or debate.

The fundamental principle is that a website should provide topic information and links to a subject. Web sites thus belong to two classes, which offer good knowledge on an issue which is called authentic and which gives the user good links to a subject such as hubs. The HITS algorithm provides a hub and a definitive meaning for any web page. In one of our previous works, we have begun to experiment with the HITS algorithm for mining informative Instagram hashtags and extended this paper to consider the HITS algorithm in actual crowd tagging environments enabled by the crowdsourcing network. We also expanded approximation numbers by 250 in each image, formed two-part charts for all pictures, and measured annotator output over all of these images. In addition, FolkRank has been used as a basis for evaluating the method's efficiency [4].

Existing Work

Several investigators tested and checked the authenticity of the crowd-sourced image annotation. Mistry et al. [25] contrasted the quality with that in experts of crowdsourced picture rating. They used 100 photographs of the retinal fundus, chosen by two researchers. Each annotator was asked to grade eighty-four retinal photographs, and 16 practice-training images first tested the capacity of annotators to properly identify these images. The researchers indicated that inexperienced individuals were equal in their success in retinal picture categories to experts.

In this challenge of leaf counts in the images of *Arabidopsis Thaliana*, Giuffrida et al. [15] calculated the inconsistency among expert or unexperienced users. Everyday people can accurately count their leaves according to their performances. The efficacy of large-scale crowdsourcing for endoscopic photographs was tested by Maier-Hein et al. [18] and the untrained staff performed comparatively with the medical expert.

Cabrall et al. [3] used the crowd to list driving scenarios such as the involvement for categorizing other road users and bikes or pedestrians. In the classification of vast volumes of videos with different scene contents they used the tool Crowdfunder.

As standard, Crowdfunder's Gold Test Questions have been used to ensure that the annotators are working properly. The findings show that Crowdfunder's crowdsourcing was successful in classifying the contents of naturalistic scenes. The HITS algorithm was initially used for the discovery and assessment of web pages that relate to a topic. The Hits algorithm is used in the social network analytics to assess the centrality of nodes, especially in networks consisting of 2 node types known as two-mode networks. It contains the hub and authority values it calculates. The bipartite networks that are typically modelled on bipartite graphs are a common example of such networks.

A bipartite graph is a graph whose nodes can be separated into two distinct classes because its boundaries are not inter-partition nodes. In order to model suggest schemes, two-mode networks [23] also are used, given that users and goods belong to two distinct kinds of organizations. Mao et al. [22] used HITS in the social identification scheme to enhance user profiling. User profiling aims at understanding and coding consumers' personal preferences so as to provide innovative, customized services. The social labelling framework has been designed as a user-tag network, PageRank and HITS is added to refining label weights. The calculated tag weights were then used for a diffusion method in the tag-item bipartite graph of the array. The tests on three various data sets revealed that the proposed solution exceeds the conventional collective filtering method used in the recommender programs. The results are dependent on tags.

Zhang et al. [17] sought to draw feedback from users on features like smart phones, laptops, etc. They built a two-mode network with features designed as bodies and metrics of interest as hubs to determine the value of those features. By thresholding respectively, the relevant authority and hub values, the HITS algorithms were utilized to classify highly relevant features and good characteristic indicators. The HITS algorithm variant called GeoHITS was used by Tri and Jung [10] for the classification of places as regards basic tags, such as food categories. Both tags and positions on social network services were gathered from geotagged tools.

A sub-set of tags is used by author writers, who were exchanged around many places and were regarded as hubs. Many experts have examined the guidelines for mobile apps. Mobile platforms offer various kinds of information and support consumers in taking decisions to satisfy the needs of different users. This kind of detail Many experts have examined the guidelines for mobile apps. Mobile platforms offer various kinds of information and support consumers in taking decisions to satisfy the needs of different users. Included in this

data are applications, customer reviews, feedback and installation numbers. The subject of research on major platforms has been on how to pick applications that are interesting for users from the vast number of applications and give users customized recommendations. Some studies have suggested the extension and adaptation of standard recommendation algorithms to the application domain.

Yan et al. proposed AppJoy, which makes use of past user history to create the preferential matrix and then makes CF suggestions based on items. The sparseness of the matrix, however, affects the precision of the recommendation. In an implementation analysis, Falaki et al. [15] used a method. However, by using the collected evidence, they did not make suggestions. Davidsson et al. [16] suggested a solution that incorporated user positions, including the fact that different apps could be used by different users in the same geographical area. This strategy does not, however, take into account consumer expectations.

Cui et al. [17] used mobile-based download guidelines to assess the applicable download habits and suggested a probabilistic proposal. The approaches described above are focused on advice of the users and past input from applications and do not take into consideration possible relations between applications. Evil users with malicious comments cannot be deleted. Liang et al. suggested a functional factorizing matrix approach that turned the user rating estimation into a rating prediction problem and coupled it with knowledge about the user preference. The matter of sparse data was also solved.

A recommendation approach was proposed by Xu et al. [18], which could produce accessibility interface recommendations. They take the functional needs of consumers into account and also recommend an efficient approach to the extraction of functions. But the impact of user ratings on the findings was not taken into account in their process. You simply think about the user's role choice and neglect the user's functions knowledge. To make suggestions on the mobile app, Wu et al. introduced two hybrid models focused on shared filtering. Our approach takes into account not only user experience, but also the importance and internally relevant relevance to various implementations, which can increase the quality of the recommendation accurately and substantially.

The algorithm for the connection rules is a classical data mining algorithm. The university is now applying affiliation guidelines to individual recommendations. Liu et al. [19] investigated how people with similar tastes and desires would receive tips on Weibo. They suggested a data mining-based suggestion algorithm. They incorporated forwarding and interacting features into their associated users, combining user knowledge characteristics, and used a refined Apriori algorithm to increase the precision of their suggestions. In order to differentiate cases of fraud from common reports, Cui et al.[6] suggested an approach for detecting

fraud focused on the trusted nature of physicians. They also developed a two-mode physician-patient network, which was seen as a bipartisan weight graph. The medical notes were used to calculate edge weight for the prescription behaviour.

According to the authors, the HITS algorithm's hub scores offer a fair assessment of doctors' credibility. The new HITS algorithm called Co-HITS was used by London and Csentes [22] for the technical assessment of wine tasters. They created a weighted two-part graph consisting of wine tasters, configured as centres, and beverages, configured as authority in order to accomplish this aim. The weights match the scores given to the wines by the wine tasters. The authors claim that calculated hub values can be used to root out inept tasters whilst being closely linked to a wine taster's skills.

Tseng et al. [14] have sought to differentiate between fake remote and regular telephone calls because the confidence value of distance telephone number has to do with the HITS algorithm hub ranking. To this end, they also used telecommunications records to provide guided two-part charts that provide incoming and incoming calls between users' contact book entries, presumed by government, and remote telephone numbers assumed as hubs. Based on time and frequency associations between a user's and a remote telephone number, the edge masses for each user pair and the telephone number are measured.

Using HITS, the trust value was calculated and used for every remote telephone number to identify remote calls as fraudulent and natural. Also, there are a few experiments in which the HITS method is used as we do in this article for the particular case of image tagging in a crowd-source context. In most cases, however, focus is given not to knowledge mining but to an assessment of quality improvement of the data from the crowdsourcing. The HITS algorithm was implemented by Sunahase et al. [16] in three separate tasks: the image classification, the logo design, and the article translation.

The HITS algorithm is applied to comparisons between them. Sunahase et al. The objective was to evaluate the data content provided by comparing the image specifications, logo designs, and article interpretations generated in parallel by two separate creators and data producers. The evaluators were then able to evaluate these data.

Proposed System

HITS refers to Hypertext Induced Topic Search, and it is an Internet knowledge discovery query-dependent rating algorithm. There are two fundamental properties on each page of the HITS algorithm: the hub domain and the authority. A page of authority applies to a page which has many page connections. A portal page is a website indicating a number of other sites. The trick to the HITS algorithm is that many good authority

pages are marked with a good hub page and many pages with good authority. There is a reciprocal strengthening partnership between authorities and hubs. The algorithm returns a high-end authority page and a high-end hub page. If you have access to the latest page q , the cumulative number of web pages is n , and the jurisdiction and hub is q .

Relationships are described by the link page:

$$\forall q, b(q) = \sum_{j=1}^m h(j)$$

$$\forall q, h(q) = \sum_{j=1}^m b(j)$$

So,

$$\sum_{j=1}^m h(j)^2 = \sum_{j=1}^m b(j)^2 = 1.$$

The weight and current iteration of the preceding is calculated. If no major modifications occur, it ensures that the calculations will be fundamentally stable and completed. There was convergence of authority and hub ratings. The objective of the algorithm for association rules is to detect the connections and associated relationships in several item sets and to define the likelihood of implied connections of various data in the database. The association rules gain support and trust within sets, according to the above meanings. The high-frequency sets are screened by raising the resource dependency, and by lowering the confidence threshold, item sets with high correlations are also chosen. The closely associated items are finally obtained in the transaction database. In this section, we first explain the general context for the HITS-based model of the recommendations for association rules and then use the HITS template to enter the user reviews matrix for the next association rule.

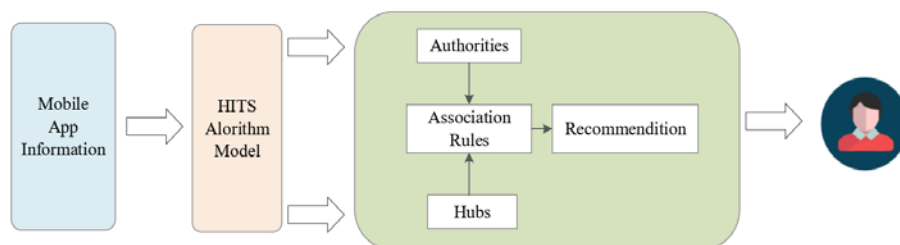


Figure 1. Proposed System architecture

Finally, in this article you can find the pseudocode of the model algorithm. A user/app network layout map was developed according to the past behavioural records of users of the app. Users were identified as hub nodes and applications as authority nodes through the concept of the HITS model. Using the relationships

between the driven edges, the HITS algorithm is used in the calculations, and the User Score matrix is applied, so that the power and the hub scores after every iteration can be updated before the authority and the hub overlap. Weighting variables for each customer and application would be authority and hub.

In order to quantify possible associations between various applications, the improved Association Rules were used and new help and confidence grade estimation formulas were defined. The application sets with a high validity are determined by applying association rules of authority and hub weights, which are more consistent with the actual suggestion. The method not only takes into account application relevance, but also user interface. After multiple iterations, the rules of partnership are obtained according to the threshold and the rules of association are established in terms of trust from the largest to the smallest.

Then we choose the top-k applications in the training collection to recommend them to each user according to the top-k theorem. This ensures that the list of users recommended is collected. Apriori is an algorithm often used to locate frequent object sets. The theory is that all its subitems are frequent items if an object is a frequent item. In the meanwhile, all the supersets are therefore non-frequent if an object is a non-frequent set. In accordance with the algorithm Apriori and the candidate item collection is scanned for iterative lay-by-layer method of searching. The Apriori algorithm uses the search utility a priori.

In the modern environment, the relevance of various applications can vary, and the user ratings relation is often different. Weighting of applications and customers is then enforced and the recovered responsibility and hub is attached to help and confidence estimation formulas by using the HITS algorithm to incorporate score information not only taking into account score information, but also introducing app and device weight information, enhancing the outcome of the recommendation with the situation of the actual recommendation. To locate any frequent object in the standard Apriori algorithm, the index must be scanned.

The utility of this costly algorithm is significantly diminished if we have a large number of records. Furthermore, if the weight of each object is not taken into account, a vast number of rules will be detected and the advice for applications is insignificant. In order to evaluate a set of web pages appropriate for one subject and find the "authoritative" ones in the topic, the HITS algorithm was initially implemented in [20] and [19]. On these web pages, it conducts connect analytics to classify them as two measures: Hub Meaning and Authority. The score of the authority estimates the content of the website, while the score of the hub estimates the consistency of its links to other sites.

Thus, a web page that has many links from other high-Hub sites is regarded as an authority, while a website with many links is a hub. In short, the HITS algorithm's key concept is that a knowledgeable hub

indicates many efficient authorities, and many informed hubs indicate an effective authority. There is also a partnership between authorities and centres of reciprocal strengthening. The HITS algorithm is used for two-mode analyses of the networks described as bipartisan, as previously stated in the introduction. In this case, the ideals of authority and hub are used as centrality measures, but their meaning varies considerably. A high-performance vertex is called a specialist, while a high-performance hub is considered to be a good recommendation.

Results and Discussion

In this segment, we track the generalized use of the Samsung mobile application market with a large number of experiments and compare the findings with the comparatively newer methods and traditional methods. Besides this, we research the effect on the precision of the exploratory forecast of various parameters. The data collection used is the download/score record for actual applications. Any record shows the one-time actions of a consumer. The dataset which we used in the investigation was derived from the actual Samsung application market download/score records. The information package contains the name of the app, user ID, ranking, time for feedback and model phones. The data collection includes information. There is a very sparse user app matrix in the real world, and people have used only a few applications. In order to resolve this issue, more than 20 smartphone applications have been downloaded and commented by each user, and over 20 users have downloaded and commented on them, which is more in line.

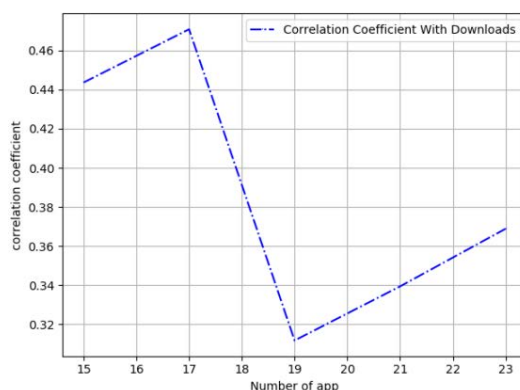


Figure 2. Performance of the proposed system

The dataset included 13869 documents after testing, including 2400 users and 768 applications. Thus, the similarity of the outcomes of algorithm which are applied to the HITS algorithm can be analysed by the above formulas. The association between rankings of authority in Algorithm, and download volume and average score on Huawei's mobile device market is measured in this paper by the Pearson Correlation Coefficient and the Spearman Correlation Coefficient. The following briefly explains how the two calculations differ. The

coefficient of the Pearson correlation is used to represent the extent to which two variables have similarities. The significance of the features and categories derived can be calculated by calculating the similarities between characteristics and categories. The following theorem explains the Pearson correlation coefficient:

$$\rho = \frac{\sum_{j=1}^m (a_j - \bar{a})(b_j - \bar{b})}{\sqrt{\sum_{j=1}^m (a_j - \bar{a})^2 \sum_{j=1}^m (b_j - \bar{b})^2}}$$

Where a and b are the two consecutive vector set, \bar{a} is the mean vector a , and \bar{b} is the average vector b , for Pearson coefficient of correlation [1,1]. Here, ρ means that there is no connection between the two vector sets, that 1 means that there is a significant correlation among the two vector sets, and that -1 means that there is a linear relationship. In the formula above, we analyse the association between the algorithm outcomes and the installs and the overall app scores. They show the coefficient of association between top-k app officials and downloading applications and average scores in Pearson.

The correspondence with the authorities and downloads between the Pearson coefficient is around 3-0.49, as seen. The association between power and the average score of Pearson is around 0.21-0.42. In summary there are some strength correlations of authorities measured using algorithm with downloads and app scores. Thus, it is worth adding user ratings to the HITS algorithm, that has a certain effect on test results and is more consistent in real life with the customized suggestions. In addition, the coefficient of Spearman correlation is used to measure the correlation of downloads and average scores between the application authorities the following is the determination of the coefficient of Spearman correlation. The coefficient of Spearman correlation is close to the coefficient of Pearson.

$$\rho_t = \frac{\sum_{j=1}^m (M_j - \bar{M})(N_j - \bar{N})}{\sqrt{\sum_{j=1}^m (M_j - \bar{M})^2 \sum_{j=1}^m (N_j - \bar{N})^2}}$$

The distinction is that the values will be changed to ranks. In this case, M_j and N_j are the respective value levels of j . \bar{M} and \bar{N} are the average levels of a and b , and t are the cumulative number of observations respectively. The coefficient of Spearman is identical to the coefficient of Pearson correlation. The distinction is that the values will be changed to ranks. In this case, M_j and N_j are the respective value levels of j . \bar{M} and \bar{N} are the average levels of a and b , and t are the cumulative number of samples respectively. They reveal that the average accuracy declines as the number of recommended applications grows and the recall increases. The suggested accuracy and reminder are interconnected and interconnected adversely.

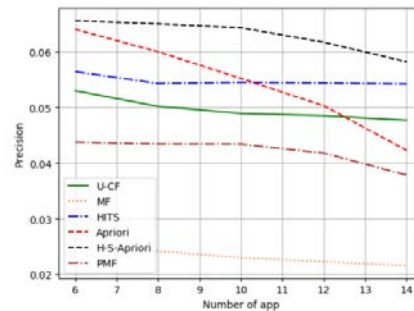


Figure 3. Comparison of proposed system with others

The increased number of suggested applications means that the user's actions list is constant in the test set and there are further applications which have been properly recommended, which contribute to a greater call-back. The proposed list of improvements in training sets determined by the usage experience and the number of accurate recommended applications does not increase, which leads to a reduction in accuracy. The approach in this paper is enhanced by association rules and we thus review the impact of sponsorship on common items. We should set a fair aid level to ensure that the laws of the mining association are more correct. The effect of the help on frequent 1-item settings, frequent 2-item sets and Association Rules apps, is mentioned.

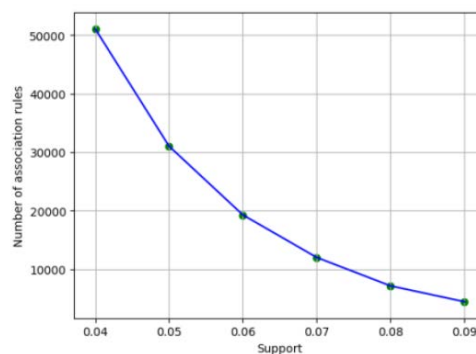


Figure 4. Learning support model

The statistic shows that the number of regular two-point sets is even higher than the number of frequent one-point sets. Increasing orders significantly increases the amount of often used items mined by the algorithm. Therefore, to remove a significant number of obsolete association regulations, the number of Frequently mined items from the algorithm should be monitored. They reveal that the average accuracy declines with the increasing number of recommended implementations as recall increases. The suggested accuracy and reminder are interconnected and interconnected adversely.

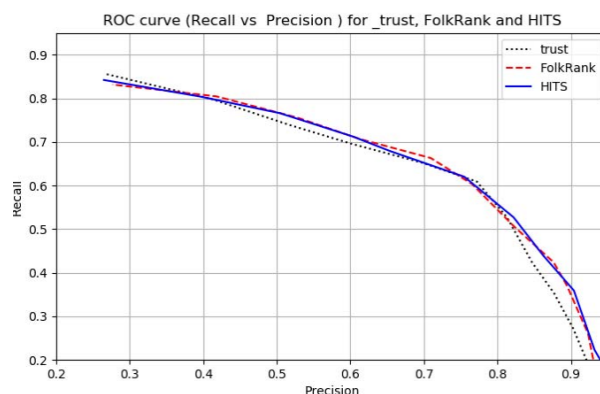


Figure 5. Precision vs recall comparison

As the number of suggested applications increases, the list of users in the test range will remain constant and the number of correctly suggested applications will increase and the callback will increase. The recommended collection of consumer habits measured in the training sets does not, however, increase or decrease the number of correctly recommended users. The approach in this paper is enhanced by association rules and we thus review the impact of sponsorship on common items. We should set a fair aid level to ensure that the laws of the mining association are more correct.

The list of the effects of assistance on the applications of regular one-point sets, frequent two-point sets and association rules. The statistic shows that the number of regular two-point sets is even higher than the number of frequent one-point sets. Increasing orders significantly increases the amount of often used items mined by the algorithm. Therefore, to remove a significant number of obsolete association regulations, the number of Frequently mined items from the algorithm should be monitored. They reveal that the average accuracy declines with the increasing number of recommended implementations as recall increases. The suggested accuracy and reminder are interlinked and interconnected adversely.

As the number of suggested applications increases, the list of users in the test range will remain constant and the number of correctly recommended applications will increase and the callback will increase. The recommended collection of consumer habits measured in the training sets does not, however, increase or decrease the number of correctly recommended users. The approach in this paper is enhanced by association rules and we thus review the impact of sponsorship on common items. We should set a fair aid level to ensure that the laws of the mining association are more correct. They list the effect of assistance on the applications of regular one-point sets, periodic two-point sets, and association law.

The statistic shows that the number of regular two-point sets is even higher than the number of frequent one-point sets. Increasing orders significantly increases the amount of often used items mined by the algorithm.

Therefore, to remove a significant number of obsolete association regulations, the number of Frequently mined items from the algorithm should be monitored. Annotators based on the hub scoring and the algorithm FolkRank are similar as we all observe major variations between them based on the same theory and the resulting rating is, in nearly all situations, in the last 200. The cumulative number of annotators $N = 249$ is recalled here.

Conclusion

There are several issues with the conventional application recommendation process. This article provides a method that incorporates the HITS algorithm with the rules of association. It incorporates user ratings in order to iteratively quantify power and hubs and then links the results to alliance laws and redefines trust and trust estimation formulas. Finally, confidence-based implementations are recommended to users to prevent the issue of high time usage of standard association rules, increase the performance of the algorithm and address the problem of fraudulent user ratings effectively. The findings show that the suggested approach has excellent success in terms of accuracy and reminder by experimentalizing the consumer records on the Samsung mobile app market. Although there are still some restrictions in our solution. First, we have strengthened our Apriori algorithm, so that the over time is still huge and the mining recommender systems are not highly efficient in the face of large data. In future work, we shall update the association rules on mining, reduce overtime and data redundancy and improve association rules' mining performance. Secondly, our approach only takes into account customer reviews with low data size and medium suggestion action. We will also be working to add additional detail on product features and customized variables for apps such as customer requirement, user details and user feedback in the future.

References

- [1] Agarwal, K. K. Bharadwaj. "A collaborative filtering framework for friends recommendation in social networks based on interaction intensity and adaptive user similarity," Social Network Analysis and Mining, 2013, 3(3): 359-379.
- [2] Ashleydejo, S. Ngwira, T. Zuva. "A survey of Context-aware Recommender System and services," in Proc. 2015 International Conference on Computing, Communication and Security (ICCCS). IEEE, 2015: 1-6.
- [3] Bellogin, J. Parapar. "Using graph partitioning techniques for neighbour selection in user-based collaborative filtering," in Proc. Proceedings of the sixth ACM conference on Recommender systems. ACM, 2012: 213-216.

- [4] Cui, K. Liang. "A Probabilistic Top-N algorithm for mobile Apps recommendation," in Proc. 2013 5th IEEE International Conference on Broadband Network & Multimedia Technology. IEEE, 2013: 129-133.
- [5] Davidsson, S. Moritz. "Utilizing implicit feedback and context to recommend mobile Apps from first use," in Proc. Proceedings of the 2011 Workshop on Context-awareness in Retrieval and Recommendation. ACM, 2011: 19-22.
- [6] Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, D. Estrin. "Diversity in smartphone usage," in Proc. Proceedings of the 8th international conference on Mobile systems, applications, and services. ACM, 2010: 179-194.
- [7] Gomezuribe, and N. Hunt. "The Netflix Recommender System: Algorithms, Business Value, and Innovation," ACM Transactions on Management Information Systems (TMIS), 2016, 6(4): 13.
- [8] Ha, D. Wagner. "Do Android users write about electric sheep? Examining consumer reviews in Google Play," in Proc. 2013 IEEE 10th Consumer Communications and Networking Conference (CCNC). IEEE, 2013: 149-157.
- [9] Liang, L. Chen, X. Ying, S. Y. Philip, J. Wu, Z. Zheng. "Mobile application rating prediction via feature-oriented matrix factorization," in Proc. 2017 IEEE International Conference on Web Services (ICWS). IEEE, 2017: 261-268.
- [10] Lim, P. J. Bentley. "Investigating app store ranking algorithms using a simulation of mobile app ecosystems," in Proc. 2013 IEEE Congress on Evolutionary Computation. IEEE, 2013: 2672-2679.
- [11] Liu, S. Yu, X. Wei, Z. Ning. "An improved Apriori-based algorithm for friends recommendation in microblog," International Journal of Communication Systems, 2018, 31(2): e3453.
- [12] Mao, K. Lu, G. Li, and M. Yi, "Profiling users with tag networks in diffusion-based personalized recommendation," J. Inf. Sci., vol. 42, no. 5, pp. 711–722, 2016.
- [13] Miller, G. Rae, and F. Schaefer, "Modifications of Kleinberg's HITS algorithm using matrix exponentiation and Web log records," in Proc. 24th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2001, pp. 444–445.
- [14] Mitry et al., "The accuracy and reliability of crowdsourcing annotations of digital retinal images," Transl. Vis. Sci. Technol., vol. 5, p. 6, Sep. 2016.
- [15] Nagasinghe, "Computing principal eigenvectors of large Web graphs: Algorithms and accelerations related to pagerank and hits," Ph.D. dissertation, Southern Methodist Univ., Dallas, TX, USA, 2010.

- [16] Ntalianis, N. Tsapatsoulis, A. Doulamis, and N. Matsatsinis, "Automatic annotation of image databases based on implicit crowdsourcing, visual concept modeling and evolution," *Multimedia Tools Appl.*, vol. 69, no. 2, pp. 397–421, 2014.
- [17] Ntalianis, N. Tsapatsoulis, A. D. Doulamis, and N. E. Mastorakis, "Social relevance feedback based on multimedia content power," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 1, pp. 109–117, Mar. 2018.
- [18] Pirasteh, J. J. Jung, D. Hwang. "Item-Based Collaborative Filtering with Attribute Correlation," in *Proc. Asian conference on intelligent information and database systems*. Springer, Cham, 2014: 245-252.
- [19] Preece et al., "Sentinel: A codesigned platform for semantic enrichment of social media streams," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 1, pp. 118–131, Mar. 2018.
- [20] Sagduyu, A. Grushin, and Y. Shi, "Synthetic social media data generation," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 3, pp. 605–620, Sep. 2018.
- [21] Schall, B. Satzger, and H. Psailer, "Crowdsourcing tasks to social networks in BPEL4People," *World Wide Web*, vol. 17, no. 1, pp. 1–32, 2014.
- [22] Snoek and M. Worring, "Concept-based video retrieval," *Found. Trends Inf. Retr.*, vol. 2, no. 4, pp. 215–322, 2009.
- [23] Sunahase, Y. Baba, and H. Kashima, "Pairwise HITS: Quality estimation from pairwise comparisons in creator-evaluator crowdsourcing process," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 977–984.
- [24] Theodosiou and N. Tsapatsoulis, "Image retrieval using keywords: The machine learning perspective," in *Semantic Multimedia Analysis and Processing*. Boca Raton, FL, USA: Taylor & Francis, 2014, pp. 3–30.
- [25] Theodosiou and N. Tsapatsoulis, "Crowdsourcing annotation: Modelling keywords using low level features," in *Proc. 5th IEEE Int. Conf. Internet Multimedia Syst. Archit. Appl.*, Dec. 2011, pp. 1–4.
- [26] Wang, J. Yin. "Enhancing accuracy of User-based Collaborative Filtering recommendation algorithm in social network," in *Proc. 2012 3rd International Conference on System Science, Engineering Design and Manufacturing Informatization*. IEEE, 2012, 1: 142-145.
- [27] Wei, N. Ye, S. Zhang, X. Huang, J. Zhu. "Item-Based Collaborative Filtering Recommendation Algorithm Combining Item Category with Interestingness Measure," in *Proc. 2012 International Conference on Computer Science and Service System*. IEEE, 2012: 2038-2041.

- [28] Xu, K. Dutta, A. Datta, C. Ge. "Identifying functional aspects from user reviews for functionality - based mobile app recommendation," Journal of the Association for Information Science and Technology, 2018, 69(2): 242-255.
- [29] Xu, J. Yin, S. Deng, N. N. Xiong, J. Huang. "Context-aware QoS prediction for web service recommendation and selection," Expert Systems with Applications, 2016, 53: 75-86.
- [30] Yin, S. Aihua, G. Min, X. Yueshen, and W. Shuoping. "QoS prediction for web service recommendation with network location aware neighbor selection," International Journal of Software Engineering and Knowledge Engineering, 2016, 26(04): 611-632.
- [31] Yang, Y. Guo, Y. Liu, H. Steck. "A survey of collaborative filtering based social recommender systems," Computer Communications, 2014, 41: 1-10.
- [32] Zhou, B. Wang, J. Guo, J. Pan. "QoS-Aware Web Service Recommendation Using Collaborative Filtering with PGraph," in Proc. 2015 IEEE International Conference on Web Services. IEEE, 2015: 392-399.
- [33] Zheng, L. Zhang, X. Xie, W. Y. Ma. "Mining interesting locations and travel sequences from GPS trajectories," in Proc. Proceedings of the 18th international conference on World wide web. ACM, 2009:791-800.