Mining User's Data Based on Customer's Rating for Prediction and Recommendation—A Comparative Analysis

Soma Bandyopadhyay, S. S. Thakur and Jyotsna Kumar Mandal

Abstract The business of E-commerce is increasingly becoming popular due to pervasive Internet technologies. It is a human tendency to rely on the data or information, which they receive from their friends and neighbours prior to taking any decision, especially before purchasing any item. Presently, people are getting vast information and worldwide data though Web. Due to information overload, customers often face difficulties to locate their item of interest. Recommender system plays a significant role, and it helps to deal with information overload and further provides personalized recommendations to customers or users. In this paper, recommendation of smartphone was given based on feedback given by customer using weighted mean approach. The prediction was calculated for untried items, based on ratings given by new user using collaborative filtering. The results of recommendation and prediction show the approach is interesting.

1 Introduction

In the last decade, recommender systems become an integral part of e-commerce business to promote product sell and thus become a popular research field in the present era. To provide recommendation services, online stores use mostly either collaborative filtering (CF) or content-based (CB) recommendation approach. Two types of collaborative filtering are used in the recommendation system, one is user-

---

S. Bandyopadhyay (✉) · S. S. Thakur
MCKV Institute of Engineering, Howrah, West-Bengal, India
e-mail: somabanmuk@yahoo.co.in

S. S. Thakur
e-mail: subroto_thakur@yahoo.com

J. K. Mandal
University of Kalyani, Nadia, West-Bengal, India
e-mail: jkm.cse@gmail.com

based CF and another one is item-based CF. In the case of user-based collaborative filtering, the customers who have purchased and rated similar items with the target user are found. In the case of item-based collaborative filtering, instead of finding, similar customers focus is given on finding similar items. The algorithms attempt to find a similar type of items for each user which they have previously purchased, used or rating has been given by them [1]. The software tools and techniques that suggest the target user to choose items of their preference are known as recommender systems [2]. It can assist target user by providing a suggestion about items like products, services, information, which they would like or suits their needs [3]. Recommendation of a particular item may be provided by the following ways: items frequently purchased together, customers who purchased a particular item also purchased earlier, customers who searched a particular item also searched and sponsored product related to any particular item. According to Liang et al. [5], Web-based recommendation systems have gained popularity due to its ability of reducing overloaded information and increasing user satisfaction [4, 5]. A lot of information is needed to personalize recommender systems. User profile, age, gender, profession, educational qualification, hobby, preference, and previous purchasing pattern can be taken into consideration. The user is the most challenging to define as every user is having individual needs and goals. The loyalty of the customers should be taken into consideration. The older rating also should be considered in a good recommender system. Many recommender systems have inbuilt functions that give less weightage to older ratings but also make the system prone to possible loss of user's long-term interests that are not in frequent use [6]. If the recommended item is purchased by the user (e.g. products in an online store), then it can be considered as cost-effective but if after browsing the item, it is not purchased or the item is not found by the user it will not be considered as cost-effective.

2 Market Trends (2013–2018)

Since the advent of mobile phones, it has increasingly become an essential device for our everyday life. With the advancement of technology and in the era of miniaturization, smartphones become an essential component in our daily life. People can get multiple features like Internet, instant messenger, and e-mail in a single handheld device. They also can use this device in business or in pleasure. Table 1 shows the number of mobile phone user from 2013 to 2018 [7]. Figure 1 shows the increasing trend of the mobile phone user.

Table 2 shows the percentage of smartphone users in India from 2014 to 2018 among mobile phone users [7]. It is evident from Fig. 2 that the smartphone penetration rate is increasing in India in the recent years. The remaining paper is organized as follows. Section 3 presents the proposed work. In Sect. 4, prediction and recommendation computation have been discussed. Finally, results and discussion are presented in Sect. 5.

Table 1 Number of mobile phone user

Year	Number of mobile phone users in million
2013	524.9
2014	581.1
2015	638.4
2016	684.1
2017	730.7
2018	775.5

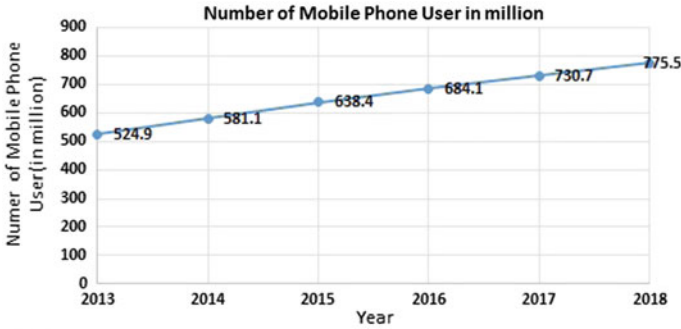**Fig. 1** Increasing trend of mobile phone user

Table 2 Smartphone user from 2014 to 2018

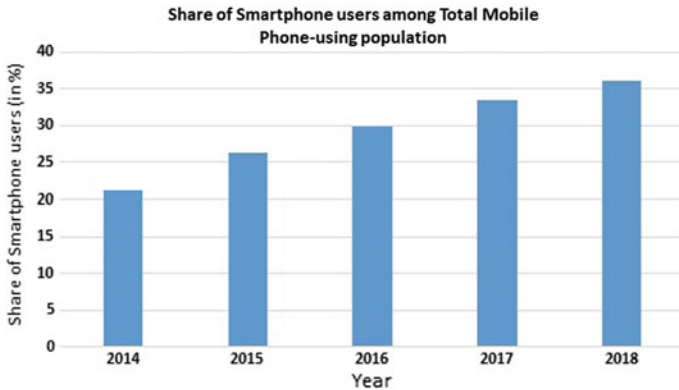| Year	Share of mobile phone user in %
2014	21.2
2015	26.3
2016	29.8
2017	33.4
2018	36

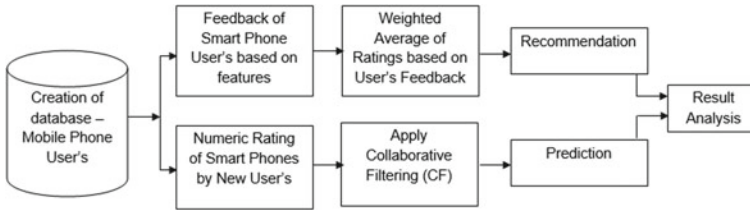**Fig. 2** Share of mobile phones users that use smartphone in India

Fig. 3 Block diagram of our proposed work

A survey was conducted in 2017 among the mobile phone user. It has been observed in India that 09% of users listen to music on their mobile phones and 40% of mobile users visited social networking sites [7]. The smartphone is also used to pay utility bills. One can click a picture by using the camera of the phone or can play game using some apps. It can also be used for navigation purpose. The smartphones are not only mere a means of communication, and it can also be used for entertainment. Various companies have specifically designed smartphones targeted towards various niches. The different features on which the smartphone companies have emphasised are camera, battery, speed, display, heating issues, etc. It has been observed that people of different sex or of different age groups have different taste in mobile usage and according to their requirement, they choose the product.

3 Proposed Work

In this work, the feedback was taken from customers related to smartphones they are using at present or used in the past. The feedback was given by them based on different features of smartphones. Similarly, new customers rating [8, 9] related to smartphone were taken, and finally, a comparison was done between two results. The approach, which is presented in this section, is based on the goal of targeted marketing.

3.1 Block Diagram of Our Proposed Work

The block diagram of our proposed work has been shown in Fig. 3.

3.2 Algorithm

Step 1: Start
   Step 2: Create initial database

Collect feedback from customers related to smartphones they are using at present or used in the past

Customer's feedback is based on ordinal rating

Step 3: Compute weighted average of ratings based on customer's feedback

Step 4: Consider this weighted average of ratings as user's recommendation

Step 5: Create another database

Collect feedback from new customers related to smartphones they may use in future

New customer's feedback is based on numeric rating

Step 6: Apply Collaborative filtering for prediction of items based on numeric ratings

Step 7: Do comparison between prediction and recommendation results

Step 8: Calculate accuracy and error estimation

Step 9: Repeat the process (i.e. Step 5 … Step 8) whenever a new customer gives the feedback, else go to Step 10

Step 10: Stop.

3.3 Memory-Based Collaborative Filtering Technique

For generating a prediction of an active user (new user), memory-based collaborative filtering algorithms use either entire database or a record of the user-item database. The people having preferences for similar kinds of products or having similar interests are treated as likeminded people. Here, neighbours of the active user (new user) are searched from the database to find the likeminded individuals and based on that prediction of preferences is produced for the new user. In case of neighbourhood-based collaborative filtering approach, first, the weight between two items is calculated. Here, $w_{i,j}$ denotes distance, correlation or weight between two users or items and i, j denote two users or items. After that, the weighted mean of all the ratings of the items are calculated for providing prediction to an active or new user searching for the similar kind of product or services [10, 11]. For generating a top-*N* recommendation, similarity computation (nearest neighbour) is done to find k most similar items or users. Then the neighbours are aggregated to find top-*N* most frequent items which will be used for recommendation.

Similarity Computation. In memory-based collaborative filtering algorithms, similarity computation between item or user is a tough task. For item-based CF algorithms, in order to find similarity, $w_{i,j}$ is calculated between two items i and j based on the users who have rated both of these items in a similar manner [10]. For user-based CF algorithm in order to find the similarity, $w_{u,v}$ is calculated between two users u and v having the same preferences of items or who have co-rated the same items. Different methodologies can be used for computing similarity or weight between items or users.

Correlation-Based Similarity. Here, similarity $w_{u,v}$ and $w_{i,j}$ are measured between two users u and v, or between two items i and j, respectively, by computing the Pearson correlation or other correlation-based similarities. To get accurate result first the co-rated cases are isolated. With the help of Pearson correlation measurement, the range to which two variables linearly relate with each other [12, 13] is computed. In case of user-based algorithm, the Pearson correlation is computed between users u and v as the following

$$w_{u,v} = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}} \tag{1}$$

where $i \in I$ summations are over the items for which both the users u and v have given rating. $\bar{r}_u$ denotes the average rating of the co-rated items of the uth user. Pearson correlation for item-based algorithm is calculated by Eq. 2.

$$w_{i,j} = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}} \tag{2}$$

where the set of users who rated both items i and j are denoted by $u \in U$. In this case, $r_{u,i}$ denotes the rating of user u on item i and $\bar{r}_i$ is the average rating of the ith item by the same user.

Vector Cosine-Based Similarity. To measure the similarity between two vectors, vector cosine-based similarity is used. In this case, each document can be treated as a vector of word frequencies and to find similarities between two vectors the cosine of the angle which is formed by the frequency vectors is computed [14]. In the case of collaborative filtering, a similar approach is adopted where users or items are used instead of documents and ratings are used instead of word frequencies. If we consider, R as the $m \times n$ user-item matrix, then to find similarity between two items, j, the cosine of the n-dimensional vectors corresponding to the ith and jth column of matrix R and I are to be determined. Equation 3 shows how the cosine similarity between items i and j is computed.

$$w_{i,j} = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\left\| \vec{i} \right\| * \left\| \vec{j} \right\|} \tag{3}$$

In this work, data samples from users were collected in both online and offline modes to calculate the rating of different smartphones. Usually, the rating can be taken from a different user in the following manner:

- Numerical rating: In this case, the rating of the product can be given in the range of 1–5.

- Ordinal rating: Qualitative rating can be given by writing some comments like awesome product, average product, poor product, or comments like strongly agree, neutral, disagree.
- Binary ratings: Here, the rating can be given by writing yes or no.

Accordingly, two types of rating data were collected from the user, one is the ordinal rating based on different features of mobile which they are presently using or have an experience of use and the other is the rating on the new item. In this case, the numerical rating was taken. In the said work, the user was mainly the students and staff members of engineering colleges. Different features of the smartphone have been studied and some of the features such as display, battery backup, and hanging issue have been taken into consideration to design the questionnaire.

A survey was done among 15 students using smartphones, before the main survey to ensure that respondents understand the questionnaire. A Google form containing the questions was sent by e-mail to the targeted user. The users were requested to participate in the survey and reply through e-mail only. Some of the incomplete responses were dropped and the rest were used in this work. A total of 6490 responses were used in this work, out of which 1640 responses were from new users. For 15 different types of smartphones, eight different features were taken into consideration which is in broad sense and further divided into subcategories for rating calculation. The names of the different mobile manufacturers have been used for the sake of research work only, and we are not biased about any particular company.

Table 3 shows the feedback of different smartphones based on their features, namely camera, battery backup, performance, and heating issues. In actual, there are 17 features, namely awesome camera, decent camera, poor camera, awesome battery backup, average battery backup, poor battery backup, awesome display, average display, poor display, awesome performance, poor performance, heating issue, and hanging issue. Due to the space constraints in the table, all the features cannot be shown but in actual while doing the calculations for recommendation all the features were taken into consideration. In this case, the user has given their remarks in different words like awesome camera, decent camera, poor camera, bad battery backup, value for money, hanging issues, heating issues, and poor performance These types of feedback were divided into 2–3 groups: good, average, and poor and according to the feedback the weighted sum was taken and rating was calculated.

Table 4 shows the user and smartphone-based rating. Actually, the rating has been taken from 1640 new users, but for the sake of simplicity, only the rating of user1, user2, and user1640 is shown in Table 4. In the first case, the rating was calculated based on the weighted average, and according to that, the recommendation was done and in the second cases, the rating was calculated based on collaborative filtering for prediction.

Table 3 Feedback of different user based on features of smartphone

Different smart-phones	Awesome camera	Decent camera	Awesome battery backup	Average battery backup	Awesome perfor-mance	Poor perfor-mance	Heating issues
Lenovo K8 Note	382	55	290	34	240	57	1240
Lenovo K8 Plus	92	14	89	6	45	39	213
Lenovo K6 Note	37	0	44	2	16	13	79
Lenovo A7000	185	27	259	14	134	46	765
Samsung Galaxy A6	68	7	20	3	14	0	6
Moto G6	67	34	44	17	16	36	0
Moto G5 S Plus	26	1	17	1	13	2	40
Realme 1	47	7	35	8	28	3	63
Honor 9 Lite	113	23	40	18	40	41	214
Nokia 6.1	10	10	10	12	27	0	18
Vivo Y71	12	0	4	0	4	0	11
Xiaomi Redmi5	13	0	8	0	6	0	14
Xiaomi Redmi Note 5	94	45	98	17	39	24	126
Honor C	23	68	24	43	2	13	19
Samsung Galaxy J7 Max	169	12	174	12	89	14	34

After that, the comparison was done based on the calculation. As the data were taken from the students and staff community, so it was possible to address the people of different age groups as well as different income groups. It is a very common practice of the customer that they want to know others opinion while purchasing the same product [15].

Table 4 New user rating of smartphone

Different smartphones	User1	User2	…	User1640
Lenovo K8 Note				3
Lenovo K8 Plus				
Lenovo K6 Note	3			
Lenovo A7000	1	2		2
Samsung Galaxy A6	4.5			
Moto G6				
Moto G5 S Plus	2			
Realme 1				
Honor 9 Lite		1		
Nokia 6.1				
Vivo Y71				
Xiaomi Redmi5		4		
Xiaomi Redmi Note 5	4	5		4
Honor C		1		
Samsung Galaxy J7 Max	5			5

4 Prediction and Recommendation Computation

Collaborating filtering system is one of the important processes for obtaining predictions or recommendations. In the neighbourhood-based collaborative filtering approach, a subset of nearest neighbours of the active user are chosen by finding their similarity with the active user. Then a weighted aggregate of their ratings is computed which is used for generating predictions for the concerned user [16, 17].

4.1 Weighted Sum of Others' Ratings

The formula for finding out the prediction for the active user, a, on a certain item, i is given below.

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U}(r_{u,i} - \bar{r}_u)w_{a,u}}{\sum_{u \in U}|w_{a,u}|} \qquad (4)$$

where $\bar{r}_a$ and $\bar{r}_u$ denote the average ratings for the user a and user u, respectively, on all other rated items, and $w_{a,u}$ indicates the weight between the user a and user u. The summations are over all the users $u \in U$ who have rated the item i.

4.2 Simply Weighted Average

Simple weighted average may be used for item-based prediction to predict the rating, $P_{u,i}$ for user u on item i [10].

$$P_{u,i} = \frac{\sum_{n \in N} r_{u,n} w_{i,n}}{\sum_{n \in N} |w_{i,n}|} \tag{5}$$

where the summations are over all other rated items $n \in N$ for user u. Here, $w_{i,n}$ is the weight between items i and n, $r_{u,n}$ denotes the rating for user u on item n.

5 Result and Discussion

The present work compares the rating system for a recommendation in two ways. While considering the rating of a particular item, attention should be given on the following things:

- Number of ratings: Number of ratings available is one of the most important data for a recommendation as this information is useful to numerous users who are searching for these items. If a user rates more items, many users can get suggestions from his or her rating.
- Degree of agreement with others: The degree of agreement is measured by comparing one user's opinion with others.
- Standard deviation of rating: It is the deviation of user's rating with his average rating. If the deviation is more, that will be given priority.
- Degree of similarity with top neighbours: The user who has similar weight can be replaced as he or she is very similar to other users while the user having greater weight will be considered as influential.
- Aggregated popularity of the rated items: The user has greater probability of having overlapped preferences when the sum of the popularities of the chosen rated items are high [13].

One can find desirable items if accurate recommendation is provided [18]. Table 5 shows the average rating (rating provided by a user on untried item) and average rating (computed based on feedback after use) of smartphones. Figure 4 shows acomparison between the rating calculations.

From the graph, it is clear that the predicted rating (rating for untried item) and calculated rating based on user feedback are closer to each other. The results of recommendation and prediction show the approach has relevance as the work is concerned and the manner the data has been dealt in for future recommendations.

Table 5 Comparison between recommendation and prediction

Smart phone	Recommendation of rating based on user feedback	Prediction of rating given by new user
Lenovo K8 Note	3.33	3.70
Lenovo K8 Plus	3.37	3.50
Lenovo K6 Note	3.58	3.50
Lenovo A7000	3.30	3.00
Samsung Galaxy J6	4.83	4.50
Moto G6	3.70	3.80
Moto G5 S Plus	4.09	4.00
Realme 1	4.11	3.00
Honor 9 Lite	3.25	3.00
Nokia 6.1	3.59	3.50
Vivo Y71	3.52	3.75
Xiaomi Redmi5	4.36	4.00
Xiaomi Redmi Note 5	4.00	4.00
Honor C	3.91	3.60
Samsung Galaxy J7 Max	4.49	4.50
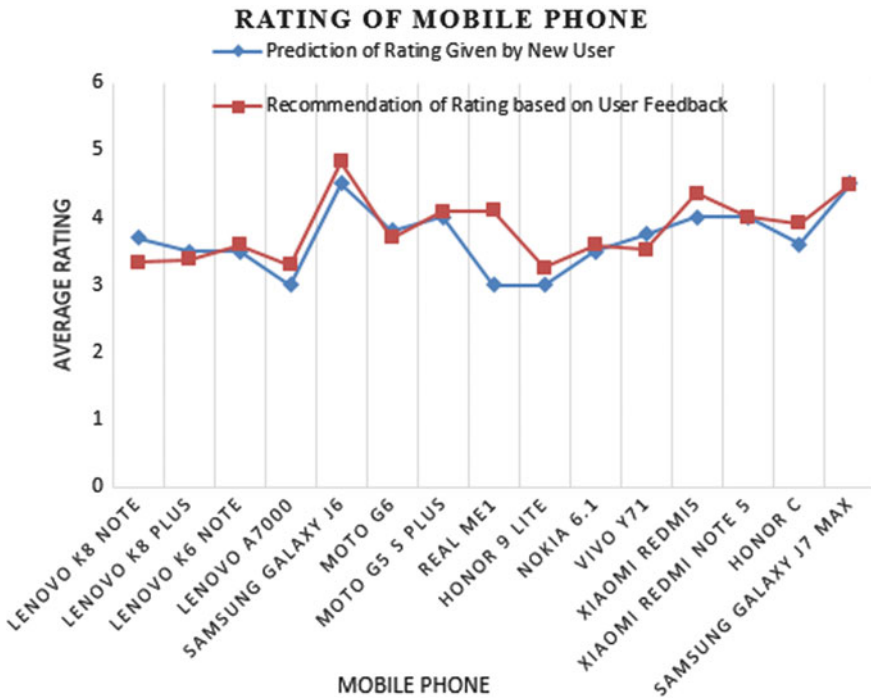
Fig. 4 Comparison between the rating calculations (recommendation vs. prediction)

References

1. Linden, G., Smith, B., York, J.: Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput. **7**(1), 76–80 (2003)
2. Rokach, L., Ricci, F., Shapira, B., & Kantor, P.B.: Recommender Systems Handbook. Springer, Berlin (2011)
3. Mahmood, T., & Ricci, F.: Improving recommender systems with adaptive conversational strategies. In: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia, pp. 73–82, Torino, Italy (2009)
4. Liang, T.P., Shaw, J.P., Wei, C.P.: A framework for managing web information: current research and future direction. In: Proceedings of the 32nd Hawaii International Conference on System Sciences (1999)
5. Liang, T.P., Hu, P.J.H., Kuo, Y.R., Chen, D.N.: A web-based recommendation system for mobile phone selection. In: 11th Pacific-Asia Conference on Information Systems, Auckland, New Zealand (2007)
6. Burke, R.: Hybrid web recommender systems. In: The Adaptive Web, pp. 377–408. Springer, Berlin (2007)
7. http://www.statista.com
8. Dhawan, S., Singh, K., Jyoti: High rating recent preferences based recommendation system. Procedia Comput. Sci. **70**, 259–264 (2015)
9. Dhawan, S., Singh, K., Kumar, N.: Rating based mechanism for effective ecommerce product recommendation in social networks. J. Netw. Commun. Emerg. Technol. (JNCET) **7**(9), 18–21 (2017)
10. Sarwar, B.M., Karypis, G., Konstan, J.A., Riedl, J.: Item based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web (WWW '01), pp. 285–295. Hongkong (2001)
11. Xie, F., Xu, M., Chen, Z.: RBRA: a simple and efficient rating-based recommender algorithm to cope with sparsity in recommender systems. In: 26th International Conference on Advanced Information Networking and Applications Workshops, pp. 306–311 (2012)
12. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: GroupLens: an open architecture for collaborative filtering of netnews. In: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, pp. 175–186. Chapel Hill, North Carolina, USA (1994)
13. Rashid, M., Karypis, G., Riedl, J.: Influence in ratings-based recommender systems: an algorithm-independent approach. In: The SIAM International Data Mining Conference, pp. 57–66. Newport Beach (2005)
14. Salton, G., McGill, M.: Introduction to Modern Information Retrieval, p. 124. McGraw-Hill, New York (1983)
15. Sohail, S.S., Siddiqui, J., Ali, R.: User feedback based evaluation of a product recommendation system using rank aggregation method. In: El-Alfy, E.S., Thampim S., Takagim H., Piramuthum S., Hanne, T. (eds.) Advances in Intelligent Informatics. Advances in Intelligent Systems and Computing, 320, pp. 349–358. Springer, Cham (2015)
16. Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J.: An algorithmic framework for performing collaborative filtering. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99), pp. 230–237. Berkeley, California, USA (1999)

17. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. Adv. Artif. Intell. **2009**, 1–19 (2009)
18. Yang, X., Guo, Y., Liu, Y., Steck, H.: A survey of collaborative filtering based social recommender systems. Comput. Commun. **41**, 1–10 (2014)