TAG BASED RECOMMENDATION SYSTEM OF A SOCIAL NETWORKING SITE*

Hatice BÜBER KAYA
Kırklareli University, Kayalı (Merkez) Yerlesshkesi, Kırklareli, Turkey

Ozan KAYA, MBA

ABSTRACT

Social networking sites (SNSs) are web-based services that allow people to create a profile and communicate with each other via their personalized web pages (profiles) .The history of SNSs begins in 1997 with the site named SixDegrees.com and from 2003 onwards too many SNSs have been launched and hence the term YASNS: "Yet Another Social Networking Service" came into our lives. In 2004 with the launch of MySpace, SNSs gained popularity remarkably and hit the mainstream .

Eventur was a tag-based event recommendation system about local events in Pittsburgh. The system used an event recommendation system to make recommendations more personalized. Recommendations are based on two of the core elements of social network - people and tags. Relationship information among people, tags, and items, is collected and aggregated across different sources within the enterprise. Tags applied on the user by other people are found to be highly effective in representing that user's topics of interest. This study was performed to understand the effectiveness of the Eventur system and this paper provides the methods, highlights the findings and suggests possible future research.

*This study was conducted as an independent study with Prof. Peter Brusilovsky, for the partial fullfilment of Master of Information Sciences degree at University of Pittsburgh, USA

Introduction

Eventur was a tag-based event recommendation system about local events in Pittsburgh. "Eventur completely relies on users to collect local cultural events and contribute to the system, serving as a content integration platform. Meanwhile, the system provides community features by giving the users online identity and provide social features, such as friending and recommending events to each other". Guy and Carmel define Social Recommender Systems (SRS) as systems to reduce the level of data exposure on social media users through its algorithm that funnels the most relevant information for users. Another major task of SRSs is to increase the total time spent, as well as the amount and frequency of activity on a social media site (Guy & Carmel, 2011). The main difference between the Social Recommender Systems and the classic Recommender Systems is, by gathering data from users' profiles and interaction between the users, social networking sites can generate a better system that can suggest more relevant recommendations (Al Falahi, Mavridis, & Atif, 2012).

| Content-based recommendation | Colaborative recommendation (including Social , demographic and group based rec.) | **HYPBRID Recommendation System** |

Content-based filtering methods use algorithms that aim to suggest items to the users based on information about the previous and current likes of the users (Brown, Robertson, & Sullivan, 2012). User profile is made up of total data collated through the user's online activities such as the items user has seen and liked/unliked and or rated in some form. During the recommendation process, the properties of unseen pages are compared with the information stored in the user profile and those resemble a higher match are recommended to the user (Mobasher, 2007). Nevertheless, we are faced with some systematic problems with the content-based filtering approaches. For example, user profiles are highly dependent on text that is made up of items which are previously visited and/or rated by the user. In this context the data misses the critical relationship between the items within the web pages (Fernández, Arias, Solla, Cabrer, & Nores, 2008). Moreover, item utilisation which is a common feature between objects in a particular context is not taken into account while formulating the user profile, which then create the base for recommendations of new items (Anand & Bamshad, 2003).

Collaborative-based (or in other words social-based) approach is an alternative to the previous approaches, targeting to overcome the restrictions of content-based approach. Rather than relying on the items seen and/or rated by the particular user, this approach collates the profiles of the users within an online community and recommends similar items to the users, based on the presumption that users share similar interests within a particular community (Amini, Ibrahim, & Othman, 2011). Collaborative filtering methods collect, process and analyse a huge chunk of information gathered about the users' online behaviour, focusing on their activities, preferences in order to predict what users will like by comparing and contrasting their behaviour vis a vis other users. Collaborative-based filtering was also augmented to include social information in order to enhance the detecting of user neighbourhood and bring a solution to data sparsity (Groh & Ehmig, 2007; Ma, Yang, Lyu, & Ki, 2008).

Known systems of recommendation like the Content-Based systems were upgraded with data and functions that include friendship and interaction data into the function formula, while designing the hybrid systems (Liu & Aberer, 2013). There are a few methods to make use of hybrid approaches; for example making content-based and collaborative based assumptions and joining those predictions (Adomavicius & Tuzhilin, 2005), secondly adding content-based capabilities and collaborative-based approach into each other (in both orders) (Hoekstra, 2010), or thirdly joining two approaches and making a new model (Boyd & Ellison, 2007).

The Eventur system had some substructure features of a content-based recommendation system at the time of study. These features were being used for narrowing down the search results by some attributes like:

-        Date based grouping (Upcoming events, grouped by daily and monthly events)

-        Type based grouping (Dance, Exhibition, Festival, Film/Video Art, Lecture, Literary Program, Music Concert, Opera/Musical, Other, Play, Sports)

-        Organization based grouping

-        Venue based grouping

-        Neighbourhood based grouping

Narrow your results by:

Organization

Advancing Black Arts in Pittsburgh (1)
Allegheny County Pennsylvania (2)
Amish Monkeys (1)
Beth Shalom Congregation (1)
Bloomfield Development Corporation (1)
Bricolage Production Company (4)
Broadway Across America: Pittsburgh (1)
Carnegie Library of Pittsburgh (20)
Carnegie Museum of Art (1)
Carnegie Science Center (1)
<< More

Venue

707 Penn Gallery (1)
709 Penn Gallery (1)
Arsenal Park (1)
August Wilson Center (1)
Benedum Center (8)
Bloomfield Little Italy (1)
Bricolage Theatre (4)
Byham Theater (1)
Carnegie Library of Pittsburgh - Main (20)
Carnegie Museum of Art (1)
<< More

Neighborhoods

Bloomfield (1)
Downtown (19)
Highland Park (5)
Oakland (35)
Squirrel Hill (9)
Other Areas (68)

Type

Dance (1)
Exhibition (9)
Festival (17)
Film/Video Art (1)
Lecture (13)
Literary Program (15)
Music Concert (38)
Opera/Musical (12)
Other (9)
Play (17)
<< More

Upcoming Events

f Share   twitter

Featured Events

- **Amish Monkeys Comedy** (Gemini Theater, 7501 Penn | Amish Monkeys | Opera/Musical)
- **Midnight Radio** (Bricolage Theatre | Bricolage Production Company | Opera/Musical)
- **Midnight Radio October** (Bricolage Theatre | Bricolage Production Company | Play)
- **Sunday Poetry and Reading Series: Carolyne Whelan** (Carnegie Library of Pittsburgh - Main | Carnegie Library of Pittsburgh | Other)
- **Sunday Poetry and Reading Series: Maureen McGranaghan** (Carnegie Library of Pittsburgh - Main | Carnegie Library of Pittsburgh | Music Concert)

Datasets

The system aimed to show different events. There were ten different types of events to be shown by the system. In the dataset that was used for this study,

there were a total of 130 events such as:

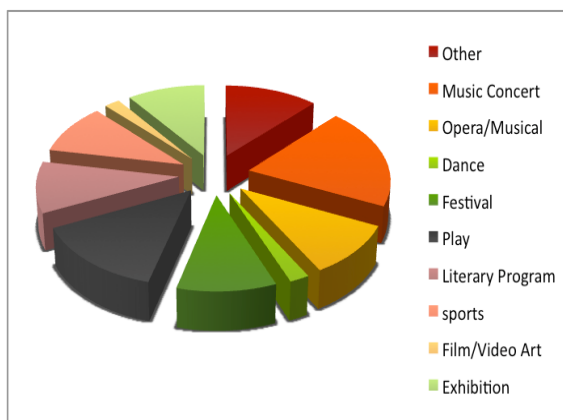|	
Other	11
Music Concert	28
Opera/Musical	16
Dance	1
Festival	17
Play	18
Literary Program	15
Sports	11
Film/Video Art	3
Exhibition	10
Total	130
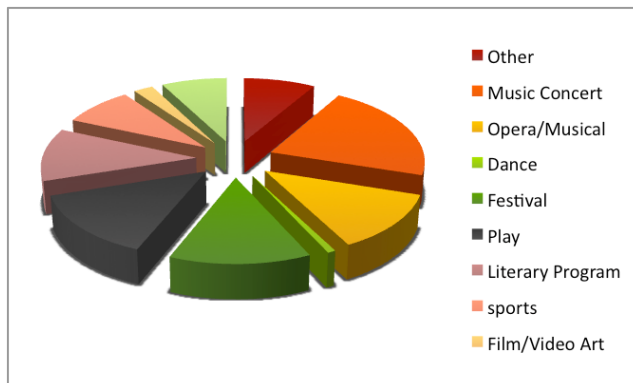
50 of them were chosen for the first phase of study. While choosing the first group of events, the number of events in each type calculated as weighted with the formula to select a diverse group:

1st Phase type n event count= round {[(total type n event count) * (1st Phase total event count)]/(total event count)}

Ex: 1st Phase "Music Concert" type event count= round {[28*50]/130} = round (10.7) = 1

	5
Music Concert	10
Opera/Musical	6
Dance	1
Festival	5
Play	7
Literary Program	5
Sports	5
Film/Video Art	1
Exhibition	5
	50

The remaining 80 of the total number of events were used in the second phase of the study.

Pre-Work

The tagging process was needed in order to run the algorithm. More than one tag submitted by more than one person was needed to have more accurate results. The problem faced at that point was where to find those people (must be objective) as soon as possible for 130 events.

Amazon Mechanical Turk (MTurk) is defined as "a crowdsourcing website for businesses (known as Requesters) to hire remotely located "crowdworkers" to perform discrete on-demand tasks that computers are currently unable to do" (mTurk).

Tagcow (by MTurk) provides descriptive tagging service to add metadata to images that helps the requester finds and organizes their image libraries more effectively (TagCow).

To use Tagcow,

-       Screenshots were taken and uploaded to an image hosting system for every single event.

-       A .CSV (A comma-separated values (CSV) file consists of any number of records, separated by line breaks of some kind; each record consists of fields, separated by some other character or string, most commonly a literal comma or tab) file created with the event url, event name and event screenshot for every single event.

-       A design layout was created for HITs.

-       After uploading the csv file, in a very short time; all 130 events were tagged by three people.

	Input.event_	Input.name	Input.image	Answer.Tag1	Answer.Tag2	Answer.Tag3	Answer.Tag4	Answer.Tag5
1	Input.event_	Input.name	Input.image	Answer.Tag1	Answer.Tag2	Answer.Tag3	Answer.Tag4	Answer.Tag5
2	http://eventu	The Addams F	http://img829	Addams	CLO	Gomez	Drama	Desk-winning
3	http://eventu	The Addams F	http://img829	series	poetry	reading	literary	sunday
4	http://eventu	The Addams F	http://img829	addams family	musical comec	nightmare		
5	http://eventu	Amish Monkey	http://img191	Amish	Pittsburgh	Monkeys	Gemini	Comedy
6	http://eventu	Amish Monkey	http://img191	amish monkey	gemini theater	comedy		
7	http://eventu	Amish Monkey	http://img191	Comady	Theater	Saturday	Pittsburgh	

Batch Results from MTurk

Algorithm

The recommendation program was coded in python (most_corresponded_tags.py - see the screenshot) and the algorithm is as follows:

1.Takes the file that was downloaded from MTurk that includes the HIT results (results_from_mt.csv - see the screenshot).

2. Cleans (takes the irrelevant information out because there were some information about the time of submission and like that) the file and removes the duplicate & empty tags for the same event then parses the first file in a new file (all_events_and_their_tags.csv - see the screenshot).

3.Then takes the event ids that the subject picked as they like from the 1st Phase dataset list.

4.Writes the set of liked events' unique tags into an array.

5.Compares events with picked events' tags and gives reputation.

6.Compares the events that are going to be recommended to subject with 1st Phase dataset list in order not to recommend the events that were already seen and rated by the subject.
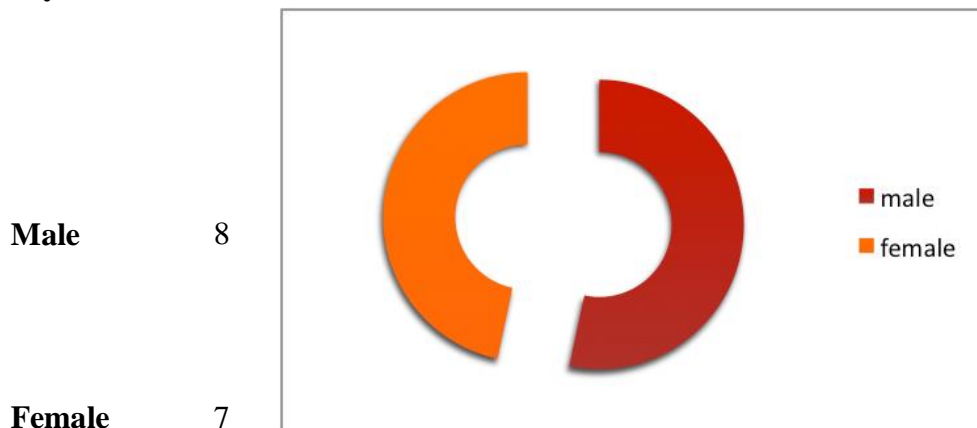
7.Sorts the events that are for recommendation by reputation.

8.Writes down the recommended event ids and reputations to a new file (reputations_of_events_which_are_the_most_corresponded_to_events_3099_44 10_4445.csv - see the screenshot).

Results_from_m t.Csv → Most_ Corresponded_ Tags.Py → All_events & tags.Csv → Reputations.Csv
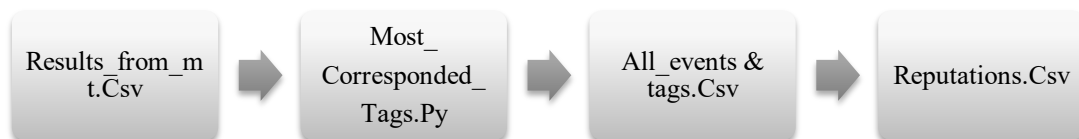
Demographics of the subjects

By Sex:

Male	8
Female	7

By Education Status:

Graduate	10
Undergraduate	5

Summary of 1st Phase User Study

Subjects were asked to tick yes if they like the event that was being showed. They were presented randomly in a mixed format to reduce the possibility of subjects' rapid/inaccurate answering due to boredom. The e-mail addresses were required from the subjects at the beginning to follow up with the study. The email submission page and an example form are shown in the pictures below. A page break was placed after every question in order not to make subjects confused.

Tag based event recommendation study

Please check out the events below and click "YES" if you are interested in that event and "NO" if you are not. If you have problem with reading you can click on the image to open event page in new window.

E-mail *   ex: myname@example.com

[Next]

Event ID 4534 *   ● YES
                      ● NO

World Kaleidoscope: Grecian Odyssey Dancers FREE

f Share   Tweet   8 +1

* This event is pending for the administrator's approval
Edit this event I Delete this event I
♥ Add to Favorites   I 📝 Create review I 👥 Share with Friends or Group I

- **Type:** Music Concert
- **Venue:** Carnegie Library of Pittsburgh - Main [map]
- **Organization:** Carnegie Library of Pittsburgh [website]
- **Source:** http://www.carnegielibrary.org/events/details.cfm?event_id=69115
- **Date/Time:**
  10-28-2012 Sunday 02:00 PM
- **Posted by:** joyhbuber

The Grecian Odyssey Dancers of the Presentation of Christ Greek Orthodox Church of Ypapanti in East Pittsburgh, PA, had their first performance at the Pittsburgh Folk Festival in 1981. Since then, they have represented the Greeks of Pittsburgh before more than 1000 audiences in Pennsylvania, Ohio, West Virginia, Michigan, Florida, New York and Greece, where they did three performance tours. In the Pittsburgh area, the Grecian Odyssey Dancers can annually be seen at Greek food festivals and international programs, as well as at academic institutions, private parties and cultural events. The Dancers range from 15 to 55 years of age and are prepared to perform 25 authentic regional dances of Greece. Their program takes you on a musical "odyssey" of Greece, demonstrating the differences of the people of the islands and the mainland. Their wardrobe boasts five sets of native costumes representing various areas of Greece. The Grecian Odyssey Dancers are honored to be perform at the Carnegie Library of Pittsburgh.

Event fee: Free
Location:
Quiet Reading Room - Main Library First Floor
4400 Forbes Ave
Pittsburgh, PA 15213

Contact:
Miguel
412-622-3151
newandfeatured@carnegielibrary.org

[Back]   [Next]

Picked events statics by gender

As it is shown below, most of the event types were picked with the same frequency by both males and females. However, while Literary Programs were more likely to be picked by male subjects, the Opera/Musical type events were liked by female subjects mostly.

Type	Male (8 people)	Female (7 people)
Other	18	24
Music Concert	35	27
Opera/Musical	9	20
Dance	5	5
Festival	18	22
Play	13	16
Literary Program	11	2
Sports	26	15
Film/Video Art	5	4
Exhibition	15	14
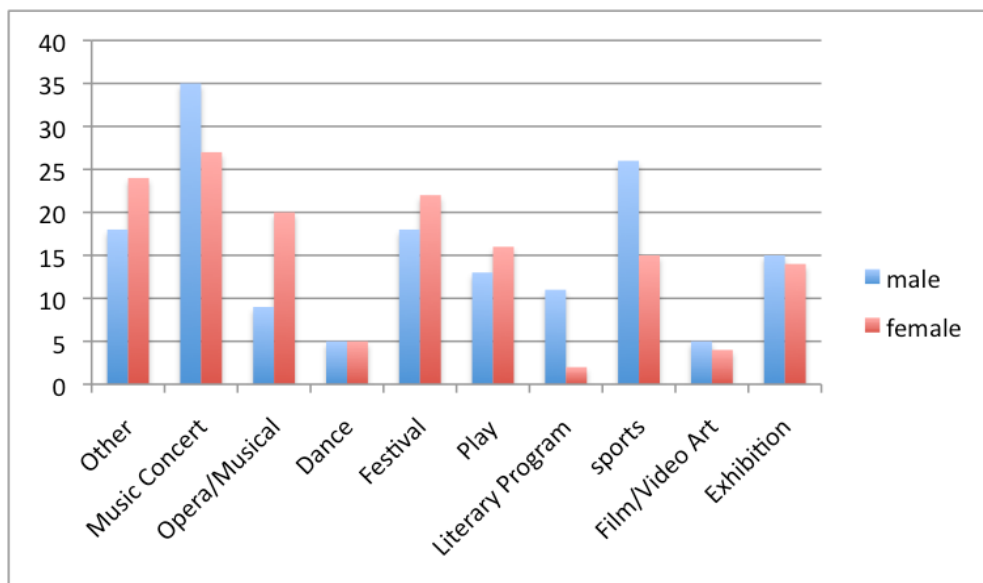
Summary of 2nd Phase User Study

After getting the results of the 1st Phase study, the code was run with the results for every subject separately. The top ten recommendations were sent to the subject as follow up study, and the subjects were asked to rate the recommendations in the same way that they picked the events that they liked.

Tag based event recommendation study

These are the events that we recommend you to see based on your previous likes..
Please check out the events below and click "YES" if you are interested in that event
and "NO" if you are not.If you have problem with reading you can click on the image
to open event page in new window.

Event ID 4536 *   ● YES
                          ● NO

**World Kaleidoscope: East End Appalachian Jam
All-Stars** FREE   f Share   ✔ Tweet g +1

* This event is pending for the administrator's approval
Edit this event l Delete this event l
♥ Add to Favorites l 📝 Create review l 👥 Share with Friends or Group l

- **Type:** Music Concert
- **Venue:** Carnegie Library of Pittsburgh - Main [map]
- **Organization:** Carnegie Library of Pittsburgh [website]
- **Source:** http://www.carnegielibrary.org/events/details.cfm?event_id=69117

Results

Unique recommendation event IDs and their occurrence
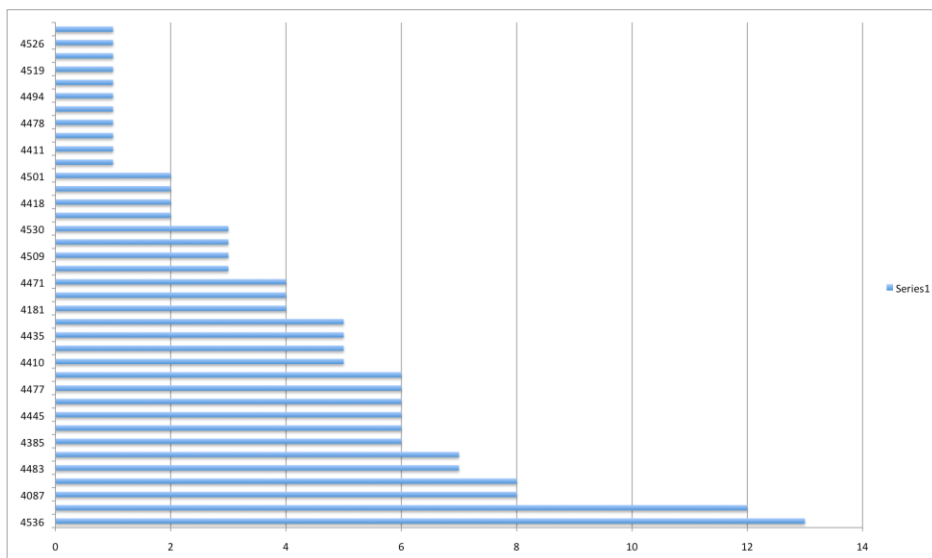
There were 38 out of 80 unique events that have been recommended to subjects.
The number of the recommended events (38) was lower than the un-
recommended events (42). Un-recommended events were not forwarded to the
subjects because they were not preferred enough based on the study done with

the 50 subjects.

Type	Times	Out of	%
Other	1	5	0.20
Music Concert	14	18	14.18
Opera/ Musical	5	11	0.45
Dance	0	0	0.00
Festival	7	12	0.58
Play	6	11	0.55
Literary Program	2	10	0.20
Sports	2	6	8.00
Film/ Video Art	0	2	0.00
Exhibition	1	5	0.20
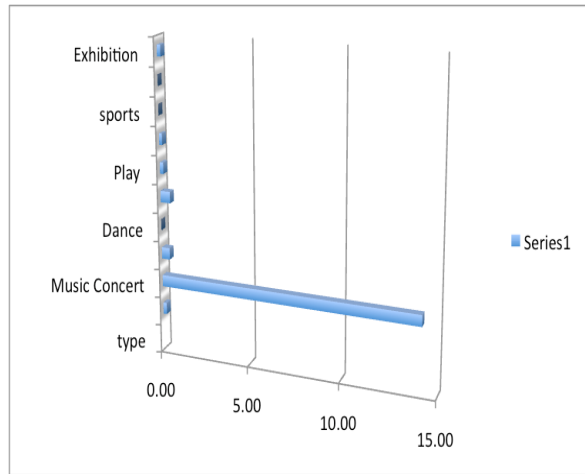
As the table above represents, the frequency of the reputations and the number of events to be recommended were proportional. The most frequent one, "Music Concert", was picked the most. Because the diversity of event types and the event counts were not even, it is wrong to make a conclusion about users' preferences.

Performance measures
Precision :

Precision is the fraction of the documents retrieved that are relevant to the user's information need. For example, average precision @ 10 can be calculated:

$$AP@10 = \frac{\sum_{r=1}^{10}(P(r) \times rel(r))}{number\ of\ relevant\ items\ @\ retrieval\ point}$$
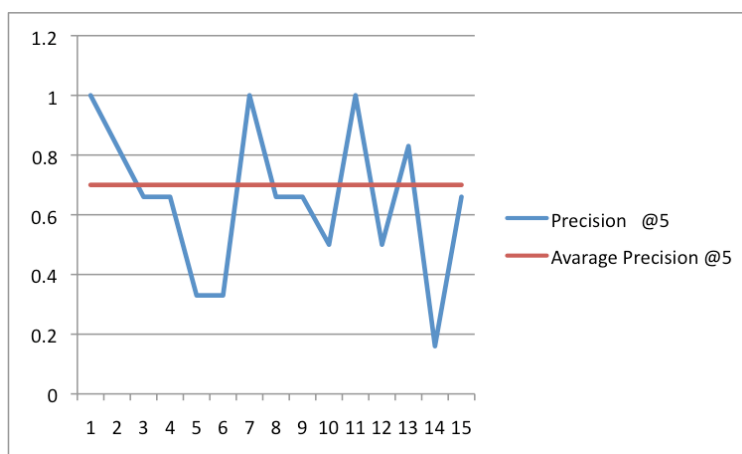
Mean average precision

Mean average precision for a set of queries is the mean of the average precision scores for each query.

$$\mathrm{MAP} = \frac{\sum_{q=1}^{Q} \mathrm{AveP}(q)}{Q}$$

where Q is the number of queries.

Precisions are calculated for every step of the follow up (2nd Phase of study).

Precision

Subjects	@1	@2	@3	@4	@5	@6	@ 7	@8	@9	@10
Sub 1	1	1	1	1	1	1	1	0.87	0.88	0.9
Sub 2	1	1	1	0.75	0.8	0.83	0.71	0.75	0.66	0.7
Sub 3	0	0.5	0.33	0.5	0.6	0.66	0.71	0.75	0.77	0.7
Sub 4	0	0	0.33	0.5	0.6	0.66	0.71	0.75	0.77	0.7
Sub 5	1	1	0.66	0.5	0.4	0.33	0.42	0.37	0.33	0.4
Sub 6	1	1	0.66	0.5	0.4	0.33	0.28	0.25	0.22	0.2
Sub 7	1	1	1	1	1	1	0.85	0.87	77	0.8
Sub 8	1	0.5	0.66	0.75	0.8	0.66	0.57	0.62	0.55	0.5
Sub 9	1	1	0.66	0.5	0.6	0.66	0.57	0.5	0.55	0.5
Sub 10	1	0.5	0.66	0.5	0.6	0.5	0.42	0.5	0.55	0.5
Sub 11	1	1	1	1	1	1	1	1	1	1
Sub 12	1	1	0.66	0.5	0.6	0.5	0.42	0.37	0.44	0.5
Sub 13	1	0.5	0.66	0.75	0.8	0.83	0.71	0.75	0.77	0.8
Sub 14	0	0	0	0	0	0.16	0.28	0.37	0.44	0.5
Sub 15	0	0.5	0.33	0.5	0.6	0.66	0.71	0.75	0.77	0.7
MAP	0.734	0.7	0.64	0.62	0.654	0.652	0.624	0.632	5.714	0.627
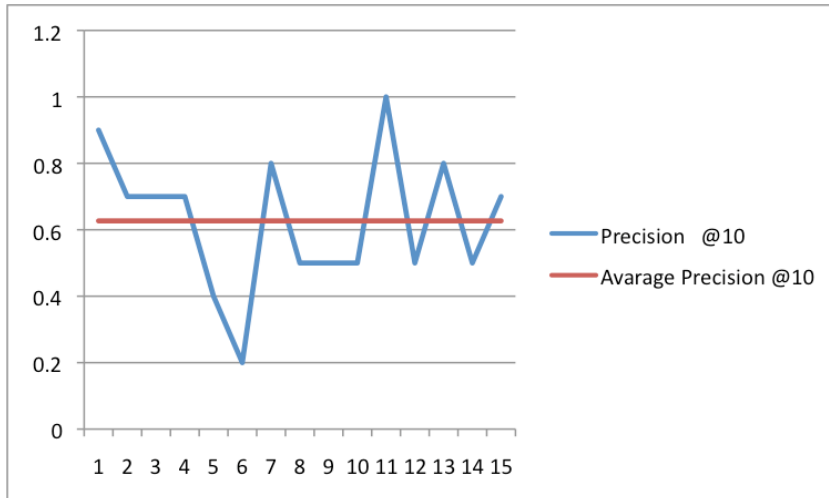
The highest mean average precisions (MAP) are at the first two recommendations, when we look at the 5th (in the middle) and the 10th (the last one) are around the same range with each other and also the average MAP, therefore we can suggest that the performance of this recommendation system is consistent.
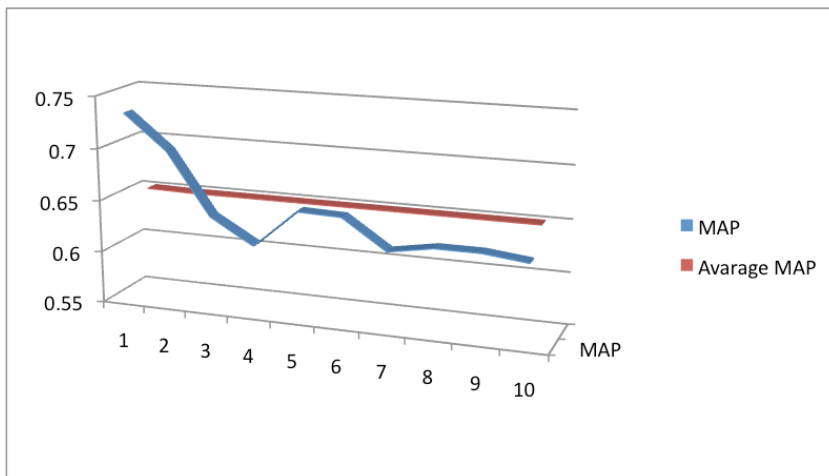
Limitations

The most important limitation was not having enough count of events in every type. Since Eventur.us was a local web site for Pittsburgh, there was not enough chance to find many events in every classification of events. Since there was time limitation for the study, we could not wait until more events to be added to underrepresented classification of events.

Due to the nature of this study, human subjects were needed to test the algorithm. However, since the subjects are asked to rate 50 events in the 1st phase of the study, every subject indicated that they got bored during selection; therefore although we even tried to encourage them for being patient and not to rush the rating process without attention, we can never be sure about it.

Conclusion

Although the recommendation algorithm was a basic algorithm, the results showed that Average MAP was above 0.5 and so we can suggest that the algorithm was successful. Since several limitations were encountered during the research; with a better-structured study, a better algorithm and higher number of events to use for survey, this research could be improved. Therefore, a future study can be designed with improved conditions for further research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Contact Information.
Email address: ozankaya@yahoo.com

References and notes:

Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering , 17* (6), 734–749.

Al Falahi, K., Mavridis, N., & Atif, Y. (2012). Social networks and recommender systems: A world of current and future synergies. In A. Abraham, & A.-E. Hassanien, *Computational Social Networks* (pp. 445–465). London: Springer.

Amini, B., Ibrahim, R., & Othman, M. (2011). Discovering the impact of knowledge in recommender systems: A comparative study. *International Journal of Computer Science and Engineering Survey (IJCSES) , 2* (3), 1-14.

Anand, S. S., & Bamshad, M. (2003). Intelligent techniques for web personalization. *IJCAI Workshop on Intelligent Techniques for Web Personalization* (pp. 1-36). Berlin, Heidelberg: Springer.

Boyd, D., & Ellison, N. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication , 13* (1), 210-230.

Brown, D., Robertson, T., & Sullivan, L. (2012). I SEE WHAT YOU DID THERE: THE USE OF SOCIAL NETWORKS TO GATHER EVIDENCE. *Southern Law Journa , 22* (2), 241 - 267.

Fernández, Y. B., Arias, J. J., Solla, A. G., Cabrer, M. R., & Nores, M. L. (2008). Semantic reasoning: a path to new possibilities of personalization. In *Proceedings of the 5th European Semantic Web Conference* (pp. 720-735). Berlin Heidelberg: Springer Berlin Heidelberg.

Groh, G., & Ehmig, C. (2007). Recommendations in taste related domains: Collaborative filtering vs. social filtering. *Proceedings of the 2007 International ACM Conference on Supporting Group Work, GROUP '07* (pp. 127–136). New York, NY, USA: ACM.

Guy, I., & Carmel, D. (2011). Social recommender systems. *Proceedings of the 20th International Conference Companion on World Wide Web, WWW '11* (pp. 283-284). New York, NY,USA: ACM.

Hoekstra, R. (2010). The Knowledge Reengineering Bottleneck. *Semantic Web – Interoperability, Usability, Applicability 1* .

Liu, X., & Aberer, K. (2013). Soco: A social network aided context-aware recommender system. *n Proceedings of the 22Nd International Conference on World Wide Web, WWW '13* (pp. 781–802). Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee.

Ma, H., Yang, H., Lyu, M. R., & Ki, I. (2008). Sorec: Social recommendation using probabilistic matrix factorization. *Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM '08* (pp. 931–940). New York, NY, USA: ACM.

Mobasher, B. (2007). Data Mining for Web Personalization. In P. Brusilovsky, & A. Kobsa, *The Adaptive Web: Methods and Strategies of Web Personalization, Lecture Notes in Computer Science, Vol. 4321.* New York, NY, USA: Springer-Verlag, Berlin Heidelberg.

mTurk. (n.d.). *Amazon Mechanical Turk FAQ Page*. Retrieved 2020, from https://www.mturk.com/worker/help#what_is

TagCow. (n.d.). *TagCow.* Retrieved 2020, from MTurk: https://requester.mturk.com/applications/app/TagCow