Review Article

Social Media Profile Recommender System using appropriate Similarity Metrics

Vigneshwar Reddy Likki

Department of Computer Science and Engineering, Bharat Institute of Engineering and Technology, Hyderabad, Telangana, India, 17e11a05a9@biet.ac.in

Abstract: This Review Article discusses about social media platforms whose fundamental idea is connecting the people and that connectivity is stronger if it is based on their common interests in different fields. if they like things commonly then there is a high probability that they are a like mind people and they can easily connect with each other. The social media helps to interact with people across the globe so, it is the social media's responsibility to give flawless and accurate profile recommendations based on user interests. This Review Article helps us to give better profile recommendations in social media sites like Facebook by applying some popular similarity calculating metrics to the attributes extracted from the individual profiles. different similarity metrics are used based on the type of attributes extracted. In this method we use specific attributes which the users do not feel insecure about. We do not use sensitive and personal information of the users, which the users are not much comfortable using those attributes to match them with other profiles such as their location, browser history, gender, religion, relationship status which they want to keep private and secure.

Keywords: Social Network, Similarity Metrics, Friends, User profile, Profile Matching

1. Introduction

Today most of the people uses Facebook in their Mobile Phones, laptops, tablets and computers. as of January 2021, the average time spent by day by American users on Facebook was 33 minutes.in other platforms like Instagram, Twitter, Snapchat the Average time spent may be more than 33 minutes. It is extremely important to engage the users with the content in the platform and give precise profile recommendations for the users. Psychologically there are many different factors which influences people consciously or sub-consciously to like or dislike anything. people often want to make friends who share common interests and has similar ideologies. This Review Article only focusses on the Categories section in Facebook which has Interest Attributes such as Sports, Movies, Music, TV shows, Books, Apps and games, hobbies. Professional attributes like education, occupation, industry and some not so personal attributes like their city, language and age. As there are vast number of attributes listed in the Facebook but most of them are very personal which the users are not much comfortable using those attributes to match their profiles with other profiles.

2. Related Work

Nagender Aneja, Sapna Gambhir in [1] presented Ad-hoc social network, where people of similar interests connect with each other using the ad-hoc communication mode of mobile devices. But in that approach, they were using the user's browser history and their GPS Location in order to match the profiles but due to the user's privacy concerns people

are not much comfortable using their browsing history to match them with other profiles and their real time GPS location due to many personal reasons. Therefore, in this Review Article I am using the professional information like (Education, Industry Occupation) and some other not so personal attributes like (language, Age, City) will be safe and effective in matching and recommending the profiles.

Rajni Ranjan Singh, Deepak Singh Tomar [2] presented an article on user profile Investigation in Orkut Social Network here, they have used some great methods to match profiles which are useful to investigate user's profiles and calculate interaction between users. but in that paper the attributes like Gender, religion, Drinking, Ethnicity Relationship Status, smoking should not be taken into consideration as two people who has similar interests can become friends although they are different genders similarly for the religion and ethnicity. the relationship status attribute should not be used as people can become friends with each other despite of their relationship status as this matching does not deal with their personal love interests as these attributes might decrease the efficiency of the matching algorithm. the attributes like drinking and smoking are completely their personal habits which my Review Article is not using to match profiles.

Sara Mazhari, Seyed Mostafa Fakhrahmad, Hoda Sadeghbeygi [3] presented an article on profiles recommendation solution in social networks where they have used the similarity metrics like levenshtein Distance and dice's Coefficient in text documents, Dice simply checks if a word from a vocabulary is present in both profiles, however not for the frequency of the occurrences. The Dice coefficient will therefore be the same between docs X = "Titanic" and Y = "Titanic Titanic Titanic" and any other reference document Z. They have used cosine similarity for the Age, Interest attributes which is good. I have used Cosine Similarity for the entire string type attributes which is more accurate. instead of doing a check if a word is present in both sets, cosine includes frequency by using the scalar product, it multiplies the frequencies of words in both profiles. Therefore, for a reference document Z, the cosine score will be different for X and Y. in their algorithm they have applied both the Levenshtein Distance and dice's Coefficient to the same set of attributes. However, from those two Similarity metrics (Levenshtein Distance, dice's Coefficient), they have compared each similarity value from both the metrics and then chose the similarity value which is greater than the other one. but this method gives us the similarity value which is greater among two metrics but not the accurate similarity value which we require to match profiles.

Vasavi Akhila Dabeeru [4] proposed an article using String Similarity Metrics for user's profile relationships where the article has the binary weight assign for mutual friends and mutual communities separately and has another new similarity score after calculating the cosine similarity which is very helpful for the accurate results.

3. Proposed Methodology

The Attributes of two different profiles can be extracted using the extracting tools then depending on the type of the Attributes calculate the similarity using similarity metrics like Cosine Similarity and Jaccard Index between the parent profile and the profiles which we have chosen to compare. After getting the similarity values compare those values with the Threshold value if the values are greater than or equal to the threshold then those profiles should be recommended to the parent profile. The profiles which are lesser than the threshold should not be taken into consideration.

3.1 Calculating the Similarity in between two profiles

The Similarity between the string attributes can be measured using the Cosine Similarity while for the numeric attributes we use the Jaccard Index to get the Similarity.

Similarity Metrics

- **Cosine Similarity**
- **Jaccard Index**

3.1.2 Cosine Similarity

Cosine similarity is a metric, helpful in determining, how similar the data objects are irrespective of their size. In cosine similarity, data objects in a dataset are treated as a vector. The formula to find the cosine similarity between two vectors is

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

3.1.3 Jaccard Index

The Jaccard distance measures the similarity of the two data set items as the intersection of those items divided by the union of the data items.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

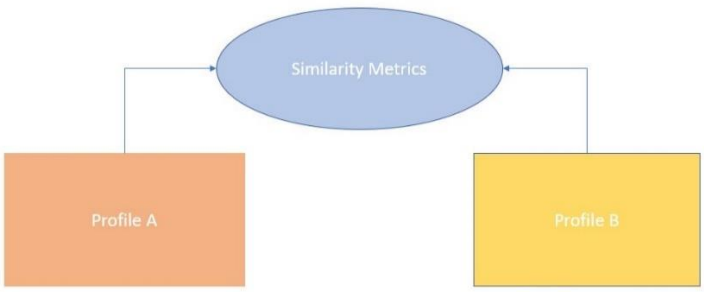

Figure 1. using of appropriate similarity metrics to match profiles

3.1.4 The Threshold Value

The Threshold value is a value which decides the similarity strength between the two profiles. The profiles can be said to be strongly similar or alike if and only if the Similarity value should be greater than or equal to the Threshold. If the similarity value is lesser than or equal to the Threshold then the two profiles might not likely to be strongly similar to each other.

Factors influencing the Threshold Value

The Threshold Value varies based on the number of similar profiles which has greater than or equal to the threshold, which we are matching with the parent profile and the different attributes on which we are depending to match those profiles.

4. Algorithms

Algorithm - 1

Begin

```
//Profiles which are selected to match with the parent profile
A = {a1, a2, a3, a4……an} //set of matching profiles
//foreach loop
Foreach i in A
Let P is a set of Professional attributes
P = {p1, p2, p3}   //set of professional attributes
Add Education on p1
Add Occupation on p2
Add Industry on p3
Let B is a set of Interest attributes
B = {b1, b2, b3, b4, b5, b6, b7}   //set of interest attributes
Add Music on b1
Add Movies on b2
Add TV shows on b3
Add Books on b4
Add Sports on b5
Add Apps and Games on b6
Add Hobbies on b7

Let S is a set of Personal attributes
S = {s1, s2, s3}   //set of personal attributes
Add Age on s1
Add City on s2
Add Language on s3
```

End

```
// apply cosine similarity for string
//apply Jaccard index for numeric type

P_set = {education, occupation, industry}
B_set = {movies, music, tv shows, sports, books, apps & games, hobbies}
S_set = {city, language, age}
```

Foreach i in A

If (P_set) is not numeric
Apply cosineSimilarity(P_set)
Else
Apply jaccardSimilarity(P_set)

If(B_set) is not numeric
Apply cosineSimilarity(B_set)
Else
Apply jaccardSimilarity(B_set)

If(S_set) is not numeric
Apply cosineSimilarity(S_set)
Else
Apply jaccardSimilarity(S_set)

End

After getting the cosine and Jaccard index similarity values we compare it with the existing Threshold and decide whether we need to match the profile with the parent profile or not.

Algorithm – 2 [decision-making algorithm]

A = {a1, a2, a3, a4……an} //set of matching profiles

Begin
Foreach i in A do
Q //Similarity Index
H //Threshold Value
If (Q >= H) then
Ans = Identical
Else
Ans = Non identical
End
Return Ans
End

After applying the decision-making algorithm Recommend the profiles whose similarity values are greater than or equal to threshold value else reject them.

5. Real-Time Example

Let us consider a real time example where we take a parent profile of Ram and he has some attributes listed and we have taken some professional, interest and personal attribute sets and we have taken five profiles and applying the Cosine Similarity and Jaccard index where ever needed to each profile then we get the following outcomes.

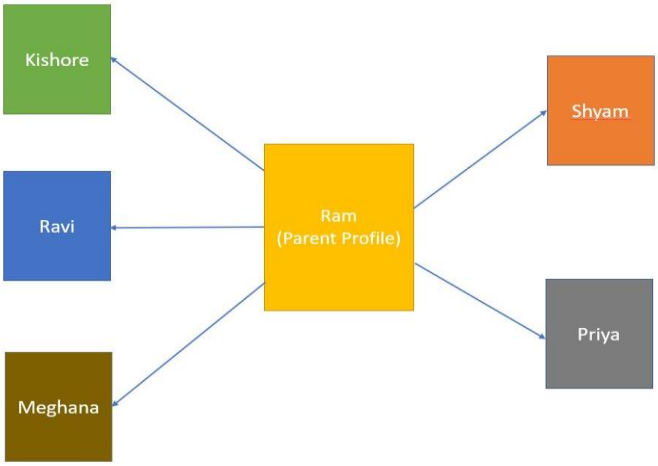

Figure 2. profiles selected to match with parent profile

Calculate the similarity values of individual profiles for each attribute using Cosine similarity or Jaccard Index formula based on the type of attribute and fill the tables below.

Table 1. Table consists of interest Attributes similarity values

Profiles	**Interest Attributes (Cosine Similarity)**						
	Movies	**Music**	**TV shows**	**Sports**	**Books**	**Apps**	**Hobbies**
Kishore	0.657	0.889	0.345	0.765	0.558	0.963	0.864
Ravi	0.867	0.823	0.768	0.813	0.800	0.912	0.789
Meghana	0.671	0.918	0.621	0.738	0.799	0.746	0.812
Shyam	0.899	0.876	0.834	0.822	0.876	0.839	0.889
Priya	0.789	0.765	0.812	0.834	0.918	0.789	0.756

Table 2. Table consists of Professional Attributes similarity values

Profiles	**Professional Attributes (Cosine Similarity)**		
	Education	**Occupation**	**Industry**
Kishore	0.558	0.963	0.864
Ravi	0.867	0.823	0.892
Meghana	0.671	0.799	0.812
Shyam	0.899	0.822	0.878
Priya	0.834	0.789	0.756

Table 3. Table consists of Personal Attributes similarity values

Profiles	**Personal Attributes**		
	City	**Language**	**Age (Jaccard index)**
Kishore	1.0	1.0	0.864
Ravi	0.0	1.0	0.839
Meghana	0.0	0.0	0.712
Shyam	1.0	0.0	0.839
Priya	1.0	1.0	0.756

Average of interest attributes of profiles:

$$\frac{\text{Set of interest attributes}}{\text{total no. of attributes in interest set}}$$

Average of professional attributes of profiles:

$$\frac{\text{Set of professional attributes}}{\text{total no. of attributes in professional set}}$$

Average of personal attributes of profiles:

$$\frac{\text{Set of personal attributes}}{\text{total no. of attributes in personal set}}$$

Example using Kishore's Profile:

Average of Interest Attributes:

$$\frac{0.657+0.889+0.345+0.765+0.558+0.963+0.864}{7} = 0.720$$

Average of Professional Attributes:

$$\frac{0.558+0.963+0.864}{3} = 0.795$$

Average of Personal Attributes:

$$\frac{1.0+1.0+0.864}{3} = 0.954$$

Average of all the Attributes:

$$\frac{0.720+0.795+0.954}{3} = 0.823$$

Similarity value of Kishore with Ram = 0.823

Similarly, we get the Similarity values of Ravi, Meghana, Shyam and Priya

After getting the overall average similarity values we then calculate the average values of all the [interest, professional, personal attributes] then we get the below values.

Table 4. Cosine Similarity and Jaccard Index of all the child profiles

Profiles	**Cosine Similarity and Jaccard Index**
Kishore	0.823
Ravi	0.768
Meghana	0.584
Shyam	0.780
Priya	0.918

Comparing with the Threshold Value

Let us assume the Threshold Value as H = 0.80 then compare the similarity index values of the profiles with the Threshold Value, if the Values are greater than or equal to the Threshold then recommend that profile to the Parent Profile else do not recommend the Profile.

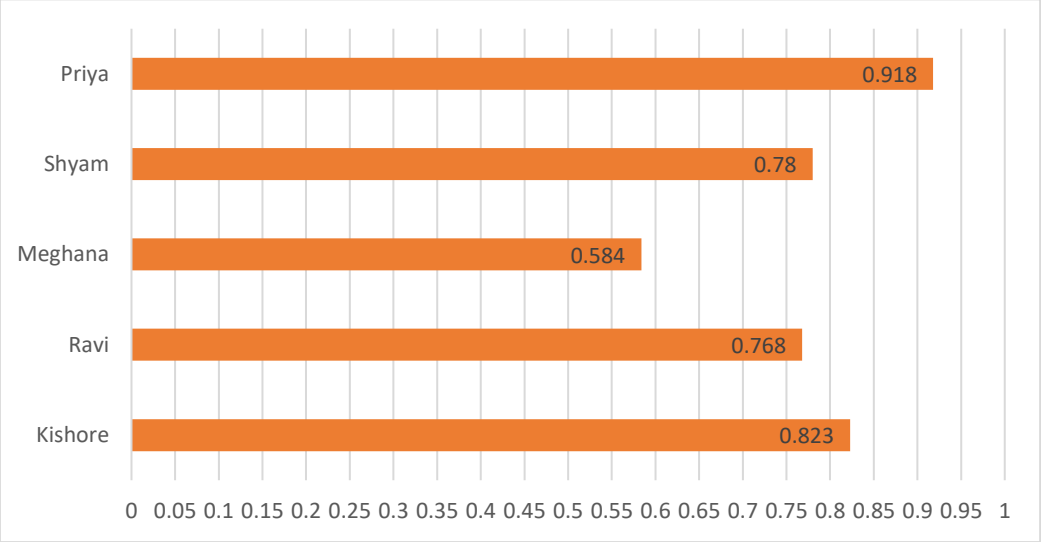

Figure 3. Similarity Chart of all the profiles

As you can observe from the above picture (Figure. 3) we can see that the profiles of Priya and Kishore are greater than the Threshold so we recommend those profiles to the Parent Profile named Ram.

6. Conclusion

This Article helps us to give better profile recommendations in social media sites like Facebook. we can match the user profiles by applying some popular similarity calculating metrics to the attributes extracted from those profiles. after getting similarity values we compare those values with the Threshold value if the values are greater than or equal to the threshold then those profiles should be recommended to the parent profile. The profiles which are lesser than the threshold should are not recommended. As future work, as social media is dynamic in nature, and this is a work on keywords matching there is a high probability of syntactical errors. so, in order to improve the accuracy of the model we have to apply this method on a large set of real time data so that we can improve the efficiency of the profile recommendations in social media applications.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Department of Computer Science and Engineering, Bharat Institute of Engineering and Technology, Hyderabad, Telangana, India.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Nagender Aneja, Sapna Gambhir, "Geo-social profile Matching Algorithm for Dynamic Interests in Ad-Hoc Social Network", Department of Computer Engineering, YMCA University of Science and Technology, Faridabad, India, 2014.

2. Rajni Ranjan Singh, Deepak Singh Tomar, "Approaches for user profile Investigation in Orkut Social Network", Maulana Azad National Institute of Technology (MANIT), Bhopal, India, 2009.

3. Sara Mazhari, Seyed Mostafa Fakhrahmad, Hoda Sadeghbeygi, "A user-profile-based friendship recommendation solution in social networks", Shiraz University, Iran, 2015.

4. Vasavi Akhila Dabeeru, "User Profile Relationships using String Similarity Metrics in Social Networks", 2014.

5. Elie Raad, Richard Chbeir, Albert Dipanda, "User Profile Matching in Social Networks", published in Network Based Information Systems, Japan 2010, pp.297-304. ffhal-00643509f.

6. Lada A. Adamic , Eytan Adar, "Friends and neighbours on the Web", HP Labs, 1501 Page Mill Road, Palo Alto, CA 94304, USA.

7. Dinithi Pallegedara, Lei Pan, "Investigating Facebook Groups through a Random Graph Model".

8. Robert Patton, "Facebook and Networked Interactivity", December 2007.

9. Facebook Network analysis using Gephi, www.gephi.com.

10. Naohiro Matsumura, David E. Goldberg, Xavier Lllora, "Mining Directed Social Network", Message Board.

11. Patrick Doreian, Tom A.B. Snijders, "Social Networks, an international journal of structural analysis".

12. William H. Hsu Joseph Lancaster Martin S.R. Paradesi Tim Weninger, "Structural Link Analysis from User Profiles and Friends Networks: A Feature Construction Approach", Department of Computing and Information Sciences, Kansas State University.

13. Minas Gjoka, Maciej Kurant, Carter T. Butts, Athina Markopoulou, "A walk in Facebook: Uniform Sampling of Users in Online Social Networks".

14. Alexander Strehl, Joydeep Ghosh, and Raymond Mooney, "Impact of similarity measures on Web-page Clustering", The University of Texas, Austin, Texas.

15. S. Kak, Feedback neural networks: new characteristics and a generalization. Circuits, Systems, and Signal Processing, vol. 12, pp. 263-278, 1993.

16. Graph Visualization of an Economic Environment using Gephi.

17. S. Kak, Self-indexing of neural memories, Physics Letters A, vol. 143, pp. 293-296, 1990.

18. D.L. Prados and S. Kak, Neural network capacity using the delta rule. Electronics Letters, vol. 25, pp. 197-199, 1989.

19. D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks.," Nature, vol. 393, no. 6684, pp. 440–2, Jun. 1998.

20. S. Kak and M.C. Stinson, A bicameral neural network where information can be indexed. Electronics Letters, vol. 25, pp. 203-205, 1989.

21. J. Kleinberg, "The small-world phenomenon: an algorithmic perspective," in Proceedings of the thirty-second annual ACM symposium on Theory of computing - STOC '00, 2000, pp. 163–170.

22. D. J. Watts, "The 'New' Science of Networks," Annu. Rev. Sociol., vol. 30, no. 1, pp. 243–270, Aug. 2004.

23. Facebook Data Team, "Anatomy of Facebook," 2012. [Online]. Available: https://www.facebook.com/notes/facebook-data-team/anatomy offacebook/10150388519243859. [Accessed: 06-Jan-2012].

24. H. Ebel, L.-I. Mielsch, and S. Bornholdt, "Scale-free topology of e-mail networks," Phys. Rev. E, vol. 66, no. 3, Sep. 2002.

25. S. Kak, On generalization by neural networks. Information Sciences, vol. 111, pp. 293-302, 1998.

26. S. Kak, New algorithms for training feedforward neural networks. Pattern Recognition Letters, vol. 15, pp. 295-298, 1994.

27. S. Kak, On training feedforward neural networks. Pramana, vol. 40, pp. 35-42, 1993.

28. K.-W. Tang and S. Kak, A new corner classification approach to neural network training. Circuits, Systems, and Signal Processing, vol. 17, pp. 459-469, 1998.

29. S. Kak, Faster web search and prediction using instantaneously trained neural networks. IEEE Intelligent Systems, vol. 14, pp. 79-82, November/December 1999.

30. K.W. Tang and S. Kak, Fast classification networks for signal processing. Circuits, Systems, Signal Processing, vol. 21, pp. 207-224, 2002.

31. S. Asur and B. A. Huberman, "Predicting the future with social media," in 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2010, pp. 492–499.

32. S. Kak, A class of instantaneously trained neural networks. Information Sciences, vol.148, pp. 97-102, 2002.

33. A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting elections with Twitter : what 140 characters reveal about political sentiment," in Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, 2010, pp. 178–185.

34. Michael H Goldhaber, "The Attention Economy and the Net," First Monday, vol. 2, no. 4, pp. 1–27, 1997.

35. N. Pope, "The Economics of Attention: Style and Substance in the Age of Information (review)," Technol. Cult., vol. 48, no. 3, pp. 673–675, 2007.

36. S. Yu and S. Kak, "A Survey of Prediction Using Social Media," 07-Mar-2012. [Online]. Available: http://arxiv.org/abs/1203.1647. [Accessed: 11-Mar-2013].

37. S. Yu and S. Kak, "Social Network Dynamics: An Attention Economics Perspective," in Social Networks: A Framework of Computational Intelligence, Edited by Witold Pedrycz and Shyi-Ming Chen. Springer, 2014.

38. Overview on Cosine Similarity and Jaccard Index, www.geeksforgeeks.org

39. H. A. Simon, "Designing organizations for an information rich world," in Computers, communications, and the public interest, 1971.

40. S. Yu and S. Kak, An empirical study on how users adopt famous entities. International Conference on Future Generation Communication Technologies, London, December 2012.