International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 9 (2017) pp. 241-248

Received: 17 April 2017; Accept: 21 August 2017; Publish 7 September 2017

Design of a Hybrid Integrator for Autonomous Social Networks

Charu Virmani[1], Dimple Juneja[2]and Anuradha Pillai[3]

[1]Computer Science and Engineering
Manav Rachna International University,
Faridabad, 121001, INDIA
E-mail:charu.fet@mriu.edu.in

[2]Department of Computer Applications,
National Institute of Technology
KURUKSHETRA, 136119, INDIA
E-mail:dimplejunejagupta@gmail.com

[3] Department of Computer Engineering,
YMCA University of Science and Technology
Faridabad, 121006, INDIA
E-mail:anuangra@yahoo.com

***Abstract*: Although the popular social networks have enhanced the interaction amongst people registered on their respective sites, however, the existing interaction rules do not allow inter-site sharing of user profiles and their activities. In fact, a user creates public profile with the intention to share activities globally through social networks. Now, since the identity of a user registered on numerous social networking sites are not integrated globally, therefore the different profiles of the user himself usually remain in isolation. It is an obvious fact that a genuine user registers with his or her unique attributes to create an identity and shall be identifiable with at least some of the common attributes across all popular social networks. The current work finds motivation from the above requirements and thus uniquely contributes a profile integrator which is able to generate a single unique profile from multiple profiles (of a user) available across different social networks. The integrator outstandingly disambiguates user profiles existing across different social networks using public attributes with decision to map the profiles using change in location of the user. The proposed model will increase the discoverability of the user, deriving new communities and promotional activities among multiple domains.**

***Keywords*:** Social Networks, Profile Integrator, Identity Resolver, Levenshtein Distance, Phonetic encoding

I. Introduction

Last decade has seen Social Network Aggregators (SNA) which aggregates the social information of users across many social networks like Hootsuite, Sco-connect, flock [1]. Owing to differences in the privacy policies (which in fact keep on evolving also) of all social networks, the existing SNAs fall short in various aspects such as resolving the identity of user i.e. ensuring that only the legitimate user profile is being integrated. Users need to register and authenticate themselves on each social network on the aggregator by providing their user-id and password to be syndicated. This paper illustrates a mechanism that connects and aggregates the user profile from various social networks.

In fact, while using social network services, the user creates his/her profile by adding, for example, his/her name, pictures, and friends; hence, diverse profiles are distributed over the network. These profiles include valuable information about the user for advertising, customer centric tasks, and a user's background check. The global information about the aggregated user profiles shall make the user understand the privacy and security issues of his open information [2].

The linking of the user's profile remains an abstract procedure for a naive user. For instance, PleaseRobMe.com integrated information from tweets and FourSquare to discover that the user was not at home [28]. The growing need of matching user profiles and linking his/her identity will not only keep the user informed, but also is the basis of new advancements in mining information about the user for personalized tasks.

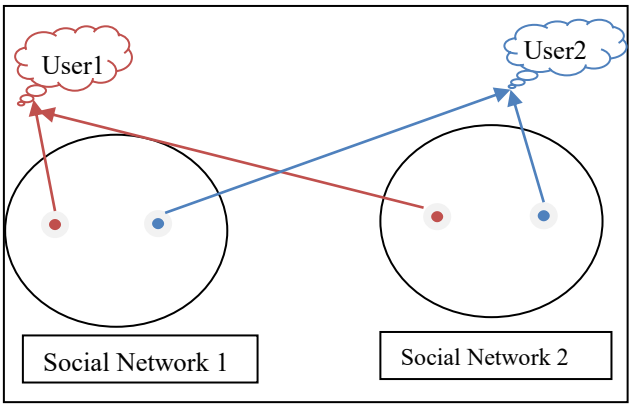

Figure 1. Desired Integration

On different social networks, a user puts variety of his personal attributes; therefore, the challenge is to map a set of these attributes with high precision and accuracy. The above discussion brings up the fact that resolving an identity is a major challenge. The current work thus uniquely contributes an *Identity Resolver* which maps user's profiles across various social networks which in turn are correlated and integrated into a single profile by *Profile Integrator* as shown in figure 1.

This paper is structured into five sections: Section 2 throws light on the work of eminent researchers highlighting their substantial contributions. The discussion in section 2 indicates that identity resolution is major hurdle towards the success of any SNA. Section 3 discusses the major challenges. The current work thus finds motivation and resolves the challenge listed above. Section 4 uniquely contributes an Identity Resolver Module (IRM) and a Profile Integrator Module (PIM). In contrast to PIM which maps the unique identity of a user profile distributed across various social sites by correlating various public attributes, IRM performs user mapping based on strengthening certain attributes such as name and location. This has been established with a data set in the evaluation section given in section 5. Section 6 finally concludes.

II. Related Work

In order to identify a user uniquely, initially a set of publicly available attributes which can find the similar account across multiple social networking sites with the claimed precision, accuracy, and recall were explored. Some of the important attributes common to most social networking sites [2][15] are user name, display name, profile image, description, location, age, sex, group of interest and connections. Although, the major attributes that distinguish a user across multiple social networks are publicly available information fields however, users may provide different information on different social networks for the same attribute. For example, same user may use the name John on Facebook and Jon on Twitter. Thus, different information about the users' same attribute from multiple social networks requires learning about the mapping of these attributes to know more about the users [8] [9]. In order to create users' aggregated profiles, Pontual et al. [10] designed a crawler that collects information from different social networks and sites using a real name. However, the same lacked in correlating the accounts that can achieve maximum accuracy.

Further, to digitize the user's identity, various attributes (public and private) [8] are being exploited to match users across social networks. The unification of accounts using graph-based techniques is discussed in [2] [3] [4]. The graph-based technique, such as Friend of Friend (FOAF) [5], links multiple user accounts based on identifiers such as email ID, Instant Messenger ID. The graph thus generated across different social networks is compared and a score is assigned to it. If the threshold score is the same in every network, the identity is considered to be of the same user. However, this technique is not scalable and relies on private information, thereby raising question on privacy policy of different social networks [6] [7].

Another popular approach connecting various user accounts is the tagging [11]. The average established accuracy for the above method is around 64.5% [11]. Correlating user id and user names is another popular option to establish a single identity. However, the accuracy rate is just 66% [3]. The user identification algorithm [13] computing the weighted score of various attributes of user profile is one of the most successful approaches in the domain. Singh et al. [14] also suggested extracting the user's birthday value resulting in exact identification when it is cross-matched with the users' name.

Malhotra et al. [9] exploits user name, name, description, location, image, connections for mapping user profiles listed on Twitter and LinkedIn. Zhang et al. [27] presented holistic supervised learning using the user's public and private information from the social network for resolving the identity. However, with the advent of Web 2.0, a user can prevent the visualization of his connections and other features which is used to identify and disambiguate users [7]. Identity search by Jain et al. [30] performs matching on the basis of the user's profile, content, network, and self-mentioning mechanism to map the user over Twitter and Facebook. Zhou et al. [20] presented a novel technique for user identification using the friendship structure and topology of the social networks which makes it an expensive approach for scant online social networks. Liu et al. [16] proposed a semi-supervised embedding algorithm which uses the capability of the network to learn the follower/followee of each user. Despite the accuracy of above discussed algorithms, the researchers have not considered the timestamp for resolving the identity and have applied the techniques concerning the network factors. Many researchers considers profile attributes in the criteria to match the identity using syntactic [17, 18, 19, 21], semantic [22, 23, 24, 25], and graph-matching techniques [2, 3, 4, 26].

Vu et al. [28] considered Facebook and Twitter as the underlying social networks for aggregation of profiles using FOAF ontology. The model does not trace the source and the time of information. Moreover, whenever there is a conflict in the information, the user is the deciding factor to choose among pieces of information and the others are deleted.

The dwelling of literature clearly indicates the fact that user profile disambiguation is achieved by using a large set of public and private attributes and in general, the three-step matching scheme [8] exists for mapping the users who deliberately create isolated profiles on different social networks.

In brief, the social network allows user to opt out of the public display of the friends list and other several attributes which is mostly used in the above techniques [7]. Since connections and the friends list information can be restricted by a user, they cannot be used as matching criteria. The exact matching of a user's profile may not be possible as users tend to isolate their identity across social networks. A limited attribute set can be explored for matching the user profile with integration of the location attribute of the newsfeed/tweets that is generated by the social network or geo-tags that are generated by device when user updates through the status/tweets. Instead of using multiple criteria to match and search, this work finds motivation for a stepwise approach to resolve the ambiguity of user profile providing better and less expensive results. The work proposes a Hybrid Integrator for Autonomous Social Integrator (HIASN) which is an amalgamation of Phonetic Encoding Score and the Levenshtein Algorithm.

While the former algorithm takes a keyword as input (person's name, location name etc) and produces a character string that identifies a set of words that are (roughly) phonetically similar, the later is being used to match user name spellings or pronunciations. The Levenshtein Algorithm is based on computing the Levenshtein distance between two strings where the Levenshtein distance is defined as the minimum number of edits needed to transform one string into the other, with the allowable edit operations being insertion, deletion, or substitution of a single character. This technique ensures that variations in user profile variables are handled correctly. The implementation of the proposed approach is given in later section. While designing the HIASN, few challenges evolved and are being discussed in the next section.

III. Design Challenges

Although during the initial phase, designing HIASN seems to be simple task. However, following issue makes profile aggregation across social networks a stimulating task:-

1 Social Networks have diverse network structures and profile attributes for serving the functionality that makes the task of linking profiles difficult.
2 Users may choose their username depending upon the functionality and service of the social network that may not be associated to their real identity.
3 It is evident challenge that many users may exist with identical usernames.
4 Users may provide false information across their profile in order to masquerade.

To the best of our knowledge, in order to identify a user, the publicly available information is used conventionally. However, few works [8] also depend on innocuous activity to identify users. Goga et al. [8] have used the location timing and writing patterns to enhance the quality of results. However, it has been observed that, rather than looking for all the locations, timing, and what the user has written, focus should be on the activity such as change in the location of the activity at one social network and same should be mapped to another.

In order to resolve the issues highlighted above, a solution addressing the needs is strongly desired. Hence, the literature was further grilled [31][32] and it was discovered that no best solution exists for mapping the user identity across social network. A hybrid solution exploiting phonetic encoding score and the Levenshtein algorithm is proposed. It is worth mentioning that prior to the Levenshtein algorithm, the Jaro-Wrinkler algorithm has been used by various authors [8][9][10] playing key role in computing string similarity. Jaro-Wrinkler is the modification of Jaro distance that calculates the string similarity as the sum of the number of common characters and the count of transposition as a weighted score for prefixes. The strings are more similar if the jaro-wrinkler distance is less.

Christen [33] has done an extensive study to compare the techniques for mapping string as personal name. It has been observed that choosing the right algorithm to match two short strings depends upon the performance of the system. In general, Jaro-wrinkler and Levenshtein distance are expensive algorithms as it involves enormous number of evaluations because each string will be equated to every other string in the dataset. Identifying similar string using phonetic encoding and then applying sophisticated string matching algorithm will provide better performance and results.

As illustrated in the previous section, Jaro-Wrinkler algorithm is expensive approach when applied for each name in millions of records and thus has not been considered while designing HIASN.

IV. The HIASN

HIASN is the combination of phonetic encoding score and the Levenshtein algorithm. HIASN determines user digital identity across several social networks targeting towards improving the search efficiently and precisely.

Exploiting the fact that each social network provides various public attributes to identify the user's digital footprints across an aggregated social network environment, the HIASN offers a three-phase solution i.e. identification of attributes contributing towards establishing user footprints, mapping of identified user profiles and finally produce a single integrated profile. Upcoming section illustrates each of the above listed phases.

4.1 Identification of Contributing Attributes

A social network runs a set of services [2] to ascertain a unique identity using publically available attributes. However, more 'mined' attributes are indispensable to govern the search. In addition to the user ID and name which have proved to be most promising attributes to recognize a user, HIASN considers location of the newsfeed/tweets of the user as an additional attribute to match the user. While considering the location, a weighted score of location and change in location of the user is evaluated. A set of twitter location – dependent users are collected using Twitter API to map Twitter users to LinkedIn users. The Twitter profiles are enriched with the results of FullContact and Bing Search. The data collected from multiple social networks is highly unstructured and thus needs

preprocessing. Following preprocessing, the relevant features are extracted like UserID, Name and Location from profile as well as tweets/newsfeeds for mapping two profiles. Algorithm 1 demonstrates the working of this phase.

```
Algorithm 1 Identification of Contributing Attributes
Input: User_Name to Social Network like Twitter
Output: Features_Extracted
{
users = searchTwitter(User_Name)
users = searchFullContact(User _Name)
users = searchBing(User_Name)
matching_users = search LinkedIn(User_Name)
p_users = preprocessing(users)
p_matching_users = preprocessing(matching_users)
Features_Extracted = Extract_Feature(p_users)
Features_Mapped = Extract_Feature(p_matching_users)
}

Extract_Feature(processed user)
{
For each processed_user in processed_users
    UserID = Extract_UserID(processed_user)
    UName = Extract_UserName(processed_user)
    ULoc = Extract_UserLocation(processed_user)
    Tweets = Extract_Tweets(processed_user)
    For each Post in SN // For Ex. Tweets for Twitter
    P_Loc = Extract_PostLocation(Post)
    return Features
}
```

4.2 Identity Resolver Module (IRM)

Since username is the unique attribute for each user across different social networks, it is possible to determine the mapping of user profiles using this as a major attribute. However, mapping the similarity for UserID is a challenging task as users may use different ID's to log on to the network such as email ID/name. The IRM employs Phonetic encoding score and Levenshtein algorithm for computing similarity between usernames/userID.

The decision whether two profiles are the same or not is taken by the change in the location factor. HIASN extracts the feeds from one social network and finds the change of the location i.e. if the user has covered a distance on the basis of longitude and latitude more than a significant threshold value. The probability of matching a profile increases if the location of the user differs with the same value on another social network. This change in location is mapped with latitude and longitude using Google APIs. IRM computes combined weighted score to determine the location, based on the Euclidean distance between two location using the latitude and longitudes. Any change in the location of the latest activity feed/tweet is mapped resolving the disambiguate user profiles. Thus, the resulting Equivalence IRM vector is

IRM_Vector：<User ID, Username, Location>

where IRM_Vector is the final Score for resolving the user's identity. The architecture of the system is shown in figure 2.Working of IRM is given in algorithm 2.

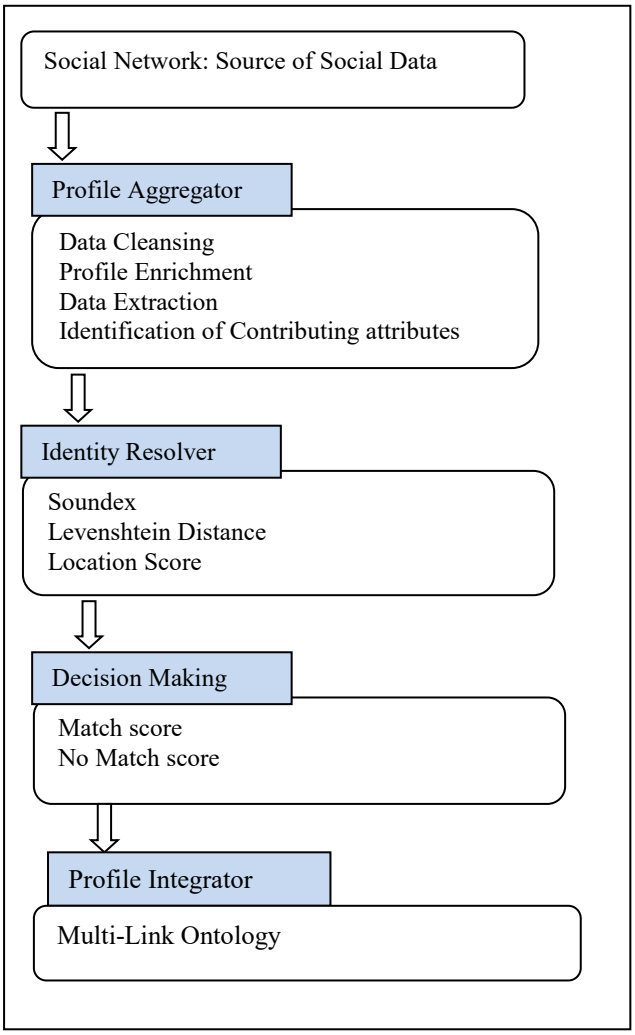

Figure 2. Architecture of Profile Integrator

The raw data is obtained from multiple sources which are highly unstructured and contain noisy information. Text cleaning was then performed on these variables. An analysis of the same was carried out using MongoDB.

Data reduplication is performed using flexible name matching techniques on entire raw data for the purpose of removal of duplicate content, unification of similar profiles, and enrichment of data metrics obtained from different sources. The name matching algorithms i.e. Phonetic Matching is applied to find names that are phonetically similar. Levenshtein distance technique ensured that variations in user profiles are handled correctly.

The HIASN is strengthened by matching the location field. Euclidean distance is applied on the location field to map the user location by extracting latitude and longitude from Google API [29]. Several sites provide the location when a user posts/tweets in the social network. The module extracts and cleans the location attribute of the profile and the posts and computes Euclidean distance between two locations of the profile. A change in the location of post more than a fixed value is observed and mapped to another network to verify the

similar change. This fixed value can be taken as a change in the region or country. Euclidian distance between these two locations of the post is calculated and a combined score is considered.

```
Algorithm 2: The Identity Resolver Module (IRM)

Input : Features_Extracted of Potential Profiles of Twitter,
FullContact, Bing and Feature_Mapped of LinkedIn
Threshold value, t
Output: Matched_Profiles

For Each Features_Extracted and every Feature_Mapped
MIDS = MatchIDScore(UserID, Feature_Mapped)
  MNS = MatchNameScore(UName, Feature_Mapped)
  MLS = MatchLocScore(ULoc, Feature_Mapped)
  Average = (MIDS+MNS+MLS)/3
 If Average > t
   Matched_profiles = Matched_Profiles + 1
return Matched_Profiles

MatchIDScore(UserID, Feature_Mapped)
P_Score= PhoneticScore (UserID, Matching_UserID)
  if (P_Score > x)
  L_Score =LevenshteinScore(UserID, Matching_UserID)
  MIDS = (P_Score + L_Score)/ 2
return MIDS

MatchNameScore(UName, Feature_Mapped)
P_Score= PhoneticScore (UName, Matching_UName)
  if (P_Score > x)
  L_Score=LevenshteinScore(UName, Matching_UName)
MIDS = (P_Score + L_Score)/ 2
return MNS

MatchLocScore(UName, Feature_Mapping)
LE_Score = Comp_Loc(ULoc,Matching_ULoc)
For each post in SN.posts
  PL_Score = Comp_Loc(P_Loc[i],P_Loc[i+1])
  if PL_Score < y
 MPL_C = FindMatchingPLoc(PLoc[i])
 For each j in MPL_C
 if (PL_Score == Comp_Loc(MPL_C[j],MPL_C[j+1])
  MPL_C1 = Comp_Loc(PLoc[i],Matching_PLoc[j])
  MPL_C2 = Comp_Loc(PLoc[i+1],Matching_PLoc[j+1])
   Change_LocScore = (MPL_C1 + MPL_C2)/2
MLS= (LE_Score+ Change_LocScore)/2
Return MLS

Note: If x is more than 0.85 means the UName and UID are
more similar whereas if y is less than 0.70 means the
location is more dissimilar. This change is to be noted in
another social network.  The variation in value of x and y
because name are more similar than location.
```

Similarity score is taken as equivalent to mean score of the proposed techniques on User ID, name, and location. A threshold value of similarity score 0.85 is derived from manual testing of results. This threshold implies that those pairs having the similarity score of greater than or equal to 0.85 are categorized as relevant matching candidates. The list of most expected profile of user is identified and presented to the user. The user is asked to choose the profile which exactly matches an account on another network. Then, the chosen profile is integrated using multilink ontology.

4.3 Profile Integration Module (PIM)

In order to develop a single unique profile for a user, the multiple ontology approach is employed to model each user data source in combination for integration. It requires mapping between multiple ontologies to provide a global view to profile. All public attributes of the profile from different social network are now made visible to user, the choice of displaying the value of attributes solely depends upon the user. PIM provisions the flexible modification of attributes.

General attributes used in most online social networking sites are personal characteristics, friends, interests, groups, studies, and user created content. A multilink data structure is used to store the information across different social networking sites and provide global as view. Table 1 provides multilink data structure across multiple online social networks. Working of profile integration is depicted in Algorithm 3

```
Algorithm 3: Profile Integration Module
Input: Matched_Profile
Output : Unique_Profile
 For Each attribute in Matched_Profiles
      If (att_SN1(value) = att_SN2(value))
        Store att_SN1
      Else
       Create
          Multi_link_att(att_SN1(value), att_SN2(value))
```

S. No.	General property or attribute	Mapping of integrated profile to individual profiles across online social networks
1	Property : Name	Name: Name (SN1)
		Name: Name (SN2)
		Name: Name (SN3)
2	Property: Gender	Gender: Gender (SN1) = Gender (SN2)
3	Property: Image	Image: Image (SN1)
		Image: Image (SN2)
		Image: Image (SN3)
4	Property: Connections	Connection: Friend (SN1)
		Connection: Friend (SN2)
		Connection: Friend (SN3)
5	Property: Location	Location: Location (SN1)
		Location: Location (SN2)
6	Property: Birthday	Birthday: Birthday (SN1) = Birthday (SN2)

Table 1. Multilink Ontology

If the location is the same for two different social networks, then the system will preserve only one location; if it is different, it will keep both the locations. Noticeably, a generalized identity is being kept by sub-grouping it with individual values of each social network. A user can select any medium of the

social network to look for his/her profile and, above all, information is preserved at one place.

V. Implementation and Evaluation

Figure 3 depicts various phases of implementation.

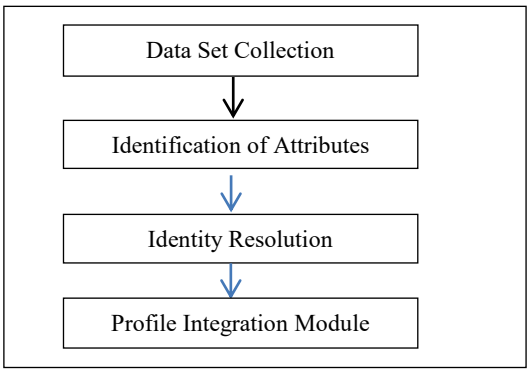

Figure 3. Phases of HIASN

HIASN was implemented on Intel Core i5 with 8GB RAM using Windows 7 Operating System. The ground truth data collection system extracts different data metrics from different social media platforms with the help of input search queries and crawling the publicly available information using Twitter API, Bing Search API, FullContact API and LinkedIn API. The obtained data is unstructured and preprocessed to filter the relevant data and then stored in the form of documents containing key-value pairs associated with different metrics. The following sources were used for data collection:

Twitter public search: Twitter public search API makes regular calls to Twitter servers and provides the past 7 days of a sample Twitter data when queried using an input query. The output documents obtained from Twitter contains both post level metrics, such as tweet text and location, as well as user level metrics, such as user name and gender. The HIASN extracted a total of 27956 documents from Twitter services.

Automated search engines: This component is built upon automated search engine queries that extract most close and related social profile URLs of a user. A screen name obtained from Twitter results was used as input to search engines that gave URL links of other channels of the same user. In order to ensure accurate results, a person name matching and deduplication mechanism were adopted. The Bing Search was used for this purpose.

Web services like FullContact was used in order to enrich the data obtained from Twitter with other channels metrics. The raw data obtained from different channels is then stored in a NoSQL database (mongoDB) hosted on cloud. The training and testing dataset is constructed on the basis of initial manual mapping for the user who exists on both the networks. The users who didn't match are taken as negative pairs for calculating precision and accuracy. The user information is collected and used for the research purpose only with the consent of the users. For evaluation of result, the proposed algorithm is applied on the data set to establish the accuracy, precision, recall, and F1 respectively with the sample data.

This section describes and evaluates the experiments on the data set from multiple social networks as discussed in previous section. The working Engine of HIASN is depicted in figure 4. The search vector used for the current search is:

IRM_Vector：<User ID, Username, Location>

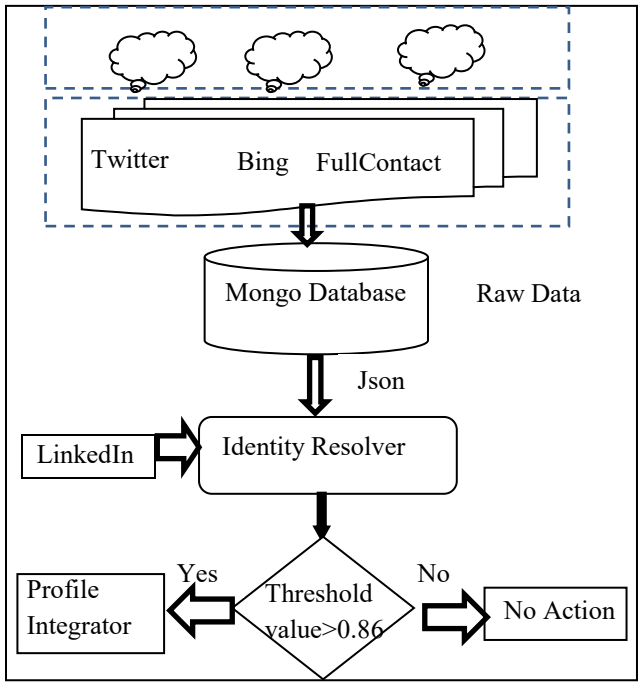

Figure 4. Working Engine of HIASN

The high quality examples contain all the similarity vectors for the profile pairs of the public profile dataset. The Same wide variety of poor examples was synthesized by arbitrarily pairing profiles that don't participate in the same end user and calculating their similarity vectors. This yielded a complete of instances. After training the classifier, Output was tested by giving as input a profile pair of two social networks to be classified as a "Match" or a "Not Match". Discriminability of an attribute is defined as the extent to which an account to other different accounts in multiple social networking sites. The function set with the finest accuracy, precision and recall using Naive Bayes was IRM_Vector ： <User ID, Username ,Location>. It has been determined that the functions UserName, UserID and Location using Geo-Location have been used on all the top 10 consequences, confirming that they are relevant attributes.

Briefly, the Phonetic Encoding and Levenshtein distance to find the similarity between names, and combined score of Euclidean distance to find the similarity between the location and change in location for the latest feed are the most promising attributes to resolve the identity of the user.

The matching score is calculated with the proposed algorithm and the results for each classifier are compared using the proposed vector achieving accuracy, precision, recall and F1 as 98%, 99%, 98%and 99% respectively. Naive Bayes classifier is trained to produce the likelihood that the comparability vector was produced by 2 profiles that have a place with a similar user. For every similarity vector of user profile, the possibility belonging to the same person is determined. Finally, we sort every one of the qualities of user's

account in diminishing request to shape a rank R. The evaluation is taken for the vector on the four classifiers Naive Bayes, Logistic Regression, SVM-Linear, and SVM-Kernel to assess the accuracy, precision, recall, and F1 score. Table 2 shows the result of multiple classifiers for the vector v with the proposed algorithm.

Classifiers	Accuracy	Precision	Recall	F1
Naive Bayes	0.987	0.998	0.983	0.990
Logistic research	0.965	0.995	0.946	0.97
SVM-Linear	0.982	0.992	0.966	0.979
SVM-Kernel	0.980	0.988	0.976	0.986

Table 2. Matching results

During the course of implementation of the work the following two vectors could also be observed and hence implemented. The Comparison of three possible vectors is as shown in the figure 5:

Vector_v1:<User ID,Username ,Location,description,image>

Vector_v2:<UserID,Username,Location,description,email, connection>

It is easy to observe from above that the proposed vector IRM_Vector：<User ID,Username ,Location> provides the best matching results. In 90% of the cases, the right profile was found at the top 5 ranks, while 50% if most of the public available attributes are used.

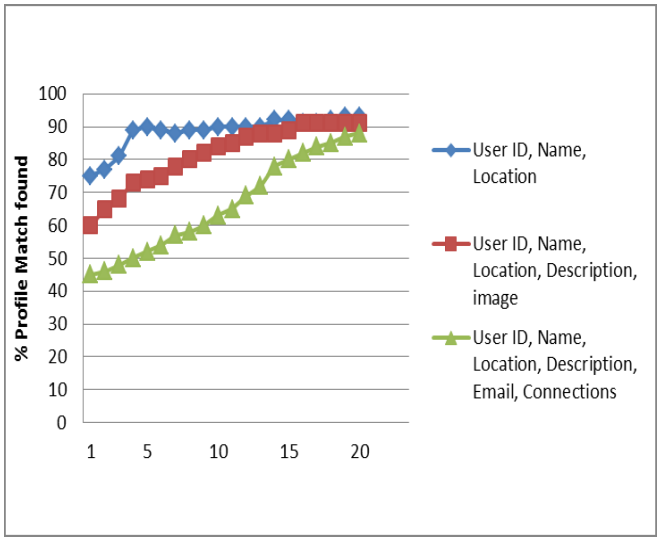

Figure 5. Comparison of Various Vectors

VI. Conclusion and Future Scope

This paper proposed to identify users from one social network to another using an efficient algorithm which strengthens name and location attributes. Only three publicly available attributes were used and achieved the best results in top ranks. User name, mined name, and mined location were analyzed and resulted in being the most discriminative features for achieving the best results. The adoption of these publicly available features allowed achieving accuracy, precision, recall, and F1 score as 98%, 99%, 98%, and 99% respectively. The system is applied to the real world user profiles extracted from various social networks and aggregators. An integrated profile is proposed that provides global-as-view to give a single profile to the user. HIASN will increase the discoverability of the user across multiple domains decreasing the email verification time or time required to fill the forms in the registration process. It will drive the associations and organizations to collaboratively execute promotions, discover new community and individuals. In the future, a query processor can be developed to extract useful information from this integrated profile.

References

[1] Schroeder, S.. "20 Ways To Aggregate Your Social Networking Profiles." Mashable. http://mashable.com/2007/07/17/social-network-aggregators/accessed October, 9, 2011.

[2] Irani, D., Webb, S., Li, K., &Pu, C. "Large online social footprints--an emerging threat". *International Conference on In Computational Science and Engineering, 2009. CSE'09.* (Vol. 3, pp. 271-276) 2009.

[3] Zafarani, R., & Liu, H. "Connecting Corresponding Identities across Communities". *ICWSM, 9,* pp. 354-357 2009.

[4] Rowe, M., &Ciravegna, F. "Harnessing the social web: The science of identity disambiguation." *Web Science Conference* 2010, April 26-27, USA.

[5] Kalemi, E., &Martiri, E. "FOAF-academic ontology: A vocabulary for the academic community." *In Intelligent Networking and Collaborative Systems (INCoS)", IEEE Third International Conference* pp. 440-445, 2011.

[6] Kabay, M. E. "Privacy issues in social-networking sites." *Network World,* 27, 2010.

[7] Korula, N., & Lattanzi, S. "An efficient reconciliation algorithm for social networks". *Proceedings of the VLDB Endowment*, 7(5), pp. 377-388, 2014.

[8] Goga, O., Lei, H., Parthasarathi, S. H. K., Friedland, G., Sommer, R., & Teixeira, R. "Exploiting innocuous activity for correlating users across sites." *In Proceedings of the 22nd ACM international conference on World Wide Web* pp. (447-458), (2013, May).

[9] Malhotra, A., Totti, L., MeiraJr, W., Kumaraguru, P., & Almeida, V. "Studying user footprints in different online social networks." *2012 IEEE/ACM International Conference on In Advances in Social Networks Analysis and Mining (ASONAM),* pp. (1065-1070), 2013, May.

[10] Pontual, M., Gampe, A., Chowdhury, O., Kone, B., Ashik, M. S., &Winsborough, W. H. "The privacy in the time of the Internet: Secrecy vs transparency". *In Proceedings of the Second ACM Conference on Data and Application Security and Privacy* pp. (133-140), 2012.

[11] Szomszor, M., Alani, H., Cantador, I., O'Hara, K., &Shadbolt, N. "Semantic modelling of user interests based on cross-folksonomy analysis." *In Proceedings of the International Semantic Web Conference Springer Berlin Heidelberg.* pp. (632-648), 2008.

[12] Surdu, A. C., & Pop, F. "Semantic approach for modeling profiles and interactions based on digital content." *19th IEEE International Conference In Control Systems and Computer Science (CSCS)"*, on pp. (232-239), (2013, May).

[13] Carmagnola, F., &Cena, F. "User identification for cross-system personalisation." *Information Sciences*, 179(1), 16-32, (2009).

[14] Singh, L., Yang, G. H., Sherr, M., Hian-Cheong, A., Tian, K., Zhu, J., & Zhang, S. "Public information exposure detection: Helping users understand their web footprints." I*n Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM),* pp. (153-161), (2015).

[15] Zheleva, E., &Getoor, L. "To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles." *In Proceedings of the 18th ACM International Conference on World Wide Web,* pp. (531-540) (2009).

[16] Liu, L., Cheung, W. K., Li, X., & Liao, L. "Aligning Users across Social Networks Using Network Embedding." *In IJCAI* pp. (1774-1780), 2016, July.

[17] Motoyama, M., & Varghese, G. "I seek you: searching and matching individuals in social networks." *In Proceedings of the ACM eleventh international workshop on Web information and data management* pp. (67-75), 2009, November.

[18] Irani, D., Webb, S., Li, K., & Pu, C. "Large online social footprints--an emerging threat." IEEE *International Conference on In Computational Science and Engineering, 2009. CSE'09.* Vol. 3, pp. (271-276), 2009, August.

[19] Perito, D., Castelluccia, C., Kaafar, M. A., &Manils, P.. "How unique and traceable are usernames?". *In International Symposium on Privacy Enhancing Technologies Symposium Springer Berlin Heidelberg*. pp. (1-17), 2011, July.

[20] 20 Zhou, X., Liang, X., Zhang, H., & Ma, Y. "Cross-platform identification of anonymous identical users in multiple social media networks." *IEEE transactions on knowledge and data engineering*, 28(2), 411-424, 2016.

[21] Iofciu, T., Fankhauser, P., Abel, F., & Bischoff, K. "Identifying Users Across Social Tagging Systems." *In ICWSM* 2011, Jul).

[22] Raad, E., Chbeir, R., & Dipanda, A. "User profile matching in social networks." *13th IEEE International Conference on In Network-Based Information Systems (NBiS),* pp. (297-304), 2010, September.

[23] Cortis, K., Scerri, S., Rivera, I., & Handschuh, S. "Discovering semantic equivalence of people behind online profiles." *In Proceedings of the Resource Discovery (RED) Workshop, ser. ESWC*, 2012.

[24] Doan, A., & Halevy, A. Y. "Semantic integration research in the database community: A brief survey." *AI magazine,* 26(1), 83, 2005.

[25] Golbeck, J., & Rothstein, M. "Linking Social Networks on the Web with FOAF: A Semantic Web Case Study." *In AAAI* Vol. 8, pp. (1138-1143), 2008, July.

[26] Narayanan, A., &Shmatikov, V. "De-anonymizing social networks." *30th IEEE Symposium* on *Security and Privacy* pp. (173-187), 2009, May.

[27] Zhang, H., Kan, M. Y., Liu, Y., & Ma, S. "Online social network profile linkage." *In Asia Information Retrieval Symposium*, *Springer, Cham.* pp. (197-208), 2014, December.

[28] Vu, X. T., Morizet-Mahoudeaux, P., & Abel, M. H. "User-centered social network profiles integration." *In 9th International Conference on Web Information Systems and Technologies*, *SciTePress.. p*p. (473-476) 2013, May.

[29] Iofciu, T., Fankhauser, P., Abel, F., & Bischoff, K. "Identifying Users Across Social Tagging Systems". *In ICWSM* 2011, July.

[30] Jain, P., Kumaraguru, P., & Joshi, A. "@ iseek'fb. Me': Identifying users across multiple online social networks." *In Proceedings of the 22nd ACM international conference on World Wide Web* pp. (1259-1268), (2013, May).

[31] Peled, O., Fire, M., Rokach, L., & Elovici, Y. "Matching entities across online social networks." *Neurocomputing*, 210, pp. 91-106, 2016.

[32] Shu, K., Wang, S., Tang, J., Zafarani, R., & Liu, H. "User Identity Linkage across Online Social Networks: A Review." *ACM SIGKDD Explorations Newsletter*, 18(2), 5-17, 2017.

[33] Christen, P. "A comparison of personal name matching: Techniques and practical issues." *IEEE International Conference on Data Mining Workshops, 2006. ICDM Workshops 2006.* pp. (290-294), (2006, December).

Author Biographies

Charu Virmani : Pursuing Ph.D. from Ymca University of Science and Technology, Faridabad, Haryana, India. Her subjects of interest include Computer Networks, Semantic Web and Data Mining.

Dr. Dimple Juneja: Post-Doctorate from Louisiana State University, USA and Ph.D. in Computer Engineering from Maharishi Dayanand University, Rohtak. Dr. Juneja is Dale-Carnegie Certified Trained Teacher and also is a recipient of Best Paper Award. Her subjects of interest include Multi-Agent Systems, Sensor Networks, Semantic Web and Distributed Operating System.

Dr. Anuradha Pillai: Ph.D. in Computer Engineering from Maharishi Dayanand University, Rohtak. Her subjects of interest include Semantic Web, Web Mining and Social Networks.