Novel Friend Recommendations Based on User-Generated Contents in Online Social Media Sites

Jiaxi Luo
Midwestern State University

The number of users joining and contributing to Social Networking Sites (SNSs) has steadily increased. Users share personal experiences such as food, shopping, travel, and leisure activities, among others, and also leave comments. The significant volume of User-generated content (UGC) offers ample data for constructing a model that accurately reflects users' sentiments towards various locations, interests, and personalities. The model should also be adaptable and scalable. These attributes can affect friend matches via the computer-supported social matching process. Regrettably, prior research in this domain omitted the use of extracted meta-text features in developing friend recommendation systems. This study proposes a text analytic framework the author applies to UGCs on SNSs. Extracting interests and personality features from UGCs enables text-based friend recommendations. The experiment results demonstrate that text-based features can enhance recommendation performance.

Keywords: text analysis, friend recommendations, social networks, user-generated content

INTRODUCTION

Social networking sites (SNSs) are online platforms designed to facilitate the development of social networks or social relations among individuals who share common interests, activities, backgrounds, or real-life connections. In recent years, SNSs have experienced a remarkable surge in popularity. For instance, as of 2022, Facebook has nearly 3 billion monthly active users (S. Dixon, 2023), while Twitter has over 500 million registered users (Dudley-Nicholson, 2013).

The rise of Social Networking Sites (SNSs) has led to collecting a tremendous volume of user-generated content (UGC), reflecting various aspects of users' lifestyles and patterns. With the rapid development of smart mobile devices and wearables, more, contextual information, such as location and health data, can now be collected. According to a survey conducted in Australia (Dudley-Nicholson 2013), 34% of social network users logged in at work, 13% at school, and 18% in the car. Additionally, 44%, 7%, and 6% of users used social networks in bed, the bathroom, and the toilet, respectively. As a result, SNSs have amassed extensive and comprehensive data on users over a long period.

UGC presents the untapped potential for various business and consumer intelligence applications. Researchers and business analytics professionals have been attracted to the possibilities of UGC and are exploring ways to use it effectively. Academic researchers study user behavior patterns, trends, and activities to integrate this information into existing social behavior theories. Meanwhile, business analytics professionals use UGC to personalize promotions and engage in customer relationship management by addressing user issues across social network channels (Woolridge 2011) to increase sales.

Unfortunately, very little research has utilized UGCs for friend recommendations, even though it is one of the most vital features of social network sites. Friend recommendation systems aim to connect people with shared interests and backgrounds to help SNSs avoid the cold-start problem, facilitate network speed, and enhance the quality of users' activities. Currently, existing friend recommendation systems rely on simple profile matching or friend network matching:

Basic profile matching has several drawbacks when recommending strangers to users in SNSs. Firstly, it does not comprehensively analyze a user's life pattern and interests since the matching is restricted to the user profiles provided by the SNS. Secondly, new users may not have complete profiles, while old users may forget to update them. These issues may lead to inaccurate recommendations (Zheleva et al., 2010).

On the other hand, the friend-of-friend method is an efficient and cost-effective way to recommend existing friends because it analyzes the entire social network and identifies overlapping links of friends, indicating real-life connectivity of users (Al Hasan et al., 2006; Lichtenwalter et al., 2010). However, this method is not useful for recommending unknown users to each other because two strangers sharing the same interests are unlikely to have any common friends. Therefore, the friend-of-friend method may not be suitable for making such recommendations.

According to Terveen and McDonald's (2005) categorization, user attributes fall under six types: demographics, social ties, geo-temporal information, interests, personality, and needs. In this article, the author proposes a novel text analytics framework: The author extracts users' writing styles, assess document readability, assigns sentiment scores based on location, and auto-recognizes personality scores by analyzing user-generated texts. Our friend recommendation system leverages these attributes to enhance the accuracy and precision of its recommendations. Our framework represents the first-known example of utilizing personality and interest attributes in text mining for friend recommendation systems.

RELATED WORK

Friend Recommendation Systems

Recommendation systems for online social networks can be categorized into two types based on the recommended objects: item recommendation systems and friend recommendation systems (Adomavicius et al., 2005; Adomavicius and Tuzhilin, 2005). Item recommendation systems suggest interesting items, such as movies, songs, books, and other products, to a user. Friend recommendation systems recommend homogeneous users in the same social network to a given user to help the user discover expertise, potential friends, old acquaintances, and so on.

Item recommendation systems have been extensively researched recently (Arazy et al., 2010; Chen, 2013; Christidis and Mentzas, 2013; Deng et al., 2013; Gavalas and Kenteris, 2011; Park et al., 2012; Sankaradass and Arputharaj, 2011). Three types of filtering methods are commonly used in item recommendation systems: collaborative, content-based, and hybrid. Collaborative filtering methods depend on user-item interactions, such as the frequency of item browsing/purchase or item rating by users. Content-based filtering methods concentrate on the attributes of an item without considering interactions between users and items. As the name suggests, hybrid methods combine collaborative and content-based filtering methods.

Despite their increasing importance to users and service providers in online social networks, friend recommendation systems have received less attention than item recommendation systems (Tian et al., 2010b). Friend recommendation systems have the potential to assist new users who join SNS without any existing friends (Adomavicius and Tuzhilin, 2005; Park et al., 2012). New users with no friends on the platform may feel isolated and disconnected, leading to a decreased likelihood of sharing activities with other users and a potential decision to leave the platform. Recommending suitable friends to new users is essential. Additionally, existing users can benefit from friend recommendation systems by finding users who share similar interests and habits, which can broaden their friend networks, enrich their social activities, and improve their satisfaction with the SNS while enhancing their loyalty to the website.

In addition to user benefits, friend recommendation systems can also be advantageous to service providers. These systems can increase social network densities and promote higher levels of active interactions among users, providing valuable channels for disseminating news, advertisements, and trends, which can translate into significant market potential. By strengthening users' social connections, service providers can increase their market share of their services, leading to potential business growth.

Major SNSs such as Facebook and LinkedIn have recently added "people you may know" features to their homepages, which suggest new connections based on users' contact information, profiles, or common friends (Scellato et al., 2011). However, these features primarily recommend people the user already knows in real life based on shared attributes. Nonetheless, there are situations where recommendation systems that include a more comprehensive range of users, including strangers who share similar interests, would be more valuable to users. For instance, when a user travels to a new city, meeting new friends with similar habits and lifestyles can be valuable since they are more familiar with the city and can offer advice catered to their interests.

Another example is an online dating website, which enables individuals to communicate and potentially develop a personal, romantic relationship over the Internet. In this context, it is more desirable to recommend someone who shares similar interests but is previously unknown rather than recommending a friend already known to the user (Menon et al., 2003).

While more comprehensive recommendation systems can be very useful in online social networks, little research has focused specifically on recommending unknown people or making inclusive recommendations. Most existing friend recommendation systems rely on simple strategies, such as suggesting a "friend-of-friend," as seen in Facebook (Chen et al., 2009b), or matching users' profiles.

A recommendation system that includes these individuals could provide even more value, an opportunity the service provider would not want to miss.

Text Features in User-Generated Contents

In machine learning and pattern recognition, a feature refers to a measurable heuristic value of the phenomenon under consideration that describes a specific aspect of an item. In our current situation, a user's interests or personality can be represented by text features. Content analysis of the text has been a fascinating research area across disciplines such as sociology and business. Researchers have gathered meta-information from numerous documents, including shallow and insightful ones. Scholars have proposed a list of text feature variables, categorizing them into the following broad feature types: 1) shallow meta-information; and 2) deeper insights. It is worth noting that the shallow meta-information, which can readily be inferred from the text, could have a powerful impact in describing a user's personality. Examples of such features include:

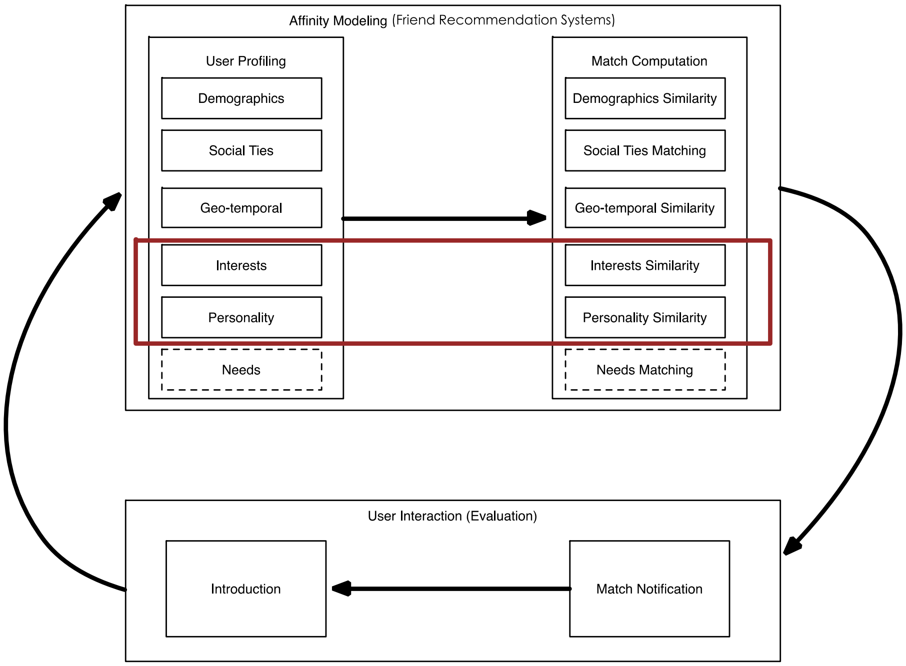
- Document Length: These features measure the document text's quantity, such as the number of words, sentences, and lines/paragraphs. The usage of few words or sentences indicates individuals who are straightforward and prefer imperative or mandatory sentences. By contrast, those who use many words may be attentive and gentle.
- Writing Style: Like document length features, these features measure the average number of syllables per word, the average number of words per sentence, and the percentage of complex words. They describe the user's writing style. A more intricate writing style may suggest someone with a higher educational background, who enjoys reading complex books, or is older. Conversely, using simple words or short sentences may indicate a younger person with a less complex nature.
- Readability: There are several indexes or scores available to measure document readability. For instance, the Fog Score developed by Gunning (1952) is well-known and has a straightforward calculation formula. This index indicates the number of years of formal education an individual of average intelligence could comprehend the text on the first reading. Scores range from 18 for unreadable, 14 for difficult, and 8 for childish. On the other hand, the Flesch-Kincaid grade level score rates text by the U.S. grade school level. A score of 8.0 means an 8th grader can understand the document. Scores between 7.0 to 8.9 are considered optimal.

The author employs text-mining techniques to extract meta-features from entire documents. Our features categories are as follows:

- Sentiment: By utilizing natural language processing, text analysis, and computational linguistics techniques, the author can identify the polarity of opinions in text. The author annotates each word in a text with its polarity strength, which the author obtains from an existing word list's polarity cues. Through sentiment analysis, the author discerns documents expressing positive or negative ideas within different contexts, such as various geographic regions. Using sentiment features, the author can identify individuals who favor or disfavor a specific location, which aids the system in determining users' interests.
- Subjectivity: As with sentiment features, subjectivity is another form of opinion-mining technique that rates text as either more subjective or more objective using text-mining algorithms.
- Personality: Prior studies have shown that psycho-linguistic attributes, frequency-based analysis of the lexical level, emotive words, and secondary clues like the number of first or second person pronouns can help in automatic personality detection. In this study, the author employs the widely utilized Big Five scheme for Personality Recognition from Text, which demonstrates consistency across age and gender and remains valid when using various tests and languages. The features in the Big Five are as follows:
    - Openness to experience (tendency for non-conventional, abstract, symbolic thinking vs. preference for non-ambiguous, familiar, and non-complex things)
    - Conscientiousness (tendency for long-term planning vs. impulsive and spontaneous behavior)
    - Extraversion (tendency for active participation in the world around vs. concentration on one's own feelings)
    - Agreeableness (tendency for eagerness to cooperate and help vs. self-interest)
    - Neuroticism (tendency to experience negative feelings and being overemotional vs. emotional stability and calmness)

By utilizing the aforementioned text features, the author can extract users' interest and personality attributes. This information and the interest and personality attributes extracted from the social matching process can help the system identify the degree of user matching. To make friend recommendations, the author proposes a text analytic framework that examines user-generated content (UGC) to extract these attributes. The following section details the workings of this framework.

FIGURE 1
COMPUTER-SUPPORTED SOCIAL MATCHING PROCESS WITH TEXT FEATURES

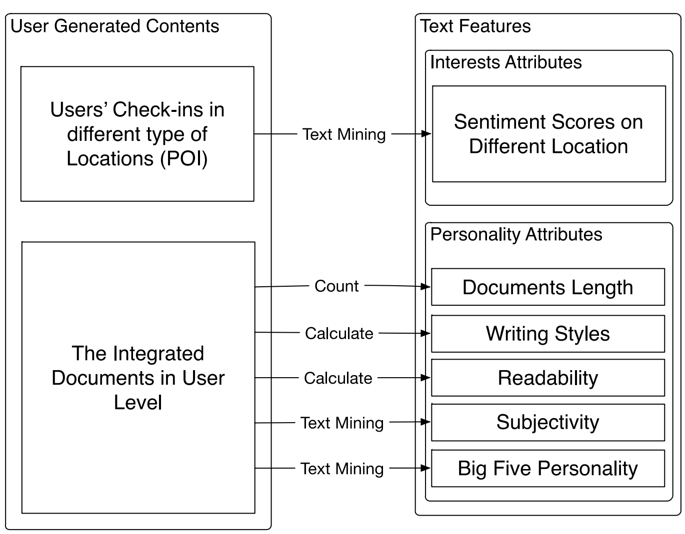


MODEL

Overview

The text analytic framework's process is based on the Computer-supported Social Matching Process theory. This theory posits a process to identify individuals with compatible interests and personalities.

FIGURE 2
THE TEXT ANALYTIC FRAMEWORK



(1) *Separate UGCs into different categories based on users' check-in location type and then integrate the UGCs into one document for analysis*. Many UGCs in social networks, such as Twitter and Foursquare, have a limitation in the number of characters because of the fact that short messages are propagated more readily. That presents some difficulties in text analysis. In this study, our solution is to integrate more than one piece of text into a document. Based on the location type, the author integrates texts into different categories. For example, the most popular location-based service, Foursquare, has nine major points of interest (POI) types: Art, College, Food, Professional, Nightlife, Outdoors, Shop, Travel, and Residence. To analyze these nine different documents, the author could extract users' interests in different locations. Then, for users' personality extraction, the author also needs to combine all comments from a certain user into one document. One document for one user will be eligible for analysis with enough words.

(2) *Count the document's length features in the integrated document*. Counting document length is quite straightforward. By splitting documents according to stop marks (periods) and blank spaces, the author can get the number of words and a number of sentences.

(3) *Calculate the writing style features in the integrated document.* According to Ott and Meurers' (2011) research study, a Perl package is available for calculating the number of syllables, which has an accuracy of approximately 90%. To analyze the UGCs, the author uses the Java Fathom Java package to calculate three features: the average number of syllables per word, the average number of words per sentence, and the percentage of complex words, which are words of three or more syllables. By leveraging these features, the author can gain valuable insights into the complexity and readability of the UGCs, which provides a better understanding of users' interests and personalities.

(4) *Calculate the readability scores for the integrated document.* Readability scores are defined as the grade level required for readers to read and understand a document. Numerous methods for calculating readability scores have been studied in research, including the Automated Readability Index (ARI) developed by Senter and Smith in 1967, the Coleman-Liau Index introduced by Coleman and Liau in 1975, the Flesch-Kincaid Readability Test created by Kincaid et al. in 1975, and the Gunning Fog Index proposed by Gunning in 1952. These methods provide different formulas to measure the complexity and readability of a document, enabling researchers to analyze UGCs and gain valuable insights into users' interests and personalities. The formula for Flesch-Kincaid Test is:

$$206.835 - 1.015\left(\frac{total\ words}{total\ sentences}\right) - 84.6\left(\frac{total\ syllables}{total\ words}\right)$$

The formula for Gunning Fog Index (Gunning 1952) is:

$$0.4\left[\left(\frac{words}{sentences}\right) + 100\left(\frac{complex\ words}{wirds}\right)\right]$$

In this study, the author uses the Gunning Fog Index and Flesch-Kincaid Test to score the text features for readability.

(5) *The author uses the text-mining package Opinion Finder from the University of Pittsburgh to analyze the text document and extract the subjectivity scores*. Subjectivity could be used to explain what influences and informs people's judgments about truth and reality. Opinion Finder uses a rule-based subjectivity classifier, which relies on manually crafted rules to tag sentences in a document as subjective or objective with high precision and low recall. The author then calculates the percentage of subjective sentences with sentiment scores ranging from 0.0 to 1.0, with 1.0 meaning all subjective and 0.0 meaning all objective.

(6) *The author uses auto-recognized personality techniques to calculate the Big Five Personality scores*. Poria et al. (2013) proposed a novel architecture for recognizing personality scores by leveraging common sense knowledge with associated sentiment polarity and affective labels. Specifically, they designed five Sequential Minimal Optimization (SMO)-based supervised classifiers to measure the Big Five personality traits, as defined by John and Naumann in 2008. The evaluation of their study yielded a precision score of approximately 0.6-0.7. In this study, the author adopts their algorithm by using the Linguistic Inquiry and Word Count (LIWC) and MRC Psycholinguistic Database and combining them with common sense knowledge-based features extracted by septic computing techniques. The author calculates the Big Five personality scores through this approach, enabling the author to gain valuable insights into users' personalities and interests.

(7) *For each type of location, the author calculates the sentiment scores of the documents by using sentiment analysis techniques*. Sentiment refers to people's attitude, opinion, or feeling towards a particular object, whether it's a person, organization, product, or location. By leveraging text-mining techniques and natural language processing, the author can determine the polarity of the text, which can be positive, negative, or neutral. In this study, the author uses AlchemyAPI to analyze the overall document and identify whether the sentiment is generally positive or negative in certain types of locations.

Following the analysis of user-generated text, the author acquires information regarding users' interests and personality traits. The author then inputs these text features into our recommendation model, enabling the author to generate personalized user recommendations. By leveraging the insights gained from the analysis of UGCs, our recommendation model can offer tailored recommendations that align with users' preferences and personalities.

Our recommendation system has the following process:

(1) The system analyzes all users' demographic attributes, social-tie attributes, and location attributes. Then it combines them with the attributes extracted from the above framework, which are interest attributes and personality attributes.

(2) The system compares a user's attributes with all other users' attributes, generates the similarities between two users, and then generates pairwise records.

The author uses Jaccard coefficient (Salton and Michael 1983) in this study, which means the distance between two users is:

$$d(a,b) = \left| \frac{a - b + \delta}{(a + b) + \delta} \right|$$

(3) The author employs data mining techniques to classify our records into two categories: Friend or Not Friend. The author wants to use probability of the classification results as the output.

(4) The system sorts the outputs and then selects the top M users as the recommendation list for the user.

EXPERIMENT

Data

Many online social networks have made their platforms accessible to developers through APIs, allowing them to access and collect data from these platforms. To obtain our test data, I applied to become a developer for Foursquare, Twitter, and Facebook.

I began collecting experiment data in October 2011 by using Foursquare's open APIs. I wrote a Java program to scan the public timeline and randomly select users who had checked in. However, because of authorization and privacy restrictions, I had to send friend requests to these users and wait for them to accept before I could collect their check-in information. Between October 2011 and February 2013, I collected data from 998 users and 6,417 check-in records. I also recorded the friendship network, which consisted of

4,074 pairs of friends. The network had low connectivity and density, with an average of four to five friends per user. Additionally, using Foursquare, I created the POI dataset, which includes 420 subtypes under nine major categories.

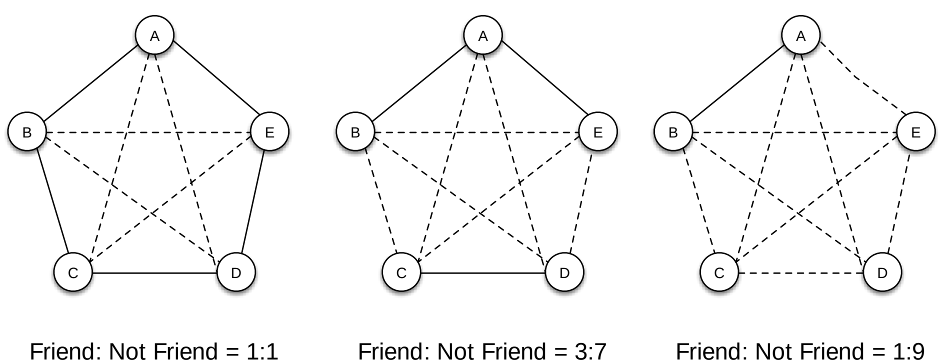
This model allowed me to collect data not only from Foursquare but also from Facebook and Twitter. Of the 998 users I collected from Foursquare, 754 also had Facebook and Twitter accounts. I used the account IDs from Foursquare to connect to the Facebook and Twitter websites. From Facebook, I obtained users' demographic information, including their religion, political orientation, age, educational background, work background, language, and favorite sports. If data was unavailable on the website, I manually visited each user's page to collect as much data as possible. From Twitter, I collected a significant amount of user-generated text data. Combined with the comments of check-ins from Foursquare, the text dataset was substantial, with a total of 17,890 pieces of text. On average, each user had 18 pieces of text available for creating a document.

Experiment Design

To control the experiment, three variables were considered: connectivity of the friend network, attribute groups, and the number of friends to recommend. After calculating the similarity between user pairs, the model transformed the recommendation query into a classification query. Each classification record contained two users, and their similarity attributes, and the dependent variable was whether the two users were friends or not.

To simulate real-world online social networking connectivity, the author aimed to create datasets with different densities of connections. As our original dataset had a relatively sparse friend network density, with only 1% of links being friend links, the author selected links in our test dataset to create social networks with different proportions of friend/non-friend links. For example, in a simple five-user network, the author selected A-B, B-C, C-D, D-E, E-A as friend links to create a 1:1 proportion of friend/non-friend links. The author used the same approach to create datasets with 1:2, 1:5, and 1:10 proportions of friend/non-friend links. The resulting datasets had different users, with the 1:1 dataset having 835 users, the 1:2 dataset having 891 users, the 1:5 dataset having 936 users, and the 1:10 dataset having 957 users. The number of users could potentially impact the experimental results, which the author will discuss later.

FIGURE 4
EXAMPLE OF PROPORTION OF FRIEND: NON-FRIEND



Second, the author wants to examine the performance between different attribute sets. To compare our model with the existing profile matching recommendation methods or "friend-of-friend" recommendation method, the author selects different groups from Table 1:

TABLE 1
TEST ATTRIBUTE GROUPS

Group	Attribute Data set
Group 1	Demographic Attributes Only
Group 2	Demographic Attributes + Interests Attributes + Personality Attributes
Group 3	Demographic Attributes + Social Ties Attributes
Group 4	Demographic Attributes + Social Ties Attributes + Interests Attributes + Personality Attributes

To evaluate whether the location attributes improve the performance of simple profile matching and "friend-of-friend" recommendations, the author compares Group 1 with Group 2 and Group 3 with Group 4, respectively.

Finally, in our evaluation, the author considers varying the number of friends recommended. If the author recommends too few friends, it may decrease the likelihood of users finding a suitable match. However, recommending too many friends could appear haphazard and create difficulty for users in making a selection.

Results

The author uses the experiment platform Weka 3.6.10; in the classification test settings, the author uses 10-fold cross-validation; and the author first uses accuracy as the evaluation result. By definition, it has:

$$accuracy = \frac{number\ of\ true\ positive + number\ of\ true\ negative}{number\ of\ records\ in\ test\ data\ set}$$

Due to the highly skewed friend/non-friend network dataset, where the proportion shifted from 1:1 to 1:10, the classifiers may have produced a negative classification output to achieve higher accuracy. For a friend network with a proportion of 1:P, the baseline accuracy rate is calculated as P/(1+P). To mitigate the impact of classification bias, the author incorporated cost sensitivity into our evaluation. The accuracy results are as follows:

TABLE 2
ACCURACY OF FRIEND RECOMMENDATION IN COST-SENSITIVE CASE

Attribute Sets	1：1	1：2	1：5	1：10
Baseline Accuracy	50.0%	66.7%	83.3%	90.9%
Group 1	52.4%	58.1%	78.4%	90.6%
Group 2	62.6%	75.3%	84.9%	91.9%
Group 3	78.4%	79.4%	80.4%	92.1%
Group 4	87.7%	89.6%	92.6%	94.7%

Based on the accuracy results, it is evident that in Group 1, where only demographic attributes were used, the recommendation outcomes were comparable to random guessing (baseline accuracy). Sparse profile attribute sets did not enhance the results. However, including textual information substantially improved accuracy, with all accuracy outputs in Group 1 exceeding the baseline accuracy. Similar trends were observed in Group 4 and Group 3, where the accuracy outputs improved significantly with the addition of textual attributes, indicating that textual information plays a crucial role in social-tie recommendations.

To conduct a more in-depth evaluation of the recommendation performance, the test must simulate the Top M recommendation process and calculate the precision. I accomplish this by using the classification probability results obtained from the Weka output to sort and select the top M users. Finally, the author calculates the accuracy rates to determine the likelihood of correctly recommending a true positive friend, meaning the system accurately predicts a friend as a friend.

To evaluate the precision of our algorithms, the author needs to calculate the baseline of precision and the optimal line of precision. Theoretically, the calculation for optimal precision only depends on the connectivity of social networks and the number of friend recommendations. But because our data set is very sparse, the author needs to consider each user's friend links.

Let us assume that the system has n users in a 1:P proportion network. For each user i, the system has the number of friend links (F_i) and non-friend links (Ni), and the system aims to recommend M friends from the list.

Baseline Precision is calculated as: For each user, if the total number of links $F_i + N_i$ is less than the number of recommendations M, then all friend links would be in the recommendation list, so the precision is F_i /M. Otherwise, the number of possible ways to select M links is C_{F+N}^{M}. The number of possible ways to select x friend links and M-x non-friend links is:$C_F^x \times C_N^{M-x}$. The expected precision of random Top M recommendation for this user is:

$$Baseline\ Precision_i = BP_i = \frac{\sum_{j=0}^{M} j \cdot C_{F_i}^{j} \cdot C_{N_i}^{M-j}}{C_{F_i+N_i}^{M} \cdot M}$$

Optimal Precisions is calculated as: For each user, and the selected friend link number will be: min (Fi, M), so, for Top M Recommendation, the optimal precision is:

$$Optimal\ Precision = (\sum_{1}^{n} \frac{\min(F_i, M)}{M}) \div n$$

Once the author has the Baseline and Optimal Precisions, I can calculate the relative positions of our recommendation precisions using the formula:

Position = (Recommendation Precision - Baseline Precision) / (Optimal Precision - Baseline Precision)

The Relative Positions of the Top 3 Friend Recommendations will be:

TABLE 3
RELATIVE POSITIONS OF TOP 3 FRIEND RECOMMENDATIONS

	1:1	1:2	1:2 Cost Sensitive	1:5	1: 5 Cost Sensitive	1:10	1:10 Cost Sensitive
Group 1	28.96%	-33.04%	32.95%	-11.17%	25.02%	-4.62%	16.61%
Group 2	37.48%	12.44%	35.86%	11.18%	27.14%	16.77%	23.24%
Group 3	58.59%	63.65%	63.74%	63.5%	61.58%	58.79%	40.68%
Group 4	68.84%	72.01%	72.30%	68.99%	65.06%	63.19%	54.30%

Based on the results, the author observed that including location attributes improved the Top 3 friend recommendation precision, as discussed earlier. Group 2 showed significantly better results than Group 1, and Group 4 exhibited slightly less improvement but still achieved significantly better results than Group 3. Furthermore, the author noted that cost-sensitive classification produced more reasonable results than the higher-biased dataset outcomes. Specifically, when only demographic attributes were considered, cost-sensitive classifiers had much higher precision than the non-cost-sensitive cases. Another trend the author observed was that as the proportion increased, the relative position decreased. This could be because as the dataset grows, it becomes harder to attain the optimal line in the Top 3 friend recommendations. The author observed these trends by manipulating the Top M friend recommendations.

DISCUSSION AND CONCLUSION

This study aimed to improve friend recommendations in a computer-supported social matching process by adding interest and personality attributes extracted from UGCs. Although many studies have explored UGCs, few have used them for friend recommendations. The text features from UGCs can help to build a user-topic model and explain users' interests and personalities. The author developed a text analytic framework to extract interest and personality attributes from users' text documents to achieve this. The author then calculated the similarity between users and made friend recommendations. The experimental results demonstrated that incorporating interest and personality attributes significantly improved recommendation performance, particularly in a sparse network. These findings have important implications for both research and practice.

First, this paper presents a novel approach to friend recommendations by incorporating a user-topic model, which, to our knowledge, has not been previously explored in this context. Prior studies have mainly focused on extracting emotions from UGCs to predict trends or facilitate information propagation. The topic-user model proposed in this study is typically used for document recommendations or expert findings. However, our study demonstrates that it can also be utilized for friend recommendations. Although this is a preliminary attempt, the author believes that further improvements could be achieved by employing more appropriate natural language processing methods, using a better-matched lexicon, or developing a more robust feature set.

Another contribution of this study is demonstrating the significance of interest and personality attributes in a computer-supported social matching process. The author can identify more relevant and interesting potential friends by integrating these attributes into a friend recommendation system. Current friend recommendation systems in social networks are limited. They primarily suggest people who already know each other and may struggle to find individuals with similar habits or interests. These limitations can be overcome by incorporating a more comprehensive set of attributes, leading to more active engagement and growth in social networks.

Additionally, our model can be extended to other applications beyond social networking, such as job matching or dating platforms, where interest and personality attributes are also important factors in matching. The algorithm and framework developed in this study can serve as a starting point for these applications and can be further customized to meet their specific needs.

Overall, our study provides a valuable contribution to the field of computer-supported social matching and has the potential to impact both academic research and practical applications.

Our research has some limitations that should be addressed in future studies:

One of the limitations is the diversity of languages used in the collected data, which makes it difficult to analyze and could impact the accuracy of the results. This issue could be addressed by limiting the data collection to a specific location or language.

Another limitation is the lack of a specific lexicon for Twitter, which could impact the natural language processing results. Future research should explore the effect of using different dictionaries and lexicons on the text analysis results.

Our recommendation system was evaluated based on existing social networks, which may limit its ability to recommend strangers. Future research should focus on recommending new acquaintances and exploring the satisfaction of recommendations among users.

Finally, our model only used five attribute sets in the computer-supported social matching process. Future research could include additional attributes, such as those related to question-answering social networks, to improve the recommendation system.

REFERENCES

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.

Adomavicius, G., Sankaranarayanan, R., Sen, S., & Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems (TOIS)*, *23*(1), 103–145.

Al Hasan, M., Chaoji, V., Salem, S., & Zaki, M. (2006). Link prediction using supervised learning. In *SDM 06: Workshop on link analysis, counter-terrorism and security*.

Arazy, O., Kumar, N., & Shapira, B. (2010). A theory-driven design framework for social recommender systems. *Journal of the Association for Information Systems*, *11*(9), 455–490.

Benchettara, N., Kanawati, R., & Rouveirol, C. (2010). *A supervised machine learning link prediction approach for academic collaboration recommendation*. ACM, New York, New York, USA.

Chen, I.B.X. (2009). A framework for context sensitive services: A knowledge discovery based approach. *Decision Support Systems*, *48*(1), 10.

Chen, J., Geyer, W., Dugan, C., Muller, M., & Guy, I. (2009). *Make new friends, but keep the old: Recommending people on social networking sites*. ACM.

Chen, X.L.H. (2013). Recommendation as link prediction in bipartite graphs: A graph kernel-based machine learning approach. *Decision Support Systems*, *54*(2), 10.

Christidis, K., & Mentzas, G. (2013). A topic-based recommender system for electronic marketplace platforms. *Expert Systems With Applications*, *40*(11), 4370–4379.

Coleman, M., & Liau, T.L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, *60*, 283–284.

Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Processes and Their Applications*, *62*(1), 139–168.

Deng, Z.-H., Wang, Z.-H., & Zhang, J. (2013). Robin: A novel personal recommendation model based on information propagation. *Expert Systems With Applications*, *40*(13), 5306–5313.

Dixon, S. (2023, February 13). Number of monthly active Facebook users worldwide as of January 2022. *Statista*. Retrieved from https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Dudley-Nicholson, J. (2013). *Australians now using social media in bedrooms and toilet cubicles*. Retrieved from http://www.news.com.au

Gavalas, D., & Kenteris, M. (2011). A web-based pervasive recommendation system for mobile tourist guides. *Personal and Ubiquitous Computing*, *15*(7), 759–770.

Gunning, R. (1952). The technique of clear writing.

Guy, I., Ronen, I., & Wilcox, E. (2009a). *Do you know?: Recommending people to invite into your social network*. ACM, New York, New York, USA.

Guy, I., Zwerdling, N., Carmel, D., Ronen, I., Uziel, E., Yogev, S., & Ofek-Koifman, S. (2009b). *Personalized recommendation of social software items based on social relations*. ACM, New York, New York, USA.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I.H. (2009). The Weka Data Mining Software: An Update. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18.

Jeh, G., & Widom, J. (2002). Simrank: A Measure of Structural-Context Similarity. *ACM*.

Jensen, C., Davis, J., & Farnham, S. (2002). Finding Others Online: Reputation Systems for Social Online Spaces. *ACM*.

John, O., & Naumann, L. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. *Handbook of personality: Theory and Research*, *3*, 114–158.

Khurri, A., & Luukkainen, S. (2009). Identification of Preconditions for an Emerging Mobile Lbs Market. *Journal of Location Based Services*, *3*(3), 188–209.

Kincaid, J., Jr., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). *Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel* (Research Branch Report).

Kullback, S., & Leibler, R.A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*.

Kuo, T.-T., Yan, R., Huang, Y.-Y., Kung, P.-H., & Lin, S.-D. (2013). Unsupervised Link Prediction Using Aggregative Statistics on Heterogeneous Social Networks. *ACM*.

Liben-Nowell, D., & Kleinberg, J. (2007). The Link-Prediction Problem for Social Networks. *Journal of the American Society for Information Science and Technology*, *58*(7), 1019–1031.

Lichtenwalter, R.N., Lussier, J.T., & Chawla, N.V. (2010). New Perspectives and Methods in Link Prediction. *ACM*.

Mayer, J.M., Motahari, S., Schuler, R.P., & Jones, Q. (2010). Common attributes in an unusual context: Predicting the desirability of a social match. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 337–340).

Menon, A.M., Deshpande, A.D., Perri III, M., & Zinkhan, G.M. (2003). Trust in Online Prescription Drug Information Among Internet Users. *Health Marketing Quarterly*, *20*(1), 17–35.

Newman, M. (2001). Clustering and Preferential Attachment in Growing Networks. *Physical Review E*, *64*(2), 025102.

O'Madadhain, J., Hutchins, J., & Smyth, P. (2005). Prediction and Ranking Algorithms for Event-Based Network Data. *ACM SIGKDD Explorations Newsletter*, *7*(2), 23–30.

Ott, N., & Meurers, D. (2011). Information Retrieval for Education: Making Search Engines Language Aware. *Themes in Science and Technology Education*, *3*, 9–30.

Park, D.H., Kim, H.K., Choi, I.Y., & Kim, J.K. (2012). *A Literature Review and Classification of Recommender Systems Research. Expert Systems With Applications*.

Poria, S., Gelbukh, A., Agarwal, B., Cambria, E., & Howard, N. (2013). Common Sense Knowledge Based Personality Recognition from Text. *Context Based Expert Finding in Online Communities using Social Network Analysis*, (8266: Chapter 42), 484–496.

Quercia, D., & Capra, L. (2009). Friendsensing: Recommending Friends Using Mobile Phones. In *Proceedings of the Third ACM Conference* (pp. 273–276).

Salton, G., & Michael, J. (1983). *Introduction to Modern Information Retrieval*.

Sankaradass, V., & Arputharaj, K. (2011). An intelligent recommendation system for web user personalization with fuzzy temporal association rules. *European Journal of Scientific Research*, *49*(1), 1–9.

Scellato, S., Noulas, A., & Mascolo, C. (2011). Exploiting place features in link prediction on location-based social networks. In *Proceedings of the 5th ACM workshop on Location-Based Social Networks* (pp. 43–52). New York, NY, USA: ACM.

Schifanella, R., Barrat, A., Cattuto, C., Markines, B., & Menczer, F. (2010). Folks in folksonomies: Social link prediction from shared metadata. In *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining* (pp. 271–280). New York, NY, USA: ACM.

Senter, R.J., & Smith, E.A. (1967). *Automated Readability Index*.

Shi, Z.W., & Andrew, B. (2013). Network structure and observational learning: Evidence from a location-based social network. *Journal of Management Information Systems*, *30*(2), 27–56.

Terveen, L., & McDonald, D.W. (2005). Social Matching: A Framework and Research Agenda. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *12*(3), 401–434.

Tian, Y., He, Q., Zhao, Q., Liu, X., & Lee, W.-c. (2010). Boosting Social Network Connectivity with Link Revival. *CIKM '10 Proceedings of the 19th ACM international conference on Information and Knowledge Management*, pp. 589–598.

Tobler, W.R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*(2), 234–240.

Wang, D., Pedreschi, D., Song, C., Giannotti, F., & Barabasi, A.-L. (2011). Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1100–1108). New York, NY, USA: ACM.

Woolridge, A. (2011). Too Much Buzz: Social Media Provides Huge Opportunities, but Will Bring Huge Problems. *Economist*.

Xu, B., Chin, A., Wang, H., & Zhang, L. (2011). Social linking and physical proximity in a mobile location-based service. In *Proceedings of the 2011 ACM international conference on Ubiquitous computing* (pp. 99–108).

Xuan Nhat Lam, Thuc Vu, Trong Duc Le, & Anh Duc Duong. (2008). Addressing cold-start problem in recommendation systems. In *Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication (ICUIMC '08)* (pp. 208–211). New York, NY, USA: ACM.

Yuan, H., Qian, Y., Yang, R., & Ren, M. (2014). Human mobility discovering and movement intention detection with GPS trajectories. *Decision Support Systems*, *63*, 39–51.

Zheleva, E., Getoor, L., Golbeck, J., & Kuter, U. (2010). Using friendship ties and family circles for link prediction. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1039–1048). Berlin, Heidelberg: Springer Berlin Heidelberg.