TWIN: Personality-based Intelligent Recommender System

Alexandra Roshchina[a], John Cardiff [a,*] and Paolo Rosso[b]
[a]*Social Media Research Group, Institute of Technology Tallaght Dublin, Ireland*
[b]*NLE Lab.-PRHLT Research Center, Universitat Politecnica de Valencia, Spain*

Abstract. This paper presents the "Tell me What I Need" (TWIN) Personality-based Intelligent Recommender System, the goal of which is to recommend items chosen by like-minded (or "twin") people with similar personality types which we estimate from their writings. In order to produce recommendations it applies the results achieved in the personality from the text recognition research field to Personality-based Recommender System user profile modelling. In this way it creates a bridge between the efforts of automatic personality score estimation from plain text and the field of Intelligent Recommender Systems. The paper describes the TWIN system architecture, and results of the experimentation with the system in the online travelling domain in order to investigate the possibility of providing valuable recommendations of hotels of the TripAdvisor website for "like-minded people". The results compare favourably with related experiments, although they demonstrate the complexity of this challenging task.

Keywords: Personality Recognition, Intelligent Recommender System, text processing, personality-based user model

[*]Corresponding author. E-mail: john.cardiff@ittdublin.ie.

1. Introduction

With the growth of the amount of information available on the Internet the possibility of manual processing and analysing of each individual piece of content posted by means of various Web 2.0 tools has become an extremely tedious task for many users. To narrow the search and personalise its results by following specific user preferences derived from the behaviour of the individual or explicitly defined by the user, the mechanism of Recommender Systems (RSs) has been explored and applied in various virtual communities.

While traditional RSs rely on demographic data or product ratings, the emerging field of Personality-based Recommender Systems (PbRSs) is developing ways of assessing the personality traits of the individual to construct the user profile. One of the challenging perspectives that we analyse in this paper is to apply automatic tools of personality recognition from the publicly available user data. The successful algorithm would minimise the efforts of the person to being correctly represented within the RS through the information automatically stored in the personality-based profile.

In this paper, we present the TWIN ("Tell me What I Need") Personality-based Intelligent Recommender System. In order to produce recommendations it applies the results achieved in the personality from the text recognition research field to Personality-based Recommender System user profile modelling. In this way it creates a bridge between the efforts of automatic personality score estimation from plain text and the field of Recommender Systems. TWIN also serves as a tool for visualizing the resulting scores to perform personality analysis. We show that the application of the TWIN in the online tourism domain produces valuable results in recommending tourist facilities to "like-minded" people.

The paper describes the architecture of the TWIN system, and the experiments conducted on the system using a dataset of reviews crawled from the TripAdvisor online tourism community website (http://tripadvisor.com). We demonstrate the construction of the personality profiles of the authors based on the NEO-Personality Inventory-Revised classification scheme (also known as the Big Five model) [1]. We then present the TWIN approach to recommending hotels to people with similar personalities, or "like-minded people", in a common situation of uncertainty.

This paper is organised as follows. The Section 2 gives an introduction to RSs and their role in the online world. In Section 3, we present and extensive review of the main approaches that exist in the field of PbRSs, and introduce the TWIN system in this context. Section 4 describes the experiments we have undertaken with the TWIN System conducted on the TripAdvisor dataset, along with an analysis of the results. Finally, the conclusions are presented in Section 5.

2. Recommender Systems

In the real world the person is surrounded by other people almost all the time so there is always the possibility to obtain help through word-of-mouth. When we are online, the process is mirrored by special types of intelligent Web services known as Recommender Systems. Ricci et al. [2] define RSs as "*software tools and techniques providing suggestions for items to be of use to a user* ". RSs help the person to make the right decision about choosing a particular piece of content (among a large number of alternatives: books, music, documents, web pages, jokes, etc.) and are applied as a part of successful marketing strategies by E-commerce firms [3].

Recently with the appearance of the various mobile devices and services such as GPS and Wi-Fi, a new dimension is being introduced to the RSs field. It provides a number of challenges due to the limitations of the mobile devices compared to PCs. Nevertheless it also brings new characteristics: *location-awareness* (the physical location of the person is known at any particular moment) and *ubiquity* (the information can be delivered at any time when it is required) [2]. Mobile Recommender Systems are appearing to serve a number of tasks including search for tourist attractions (services, places, etc.), route recommendations, news and multimedia recommendations, etc.

B & Kant [4] reviewed the existing approaches for constructing e-commerce recommender systems. Systems that perform the analysis of user reviews to calculate products recommendations face a problem of extracting relevant information (customers' sentiments and opinions) from relatively short textual abstracts. Some approaches concentrate on compositional semantic rules while other evaluate sentiment holding words numerically and try to rank their relative importance within

the sentence or additionally rely on the word order and the meaning of emoticons.

There are two main types of Recommender Systems: *collaborative filtering*, *content-based*. They are described in more detail below.

2.1. Collaborative filtering Recommender Systems

This is one of the most popular types of RSs that is widely used nowadays [5]. Collaborative filtering systems rely on ratings given by the users in order to provide recommendations. This approach allows the recommendation of items highly rated by similar individuals and does not require extensive knowledge about items themselves.

The growth of the amounts of data that should be analysed to get adequate recommendations requires specific mechanisms. For example, the MapReduce methodology [6] provides a way to collect information from various distributed sources. Dimensionality reduction strategies such as the matrix factorization method of singular value decomposition (SVD) [7] help to draw efficient generalities from the compressed data to improve recommendations. Babak et al. [8] also exploit cross-domain collaborative filtering techniques. This approach diminishes the influence of lacking preferences in the target domain by using the controlled flow of information from other fields of user's interests. The authors performed experiments with Amazon dataset that included ratings of users across four disparate domains (books, music CDs and DVDs, and videotapes). They utilised factorization machines to model the importance of the information in the target domain comparing to auxiliary domains and achieved a better than state-of-the-art performance for the cross-domain collaborative filtering system.

The classical example of a collaborative-filtering RS is Amazon (http://www.amazon.com) that recommends books and other goods based on the user's purchase behaviour and its similarity with what other people prefer to buy. Among the other examples is MovieLens (http://movielens.umn.edu). Users provide information about the movies they love or hate and the system is recommending new movies to watch based on the preferences of the people with the same taste. Another example is Netflix (https://signup.netflix.com) that recommends movies and TV shows that are rated by its users.

2.2. Content-based Recommender Systems

Content-based filtering is the oldest type of recommendation approach [9]. The user is supposed to feed the system with his or her initial ratings on a number of items for the system to be able to recommend pieces of content that have similar attributes to those that the user had already seen.

RSs of this type do not take other users' ratings into account. Therefore one of the main advantages of content-based RSs is that the user's unique taste is not smoothed by preferences of people from like-minded groups [10]. Thus, people with extreme likes will still receive appropriate recommendations. However, sometimes this advantage turns into drawback due to the *overspecialisation* - the emergence of the unbreakable circle of the very similar recommendations based on the unchanging interests of the person [11].

Content-based RSs try to provide recommendations based only on the attributes (author, title and other metadata) of the item or its content. Obviously, not all of the attributes have the same importance. When dealing with textual metadata, approaches such as Term Frequency-Inverse Document Frequency (TF-IDF) can provide a scheme to assign weights to specific features [12]. In the case of multivalued features Castro et al. [12] obtain weights by combining the inter-user dissimilarity and intra-user similarity of the TF-IDF scheme calculated by means of entropy and correlation between user ratings and features (Pearson correlation coefficient for numeric values and Cramer V contingency coefficient for nominal values). The authors claim that while TF-IDF is a commonly used approach that shows good results on boolean features, their scheme works on multivalued features and features from various domains. The scheme outperforms TF-IDF when the number of produced recommendations is less than or equal to 4 and requires less time for model building and producing a response.

The presence of mandatory metadata leads to another drawback of content-based RSs: if items do not contain any metadata initially (music, video), it is necessary to explicitly define it [10]. Recently the elegant solution to the above mentioned problem has been found with the appearance of social tagging facilities allowing people to explicitly annotate the content they are uploading to the system (as well as associating the existing item with the concept represented by the tag). The re-

sulting structure provides a simple self-organising classification approach known as *folksonomy* [13]. Folksonomy can be utilised for the construction of content-based Recommender System profiles describing items by the tags attached to them [14].

The lack of ratings in content-based and collaborative filtering approaches results in the incomplete user profile and the new user problem [15]. Such approaches as multi-attribute utility theory allow the user to evaluate only a few items based on some criteria to get the overall representation of preferences. Huang [15] followed this approach in order to construct a utility-based recommender system. The main problems were to minimise the effort of the user at the stage of weighting the importance of attributes for the utility function (usually created via multiple regression analysis, artificial neural networks, etc.) and to evaluate the resulting functions in different contexts. The authors exploited simple multi-attribute rating technique (SMART) and radial basis function network (RBFN) as they show better accuracy and require less effort from the user.

One example of a content-based RS is ACR News (http://www.acr-news.com). The site serves as an entry point for air conditioning and refrigeration professionals. The system provides daily news in the field by means of the Usenet news RS and utilises a content-based filtering algorithm. There are many examples of content-based RSs in the field of music and films recommendations such as music RS Pandora Radio (http://www.pandora.com) that suggests the music composition similar to the one the user already likes.

2.3. Personality-based Recommender Systems

The idea of Personality-based Recommender Systems is still an emerging trend. Such system is able to match people's interests at a deeper personal level through an introduction of a psychological profile to RSs. At present the variety of the proposed systems of this type is not extensive.

Nunes [16] in her research provides an overview of the state of the art of PbRSs. She claims that one interesting outcome of introducing a psychological dimension into the RS could be the possibility of products categorisation based not only on their attributes (price, physical parameters, etc.) but also on the effect they may have on the consumer.

Nunes mentions three pioneering works in the field. The concept of the emotional intelligence was incorporated into the user profile by Gonzalez et al [17] to create the Smart User Profile (which includes feelings, impressions and emotional states of the users as well as their moods retrieved by means of machine learning techniques). Masthoff and Gatt [18] experiment with predicting group satisfaction based on the individual user satisfaction (the more satisfaction and positive emotions the first purchase brings to the user, the more likely he will return to make another purchase). Saari et al. [19] have introduced the concept of psychological customisation to extract the user's emotional state, attention, learning abilities, etc. at the particular moment of working with the system.

Tkalčič et al. [20] proposed a personality-based approach for collaborative filtering RSs that follows the Big Five model. They make an assumption that the personality of the user doesn't change significantly over time and thus the neighborhood of the user can be calculated in advance (lower real-time computational effort). The authors analysed the influence of affective metadata on the performance of a content-based recommender system for images. The authors made an assumption that the specific mood the person desires to feel affects the type of the content the user is searching for. Therefore the addition of the affective information – personality types and emotive states – would help to increase the precision of recommendations.

A Personality-based music Recommender System was introduced by Hu and Pu [21] which was based on the correlations between musical preferences and personality types that follow the finding of Rentfrow and Gosling [22] of four preference groups constructed according to various styles of music compositions that people are fond of. For example, the "reflective and complex" group (which prefers jazz, blues and classical music) has correlations with openness to new experience Big Five dimension and "energetic and rhythmic" group (which tends to appreciate rap, hip-hop, funk and electronic music) correlates positively with extraversion and agreeableness. The personality data of the music RS has been calculated by means of the Big Five questionnaire. The similarity between the two users has been estimated based on the Pearson correlation coefficient with the personality scores. The authors have also tested the combined similarity measure that incorporates ratings data. They have shown that the personality-

based approach achieves a significant improvement over the baseline of considering only ratings data.

Elahi et al. [23] propose the enrichment of a variety of active learning techniques with personality data. The process of active learning involves the step of gathering preliminary information and the learning phase itself that involves classification of properties and recognition of patterns. In the field of RSs, active learning approach takes the form of explicitly asking the user to rate a few items in order to get initial ratings and avoid the cold- start problem. The authors use personality data collected through simple questionnaires to collect preliminary information about possible tastes of the users, avoid the cold-start and improve the quality of recommendations.

Tkalcic et al. [24] investigated the possibility of improving the performance of a content-based recommender system when labelling items with implicit affective metadata. In traditional recommender systems items have manual or automatic genre annotations (generic metadata) to distinguish one item profile from the other. Affective space with valence-arousal-dominance dimensions provides a tool for specifying particular emotions that items induct in different people while those are watching images and their facial expressions are being recorded. Emotions detected from such recordings are then converted into affective labels – affective metadata included into each item profile. The construction of the user profile requires the following: users (52 people in the conducted experiment) must provide explicit *binary ratings* (relevant/non-relevant) of a number of images (70 images from the IAPS dataset in the conducted experiment); *machine learning algorithm* must be trained to learn the user model from the training set – includes explicit ratings provided by the users (classes) and affective metadata of the particular image (features). Tkalcic et al. [24] claim that a recommender system (for predicting the relevance of the particular image) that exploits implicit affective labels acquired automatically performs *significantly better* than a recommender system based on generic labels and *significantly worse* than the one based on explicit affective labeling.

Quercia et al. [25] analyse the correlation between types of Twitter users (listeners, popular, highly-read and two types of influential users) and Big Five personality types and use their findings to predict personality based on the features available in any Twitter profile. They also discuss serious privacy issues that arise from the ability of automatically calculating precise enough reflections of people's tastes and behaviours.

It was found that most of the existing PbRSs utilise questionnaires (often following the Big Five model) in order to construct the personality profile of the user. One of the drawbacks of this approach is the amount of time required to complete the task. The other important factor is the correctness of the information provided by the user as questionnaire statements may be misinterpreted or answered inaccurately. Therefore an interesting task of automatic construction of the personality profile arises.

3. Towards an automatically constructed personality-based user profile

3.1. Personality estimation

The classification of personality types has always been among the widely addressed philosophical questions. The notion of personality traits' variance and its influence on people's behaviour was being discussed in works of Aristotle and his student Theophrastus in the fourth century BC [26].

The appearance of the scientific trait theory became possible from the beginning of the 20th century through systematic data collection and the development of statistical methods such as data correlation techniques and factor analysis. The process of personality modelling includes the construction of the basic classification dimensions and a questionnaire for measuring them.

A number of statistical approaches are used to find correlations between various traits and factor analysis techniques are applied to group positively correlated traits into larger groups. Each dimension consists of a number of traits that are related to each other and thus if a person has one of the traits in a particular dimension he is likely to have other traits from the same group. There are a number of traits' classifications that are used worldwide [26]. The *Big Five Inventory* (*BFI*)[1] has 44 questions and is widely used in research as it requires a small amount of time to be filled in and prevents the participants from becoming bored [27].

The Big Five model includes the dimensions of *Openness to experience*, *Consciousness*, *Extraversion*, *Agreeableness* and *Neuroticism* (abbreviated as *OCEAN*):

[1] http://www.ocf.berkeley.edu/~johnlab/bfi.htm

Extraversion refers to the desire of active and energetic participation in the world around. Such people are open to communications, talkative and tend to experience positive emotions. Introverts on the contrary are more focused on their own feelings and do not need so much external stimulation which leads to the comfortableness of being alone.

Agreeable people tend to eagerly cooperate with others and are generally seen as helpful and generous compared to people with disagreeable behaviour that includes self-interest and sometimes even unfriendliness. More generally, agreeable people are often unable to make tough decisions because they tend to care about other people's interests more.

Conscientiousness is seen as the ability to control impulses and to hold to long-term plans as well as being able to foresee the consequences of one's behaviour. Such people are usually perceived as intelligent and wise. In extreme cases it could lead to perfectionism and tendency to become a workaholic. Unconscientious people tend to enjoy things that bring immediate satisfaction and perceived as are spontaneous, joyful, impulsive and unreliable.

Neuroticism is positively correlated with the susceptibility to experiencing negative feelings such as anxiety, anger and depression. Neurotics respond very emotionally and tend to perceive each situation as threatening which causes an inability to think clearly and make right decisions under stress. Low levels of neuroticism usually mean emotional stability and calmness as well as low exposure to negative thoughts.

Openness to experience represents the tendency of the person to be sensitive to new ideas, non-conventional thinking and to being intellectually curious. It also includes the ability of symbolic thinking on the high level of abstraction. When scored low on this trait, people tend to have more common interests with which they are more familiar and do not like complex ambiguous things.

Slight variations present in the naming of the five factors in the Big Five model and some critics argue [26] that the above mentioned number of traits is not sufficient to reflect all the diversity of existing traits while others suppose that five could be further reduced to a smaller amount of traits. Nevertheless, the Big Five personality traits classification is one of the most widely used and recognised [26]. It covers most of the traits' classification schemes.

3.2. Personality and Social Media

The volume of social services on the Web is constantly growing, including *blogs* (a resource to express private, scientific or other means of opinions, posing questions, discussing problems, etc.), *wikis* (online text collections that allow users to write their own articles as well as edit other authors' articles; these are used for setting up digital libraries, knowledge repositories of large companies and institutions, and for developing technical documentation), *file sharing tools* (services to share things such as photos (Flickr - http://flickr.com) and bookmarks (del.icio.us - http://delicious.com) and to conveniently organise them) and *Social Networks* (services such as Facebook - http://facebook.com, Bebo - http://bebo.com, Twitter - http://twitter.com, etc. that offer a space for the user to organise personal information, maintain existing relationships and find new friends). The data available from the above mentioned services combined with data collected through thoroughly organised laboratory studies provides a broader view of the interconnection between language and human behaviour.

Quercia et al. [25] conduct their research in the area of social media (Twitter posts in particular) to study the personality of the users. They have established connections between Twitter users and their Facebook profiles working with the Facebook application myPersonality (some of the users provided the links to their Twitter accounts) developed by the "myPersonality Research" project[2].

Oberlander and Nowson [28] have focused on classifying the personality of the authors of weblogs, based on the idiolects (individual words usage) of the people. They have been using bi- and tri-grams to extract the personality from the text (all nouns were tagged via CLAWS). The authors have found that even the small set of extracted features shows the correlation between the features and particular personality traits.

Celli [29] has proposed an unsupervised approach to personality estimation based on linguistic cues. He analysed the data of FriendFeed (http://friendfeed.com) - the popular Italian social network - in order to construct the personality model of the authors, selecting 22 features mentioned in Mairesse et al. [30]. The evaluation procedure involved the comparison of the scores pro-

[2] http://mypersonality.org/research/

duced for various posts of the same user. Two measures are introduced – accuracy (showing the reliability of the model) and validity (the variability of the personality type among the posts of the same author) - which are calculated in an unsupervised way. The author has found that the most common personality type on FriendFeed is an *extrovert, insecure, agreeable, organised* and *unimaginative* person. The average accuracy of the results is 0.631 and the average validity is 0.729 [29].

Golbeck et al. [31] estimated the personality of Facebook users. Each Facebook profile provides a lot of valuable information (birthday, location, the number of education places and job positions, last profile update, the date the user joined the Facebook, etc.). One of the features of the profile is the availability of "About Me" and status updates texts. These were used as a source of estimating the personality based on linguistic cues. The authors found that the Conscientiousness factor has the largest number of correlations with linguistic categories. To predict the personality, the authors considered three groups of features apart from linguistic ones. In order to predict the personality score, the M5'Rules and Gaussian Processes algorithms of the WEKA [32] tool were applied. Results have shown that the personality type based on the Facebook profile features can be predicted to within the 11% of its actual value.

3.3. Personality from the text

The ideal situation for the individual working with the PbRS is the implicit gathering of his or her personality characteristics, their automatic analysis and the provision of the list of most relevant items based on those characteristics. As the automatic and effortless approach is the preferred one, personality could be estimated from online sources that already have some of the user contributions available for analysis. Existing online communities (Facebook, TripAdvisor, Twitter, etc.) provide a large amount of such freely accessible information that can be collected and processed to produce personality scores for the individual.

With the appearance of social data on the Internet, linguists and psychologists have gained access to large corpora of texts reflecting the way people talk naturally (which typically also include information about the author such as age, gender, social status, interests, etc.). From the 1970's computerised text analysis tools have started to appear. While the first programs used to implement sophisticated algorithms were based on language variables not directly visible to users, subsequently created tools became more and more transparent.

Research has shown that there is a correlation between the Big Five dimensions and linguistic features found in texts. Tausczik and Pennebaker [33] have discovered that the use of first-person singular pronouns correlates with depression levels, while the volume of positive emotions words reveals extraversion. Mairesse et al. [30] has shown that emotional stability (the opposite of neuroticism) is correlated with the amount of swearing and anger words used by the person while agreeableness is associated with back-channelling (personality types were estimated from self-reports and observers' reports). Some of the traits were studied more thoroughly (for example, extraversion) which could be caused by a higher level of representativeness of the particular trait related linguistic cues [30].

In this paper we have made an attempt to apply the tool created by Mairesse in order to automatically construct the personality profile of the individual through the analysis of the natural style of the author (words he or she uses). This approach eliminates the subjectivity and interference that could be introduced by the user evaluating or describing content.

3.4. The TWIN System

In order to validate the personality-based user profile construction approach we have created the TWIN ("Tell me What I Need") Personality-based Intelligent Recommender System, which follows a combination of the content-based and user-based collaborative filtering approaches. We make an assumption that the "similarity" between people can be established by analysing the context of the words they are using. Accordingly, the occurrence of the particular words in the particular text reflects the personality of the author. This suggestion leads to the possibility of the text-based detection of a circle of "twin-minded" authors whose choices could be quite similar and thus could be recommended to each other.

The first step the system performs when the user logs in is the crawling of resources already created by the person. As mentioned previously, we are estimating the personality from the textual content automatically using the tool created in previous

research [30]. In this way, the individual is not obliged to explicitly provide his personal information, ratings, answer any questions, etc., which saves him time and minimises the effort. When the personality analysis is over, the information is stored in the database and used to construct and visualise the user profile. To generate the precise description of the user information an ontology is constructed based on the general vocabularies freely available on the Semantic Web [34].

The TWIN system can be utilised in two different ways: as a Recommender System or as a personality visualiser.

The main approach is to use TWIN as a Recommender System (asking to actually perform the recommendations). In order to produce recommendations, TWIN searches for the profiles of people with personality types similar to that of the target user, and creates a list of items most favored by them. The list of recommendations is further visualised for the user at the final stage of the process.

The second manner of usage of the TWIN system application involves only the analysis of the calculated personality scores (through the visual interface). The main advantage of this approach is that it can be utilised to tune and improve the performance of the underlying personality estimation algorithm (providing a visual representation of the resulting scores).

The structure of the TWIN system was described in Roshchina et al. [35], [36]. The system includes the following components: Data Processor, Personality Recogniser, Profile creator, Similarity Estimator and Results Visualiser (Fig. 1).

3.4.1. Data Preprocessor

The Data Preprocessor component retrieves the textual data written by the user from the online community service (Facebook, TripAdvisor, etc.) by the provided username. This raw data contains special characters (e.g. apostrophes, semicolons) that should be represented properly when being saved in the database and we do not take synonyms, acronyms and similar issues into account. Therefore the Data Preprocessor performs the task of replacing the above mentioned symbols with the corresponding html characters.

3.4.2. Personality Recogniser

To calculate the personality scores from the user-generated text the User Data Processor utilises the Personality Recogniser tool [30]. The Personality Recogniser is based on the Linguistic Inquiry and Word Count (LIWC) dictionary [33] and on the Medical Research Council (MRC)[3] categories dictionary.

The Personality Recogniser processes the text word by word getting the category of each word and calculating the overall percentage of each of the discovered categories. Five scores (corresponding to each of the Big Five dimensions) are produced for the processed text - each of them ranging from 1 to 7 (where 1 means that the trait is weekly expressed and 7 means strong expressiveness).

3.4.3. Profile Creator

The common way to save information about people and to model their identity within Recommender Systems is to create User Profiles. These profiles can be knowledge-based (if person's details are acquired through questionnaires) or behaviour-based (extracted by means of various natural language processing techniques) [16]. Here we follow the behaviour-based approach, retrieving the profile data implicitly through the analysis of the text written by the particular person.

We applied the principles of the Semantic Web and reuse existing ontologies to construct the user model of the TWIN system. This format allows the sharing of the information in a meaningful way between various applications and such a choice is a step towards a RS in that applies the ontology-based knowledge representation approach proposed in recent research [14].

3.4.4. Similarity Estimator

The goal of the Similarity Estimator is to search for similar-typed people among all of the users profiles within the system, based on the assigned personality scores. The final recommendations are calculated considering the items liked by the community of discovered people.

As the k-nearest neighbour approach (kNN) shows high performance [37] and is relatively easy to apply for the RS, we have chosen to implement it for the Similarity Estimator component. The WEKA kNN algorithm calculates the value of the distance function in order to select k-nearest neighbours with personality profiles similar to the current user, in order to identify the circle of like-minded people. To estimate the similarity of user

[3]http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc. htm

profiles, we have chosen the Euclidian distance illustrated in Eq. (1) [38] as it is most commonly used in personality recognition tasks ([39],[40], [41]):

(1)

where n is the number of attributes (in our case the maximum was 5) and a_n are attribute values.

3.4.5. Results Visualiser

The Results Visualiser is constructed as a Flash application to represent the results of the recommendation for the user. It requires the user profile name of the particular social media site to retrieve the textual information written by the user. After the calculation of the personality performed by the Profile creator, its visual representation is provided by the Results Visualiser. Therefore the user can analyse his personality scores and compare them to other users' scores.

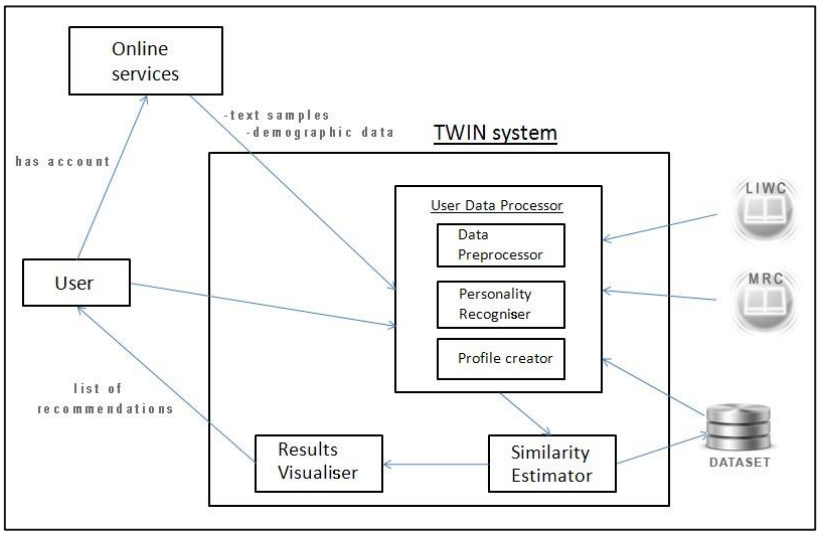

Fig. 1. TWIN system architecture

4. The TWIN System in the Online Travelling Domain

In order to evaluate the performance of the TWIN system, we have decided to apply it in the travelling domain as recently social sites such as TripAdvisor have started to emerge to allow their users to publish reviews of the places to which they have travelled. TripAdvisor provides the interface to search through the travel facilities (hotels, restaurants, etc.), check their availability for a specific date and read the reviews associated with them.

As the volume of the available reviews is growing in size every day, it is impractical for users to manually retrieve and consider each review. This is where the necessity of constructing a Recommender System appears to provide automatic filtering of relevant touristic places through the analysis of reviews texts and travellers profiles. In order to establish the similarity between people we have investigated the possibility of constructing a user profile by modelling the user's personality (according to the Big Five model) based on linguistic cues collected from the user-generated text of the reviews.

The TWIN system is utilised to visualize a list of hotels for the user on the map based on the destination of interest and the personality similarity between the particular user and other users of the system. The user profile is constructed to reflect the personality calculated from the reviews the user has written, crawled from the TripAdvisor site, and personality scores are stored in the database following the constructed user ontology.

4.1. TripAdvisor dataset

We built a Java crawler and constructed a dataset based on reviews submitted to the TripAdvisor website. The TripAdvisor site provides a large variety of user-generated content. For the purposes of our research we have utilised the dataset that contains information about hotels, reviews and people [35], updated in 2013[4]. Table shows the available fields included in the dataset.

Table 1
TripAdvisor dataset fields

Field Name	Description
Title	Title of the review
hotelRating	The overall hotel rating
content	Textual content of the review
value_rating	Hotel value rating
rooms_rating	Hotel rooms rating
location_rating	Hotel location rating
cleanliness	Hotel cleanliness rating
service_rating	Hotel service rating
sleep_rating	Hotel sleep rating

We processed review data with the Personality Recogniser tool to produce the Big Five scores for each of the reviews texts. The overview of the dataset is presented in Table 2.

Table 2
TripAdvisor dataset parameters

Parameter	Value
Number of reviews	14,000
Number of people	1,030
Average number of reviews per person	13.8
Minimum number of reviews per person	5
Maximum number of reviews per person	40
Number of words in all reviews	2.9 million
Average number of words per review	210.8
Average number of words per sentence	16.6
Minimum number of words per sentence	3
Maximum number of words per sentence	39.7

We found that scores for each of the Big Five dimensions have close to normal distributions with the highest standard deviation for the Consciousness trait and the lowest for the Neuroticism trait. This shows that people in the constructed dataset differ by Consciousness parameter more compared to the differences in the other four traits while Neuroticism scores are less variable across the dataset. Neuroticism scores also have the smallest maximum and mean values comparing to scores of other traits. Openness to experience trait has the largest mean value – this fact is interesting considering that the processed dataset consists of travellers' reviews showing the tendency for those people to be more open-minded and curious to have more diverse life experiences.

4.2. User profile construction

To model the user's personality, we store the mean score of each of the Big Five parameters calculated from the text of each of the reviews. We randomly selected 15 people from our dataset who contributed more than 30 reviews. Using the Personality Recogniser we obtained personality scores for each of the texts written by each individual. As each score is calculated from the text of the review independently we analysed them separately. The visualised scores per each of the Big Five dimensions are presented in Figures 2-6. Each grey dot represents a value of the particular trait score per each review of the current person. Dots with a "+" symbol in the middle show the mean score of all the reviews of the person. Line that joins all of such dots helps to see more clearly the variability in the mean scores.

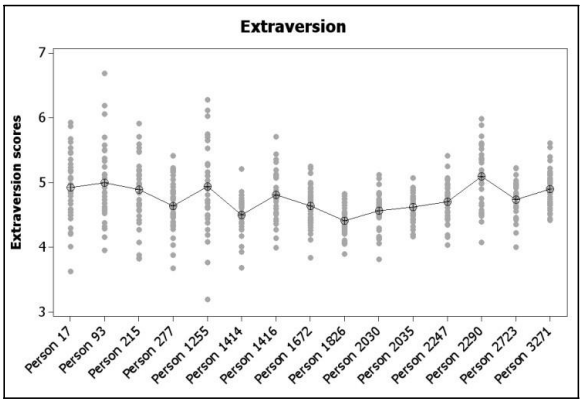

Fig. 2. Extraversion scores distribution with means for each review set

[4] http://twin-persona.org/resources.jsp

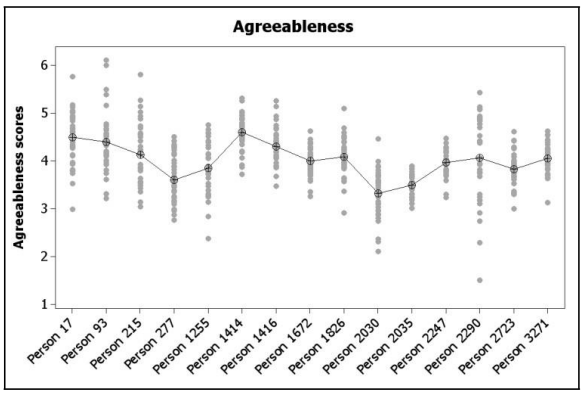

Fig. 3. Agreeableness scores distribution with means for each review set

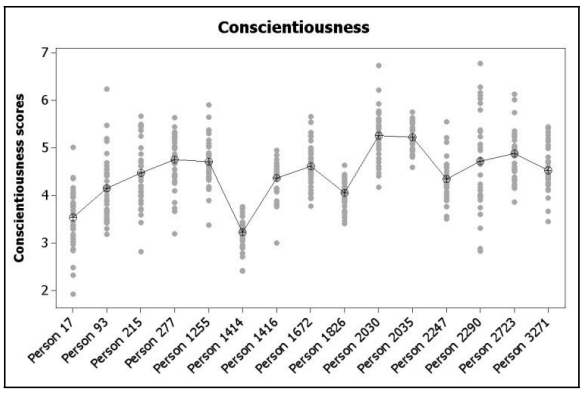

Fig. 4. Conscientiousness scores distribution with means for each review set

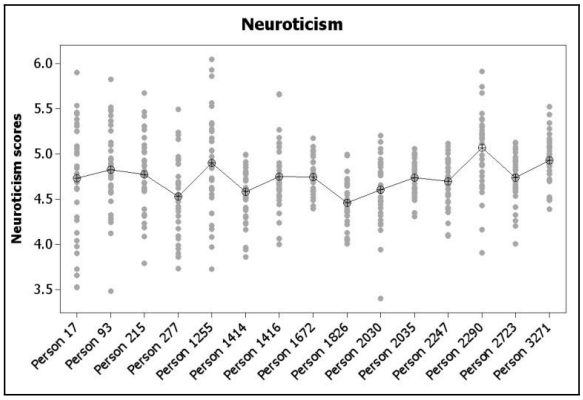

Fig. 5. Neuroticism scores distribution with means for each review set

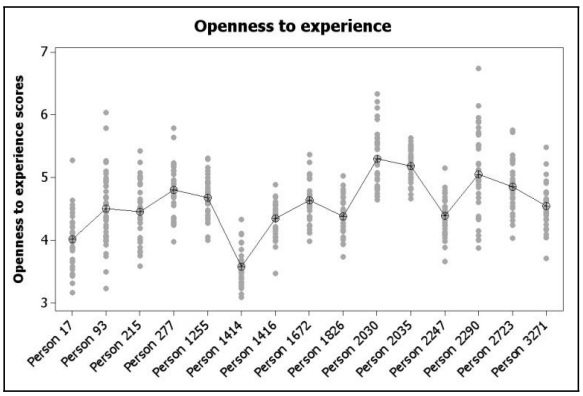

Fig. 6. Openness to experience scores distribution with means for each review set

In order to check whether the means of personality scores per person differ from each other we have performed the ANOVA (Analysis Of Variance) test. Normally when there are only two samples the standard t-test is applied but here we were comparing the variance of mean values of 15 samples representing 15 different people. The test has shown significant differences ($p < 0.001$) between persons in each of the Big Five categories. Thus it can be concluded that mean scores vary sufficiently from one person to another showing different personality patterns for each person. This fact enables us to use the mean score as the estimation of the personality in each of the 5 dimensions.

It can be seen that openness to experience scores have the highest variability in means which suggests that this trait may be the easiest to detect. This result is in agreement with Mairesse et al. [30] who had also found that openness to experience is the easiest trait to model.

4.3. TWIN System performance evaluation

In order to evaluate the performance of the kNN algorithm that forms the basis of the Similarity Estimator component, we designed the following experiment.

We selected 26 people from the TripAdvisor dataset who have contributed more than 35 reviews. We split the data into two parts: training and test sets. For the test set, we selected 5 reviews from the list of reviews of each person in the training set (at the same time, deleting those reviews from the training set). The task of the system was to identify the circle of reviews (of like-minded authors) that are similar to the given review. We experimented with two types of

training sets: with reviews' scores (821 instances) viewing each review as a single personality type and with mean vectors of reviews' scores per person (26 instances) that represent the overall personality of the author.

As we found that different traits of the Big Five have different levels of estimation complexity, we experimented with all 31 combinations of the Big Five parameters to feed the kNN algorithm in order to select the combination that would produce better results. The results are summarised in Figure 7.

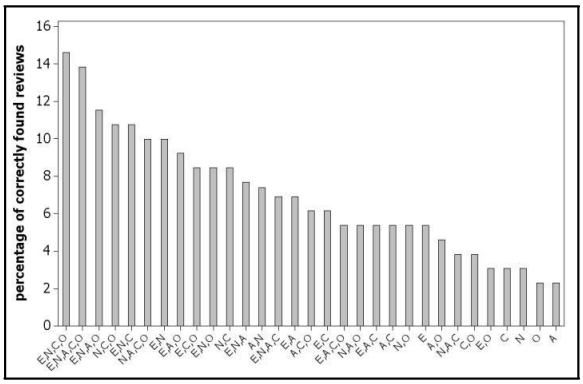

Fig. 7: The percentage of correctly found reviews considering plain personality scores.

Although the results of the classification are not very optimistic, we consider that this demonstrates the complexity of the task. Similar results were obtained by Mairesse [30] for the regression classification task. He stated that "the improvements seem relatively small" (over the baseline) explaining that the essence of the regression algorithm requires "the association of an exact scalar value with each individual". The results we obtained can be considered promising taking into account the difficulty of the personality from the text estimation on real-world data. One of the key points in personality score calculation is the recognition of sentence and word boundaries which are not always defined strictly in reviews dataset (missing punctuation marks, misspelled words, etc.). Thus, one of the approaches to improve the results is to incorporate additional information available in the TripAdvisor user profile, e.g. age, gender, location, etc., as well as the experimentation with the personality from the text recognition algorithm itself.

It can be concluded that the kNN algorithm performs better when considering the mean vectors of reviews' scores that represent the overall personality of the author. Therefore it was reasonable to select these mean vectors to be saved as the user profile information in the TWIN system. It can be seen that not all combinations of the Big Five parameters produce the same results. Figure 8 shows that the combination "Extraversion-Neuroticism-Consciousness-Openness to Experience" performs better than any other combination, with the combination of all the five traits occupying second place.

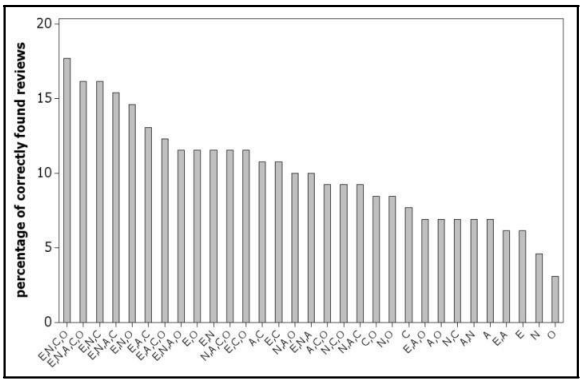

Fig. 8: The percentage of correctly found reviews considering plain personality scores.

5. Conclusions

In this paper we have investigated the challenging task of personality profile construction based on the automatic estimation of personality from the text. We have constructed the TWIN Personality-based Intelligent Recommender System and performed a number of experiments in the online travelling domain by utilising the TripAdvisor dataset.

In order to validate the methods forming the base of the TWIN system we have constructed a TripAdvisor dataset consisting of 14,000 reviews written by 1,030 people. We calculated the personality scores using the Personality Recogniser tool.

One of the major tasks of the Recommender System construction is the development of the user profile [36]. In order to find the appropriate structure for the profile we have chosen 15 people from the TripAdvisor dataset who contributed more than 35 reviews. We have performed the ANOVA test to see whether the personality scores pattern varies sufficiently from one person to another. As the answer was positive we decided to use the mean score of all the reviews (for each of the Big Five

parameters) written by the particular person as the estimator of his personality.

In order to evaluate the Similarity Estimator – the main component of the TWIN system that provides the Recommender System functionality – we have focused on the scores of each individual review [35]. With the 26 people chosen, we found that the percentage of the correctly identified reviews was 10% on average. Considering the difficulty of the task of the personality from the text recognition and especially the fact that the research was based on the real world data from the TripAdvisor, we concluded that the results are satisfactory [35].

There are a number of experiments that we consider to investigate in the future research. As the algorithm implemented in the Personality Recogniser demonstrate how challenging the task is, we are designing experiments with some of the approaches presented in [42] and at the Workshop on Computational Personality Recognition 2014[5]. In order to compare the performance of available algorithms we plan to apply convergence analysis. It will allow us to track how fast the computation of the personality values approaches the real scores when adding one review of the same author at each step in time.

Since the experiments reported in [43] and [35] we have performed a number of code refactoring tasks for the TWIN system. The new version of the system will soon be available online[6].

Acknowledgements

The research work of the third author is partially funded by the WIQ-EI (IRSES grant n. 269180) and DIANA APPLICATIONS (TIN2012-38603-C02-01), and done in the framework of the VLC/Campus Microcluster on Multimodal Interaction in Intelligent Systems.

References

[1] R. R. McCrae and P. T. Costa, "Toward a new generation of personality theories: Theoretical contexts for the five-factor model," *FiveFactor Model Personal. Theor. Percpetives*, pp. 51–87, 1996.

[2] F. Ricci, L. Rikach, B. Shapira, and P. Kantor, *Recommender Systems Handbook*. US: Springer, 2010, p. 62.

[3] A. V Bodapati, "Recommendation Systems with Purchase Data," *J. Mark. Res.*, vol. 45, no. 1, pp. 77–93, Feb. 2008.

[4] V. B and S. Kant, "Analytical Appraisal of Recommender Systems in E-Commerce," *Int. J. Comput. Sci. Inf. Technol. Secur.*, vol. 2, no. 4, pp. 747–752, 2012.

[5] J. Gemmell, T. Schimoler, M. Ramezani, L. Christiansen, and B. Mobasher, "Improving folkrank with item-based collaborative filtering," *CEUR Workshop Proc.*, vol. 532, pp. 17–24, 2009.

[6] J. Dean and S. Ghemawat, "MapReduce□: Simplified Data Processing on Large Clusters," *Commun. ACM*, vol. 51, no. 1, p. 107, Jan. 2008.

[7] R. Kumar, B. K. Verma, and S. S. Rastogi, "Social Popularity based SVD++ Recommender System," *Int. J. Comput. Appl.*, vol. 87, no. 14, pp. 33–37, 2014.

[8] L. Babak, Y. Shi, L. Martha, and A. Hanjalic, "Cross-Domain Collaborative Filtering with Factorization Machines," in *Advances in Information Retrieval*, Springer International Publishing, 2014, pp. 656–661.

[9] T. Olsson, "Bootstrapping and Decentralizing Recommender Systems," 2003.

[10] R. Nageswara and K. Talwar, "Application domain and functional classification of recommender systems a survey," *DESIDOC J. Libr. Inf. Technol.*, vol. 28, no. 3, pp. 17–35, 2008.

[11] G. Semeraro, "Content-based Recommender Systems: problems, challenges and research directions," in *8th Workshop on Intelligent Techniques for Web Personalization & Recommender Systems*, 2010.

[12] J. Castro, R. M. Rodriguez, and M. J. Barranco, "Weighting of Features in Content-Based Filtering with Entropy and Dependence Measures," *Int. J. Comput. Intell. Syst.*, vol. 7, no. 1, pp. 80–89, Jan. 2014.

[13] A. Mathes, "Folksonomies - Cooperative Classification and Communication Through Shared Metadata," 2004.

[14] I. Cantador, A. Bellogín, and D. Vallet, "Content-based recommendation in social tagging systems," in *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10*, 2010, p. 237.

[15] S. Huang, "Designing utility-based recommender systems for e-commerce: Evaluation of preference-elicitation methods," *Electron. Commer. Res. Appl.*, vol. 10, no. 4, pp. 398–407, Jul. 2011.

[16] M. A. S. . Nunes, "Towards to Psychological-based Recommenders Systems: A survey on Recommender Systems," *Sci. Plena*, vol. 6, no. 8, p. 1028, 2010.

[17] G. González, J. L. De La Rosa, M. Montaner, and S. Delfin, "Embedding emotional context in recommender systems," in *Proceedings - International Conference on Data Engineering*, 2007, pp. 845–852.

[18] J. Masthoff, "The pursuit of satisfaction: Affective state in group recommender systems," *User Model. 2005*, pp. 152–152, 2005.

[19] T. Saari, N. Ravaja, J. Laarni, and M. Turpeinen, "Towards Emotionally Adapted Games based on User Controlled Emotion Knobs," *DiGRA 2005 Conf. Chang. ViewsWorlds Play*, 2005.

[20] M. Tkalčič, U. Burnik, and A. Košir, "Using affective parameters in a content-based recommender system for images," *User Model. User-adapt. Interact.*, vol. 20, no. 4, pp. 279–311, Sep. 2010.

[21] R. Hu and P. Pu, *User Modeling, Adaptation, and Personalization*, vol. 6075. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 291–302.

[22] P. J. Rentfrow and S. D. Gosling, "The do re mi's of everyday life: The structure and personality correlates of

[5] https://sites.google.com/site/wcprst/home/wcpr14

[6] http://twin-persona.org/twin.jsp

music preferences.," *J. Pers. Soc. Psychol.*, vol. 84, no. 6, pp. 1236–1256, 2003.
[23] M. Elahi, M. Braunhofer, F. Ricci, and M. Tkalčič, "Personality-Based Active Learning for Collaborative Filtering Recommender Systems," *AI*IA 2013 Adv. Artif. Intell.*, vol. 8249, pp. 360–371, 2013.
[24] M. Tkalcic, A. Odic, A. Kosir, and J. Tasic, "Affective Labeling in a Content-Based Recommender System for Images," *IEEE Trans. Multimed.*, vol. 15, no. 2, pp. 391–400, Feb. 2013.
[25] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our Twitter Profiles, Our Selves: Predicting Personality with Twitter," in *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, 2011, vol. 2, no. 6, pp. 180–185.
[26] G. Matthews, I. J. Deary, and M. C. Whiteman, *Personality Traits*, Cambridge . Cambridge, UK, 2009, pp. 23–26.
[27] O. P. John, L. P. Naumann, and C. J. Soto, "Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues.," *Handb. Personal. Theory Res.*, pp. 114–158, 2008.
[28] J. Oberlander and S. Nowson, "Whose thumb is it anyway? Classifying author personality from weblog text," *Comput. Linguist.*, pp. 627–634, 2006.
[29] F. Celli, "Unsupervised Personality Recognition for Social Network Sites," *ICDS 2012 sixth Int. Conf. Digit. Soc.*, no. c, pp. 59–62, 2012.
[30] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text," *J. Artif. Intell. Res.*, vol. 30, no. 1, pp. 457–500, 2007.
[31] J. Golbeck, C. Robles, and K. Turner, "Predicting personality with social media," in Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11, 2011, p. 253.
[32] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.
[33] Y. R. Tausczik and J. W. Pennebaker, "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods," *J. Lang. Soc. Psychol.*, vol. 29, no. 1, pp. 24–54, Dec. 2009.
[34] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Sci. Am.*, pp. 29–37, 2001.
[35] A. Roshchina, J. Cardiff, and R. Paolo, "Evaluating the Similarity Estimator Component of the TWIN Personality-based Recommender System," in LREC 2012 - *8th International Conference on Language Resources and Evaluation*, Istanbul, 2012.
[36] A. Roshchina, J. Cardiff, and R. Paolo, "User Profile Construction in the TWIN Personality-based Recommender System," *IJCNLP Workshop on Sentiment Analysis where AI meets Psychology*, SAAIP-2011, 5th Int. Joint Conf. on Natural Language Processing, Chiang Mai, Thailand, 2011.
[37] M. J. Islam, Q. M. J. Wu, M. Ahmadi, and M. A. Sid-Ahmed, "Investigating the performance of Naive- Bayes classifiers and K- nearest neighbor classifiers," in *2007 International Conference on Convergence Information Technology, ICCIT 2007*, 2007, pp. 1541–1546.
[38] I. H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques. 2005, p. 560.
[39] M. Tkalčič, M. Kunaver, J. Tasič, and A. Košir, "Personality Based User Similarity Measure for a Collaborative Recommender System," *5th Work. Emot. HumanComputer Interact. World Challenges*, p. 30, 2009.
[40] A. Tropchenko and A. Tropchenko, "Neural Network Method for Person's Personality Recognition on the Face Image," *TEM J.*, vol. 1, no. 4, 2012.
[41] I. Cantador, I. Fernández-tobías, and R. Bellogín, "Relating Personality Types with User Preferences in Multiple Entertainment Domains," in *UMAP Workshops*, 2013.
[42] F. Celli, F. Pianesi, D. Stillwell, and M. Kosinski, "Workshop on Computational Personality Recognition: Shared Task," 2013.
[43] A. Roshchina, J. Cardiff, and R. Paolo, "A Comparative Evaluation of Personality Algorithms for the TWIN Recommender System," CIKM 3rd Int. *Workshop on Search and Mining User-generated Contents*, SMUC-2011, UK. 2011.