A Comparative Evaluation of Personality Estimation Algorithms for the TWIN Recommender System

Alexandra Roshchina
Social Media Research Group
ITT Dublin, Ireland
sasharo@itnet.ie

John Cardiff
Social Media Research Group
ITT Dublin, Ireland
John.Cardiff@ittdublin.ie

Paolo Rosso
NLE Lab-ELiRF, Universidad Politécnica de Valencia
Spain
prosso@dsic.upv.es

ABSTRACT
The appearance of the so-called recommender systems has led to the possibility of reducing the information overload experienced by individuals searching among online resources. One of the areas of application of recommender systems is the online tourism domain where sites like TripAdvisor allow people to post reviews of various hotels to help others make a good choice when planning their trip. As the number of such reviews grows in size every day, clearly it is impractical for the individual to go through all of them. We propose the TWIN ("Tell me What I Need") Personality-based Recommender System that analyzes the textual content of the reviews and estimates the personality of the user according to the Big Five model to suggest the reviews written by "twin-minded" people. In this paper we compare a number of algorithms to select the better option for personality estimation in the task of user profile construction.

Categories and Subject Descriptors
I.2.7 [**Artificial Intelligence**]: Natural Language Processing - *language parsing and understanding, text analysis.*

General Terms
Languages, algorithms.

Keywords
Recommender system, personality-based profile, natural language processing, personality from the text.

1. INTRODUCTION
With the transformation of the Web from the provision of "brochure-like" access to the information owned and edited by companies and authorities, global net developers have realized the importance of the previously unnoticed individual opinions of Web content readers and the impact they could make on the process of information dissemination. Due to the appearance of the more intuitive tools of content manipulation and administration, the model of the person's interaction with the Web has changed. The user and her interests and needs have become the main starting point of all actions, held by online intelligent applications appearing elsewhere. By exploiting the "wisdom of the crowds" [14] such applications make use of the collective intelligence to get the broader view over the particular area of knowledge.

In the real world the person is surrounded by other people almost all the time so there is always a potential possibility to obtain help. And when going online the process is imitated by special types of intelligent Web services known as recommender systems [12]. Such systems automatically provide expert opinions for people who tend to rely on them when choosing from the large number of alternatives (e.g. Movies2Go[1], Last.fm[2], etc.).

In particular people are faced with multiple choices when trying to select a place to stay for travelling purposes. Some social sites such as TripAdvisor[3] try to make this task easier through the possibility of online reviewing of hotels, restaurants, etc. by peer travellers. As the users' interest in social sites is rapidly growing the importance of such user provided travel related content has become obvious to travel marketers. It has proved to be useful for promotion purposes, creation of the positive company and destination image, etc [19]. However the rapidly increasing amount of information comprising users' opinions and assessments over the various travelling destinations prevents users from considering of all the available variants and becomes a separate problem of appropriate knowledge extraction from available reviews data that can be solved by means of the recommender system.

One of the main tasks in the construction of the recommender system is to collect the information that will represent the user and store it into the user profile. The general approach is to track user's behavior, collect his preferences and make content predictions based on that knowledge. In the majority of such systems the user profile generation is based on the explicit personal facts provided at the registration step (such as name, age, etc.) and on the constant updates of the person's interactions with the system.

In order to perform TripAdvisor hotel recommendations we have proposed the "Tell me What I Need" (TWIN) recommender system [13]. We make an assumption that people having similar personality types would prefer to choose the same hotels as their twin-minded peer travellers. As significant correlations were found between words that people use [11] and specific personality characteristics of the authors (here we consider the Big Five personality model) there is a possibility of creating a personality-based user profile that estimates personality traits of the particular individual from the text written by him.

In this paper we compare the performance of 4 algorithms for estimating the author's personality from the text, using a dataset of user-generated reviews crawled from the TripAdvisor [13]. They are provided by the Personality Recognizer tool utilized by the TWIN system and include linear regression , M5' model tree, M5' regression tree and support vector machines. Our goal is to

[1] http://www.movies2go.net

[2] http://www.last.fm

[3] http://www.tripadvisor.com

determine an approach which will provide better results for the purposes of personality-based user profile construction.

The paper is organized as follows. In Section 2 we provide the psychological background of the estimation of the personality from the text and the description of the algorithms applicable. We proceed with a brief overview of recommender systems. In Section 3 we give an overview of the TWIN system. In section 4 we present the comparison of the results of 4 personality estimation algorithms on the TripAdvisor dataset. In Section 5 we draw our conclusions.

2. BACKGROUND

2.1 Personality traits

The classification of people according to their personality types has always been one of the widely discussed psychological topics. Traditionally the process of personality modeling includes the construction of the basic classification dimensions and the questionnaire for measuring them. A number of statistical approaches are used to find correlations between various traits and then factor analysis techniques are applied to group positively correlated traits into larger groups [6].

2.1.1 Personality traits classification

A number of different classification schemes were constructed through the psychological research. They differ in the amount of basic traits (or factors) allocated and the area of application. The most widely used schemes are [6]:

1. *16PF* (Sixteen Personality Factor Questionnaire). For a long time this scheme was used for research purposes. More recently its consistency has been criticized [6];
2. *CPI* (California Psychological Inventory). Takes 20 traits into account and is applied in the industry;
3. *OPQ* (Occupational Personality Questionnaire). This scheme has 31 traits and is used for personnel recruitment purposes;
4. *EPQ-R* (Eysenck Personality Questionnaire-Revised). The scheme considers 3 factors: neuroticism, extraversion and psychoticism;
5. *NEO-PI-R* (NEO-Personality Inventory-Revised). It is also known as “The Big Five” model. It considers 5 factors and the questionnaire consists of 240 questions. This scheme is most widely used for research purposes. It also has applications in the industry for staff recruitment purposes.

2.1.2 Big Five model

The Big Five model is seen as the most recognized in the psychology research personality construction instrument [6]. There is a general agreement in the number of factors to take into consideration but the actual dimensions can vary from one model to another. Most of the empirical research was done through the application of the model of Costa and McCrae [7] including the dimensions of openness to experience, consciousness, extraversion, agreeableness and neuroticism (abbreviated as OCEAN).

Openness to experience

This trait 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 openness to experience, people tend to have more common interests with which they are more familiar and tend not like complex ambiguous things.

Consciousness

Consciousness is seen as the ability to control impulses and to hold to the 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. At the extreme it could lead to perfectionism and tendency to become a workaholic. A high score suggests people who tend to enjoy things that bring immediate satisfaction and perceived as spontaneous, joyful, impulsive and unreliable.

Agreeableness

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

Extraversion

Extraversion refers to the desire of active and energetic participation in the world around them. 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 lead to the comfortableness of being alone.

Neuroticism

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

2.2 Recommender Systems

The general model of the recommendation process involves a Recommendation Seeker, a person who needs to make a selection between a number of possible alternatives in the Universe of Alternatives (see Figure 1) [16]. When the person is not an expert in the field or the amount of different choices is too large he seeks recommendations provided by the specific recommender system.

2.2.1 User profile

Recommender systems collect the information about the particular user either explicitly (asking individuals to provide their preferences) or implicitly (automatically analyzing the data or the activity of the user) and store it in the Preference Provider (or simply, the user profile). The appearance of various social services on the web brought the possibility of gathering a large

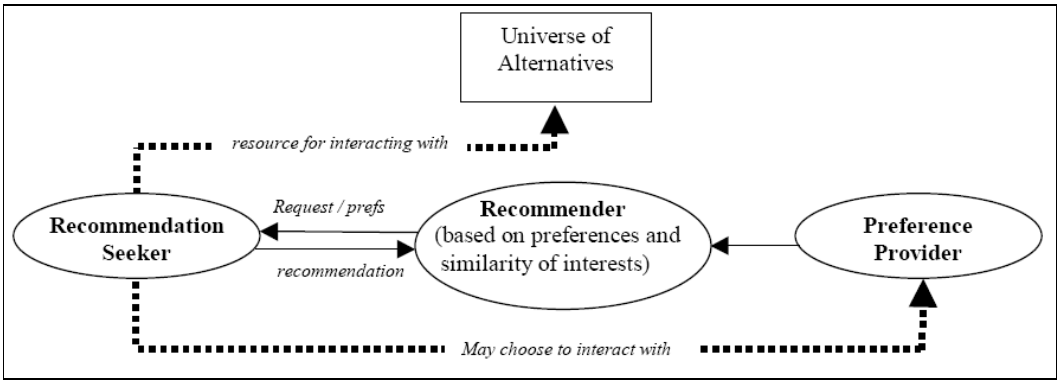

Figure 1. The process of recommendation

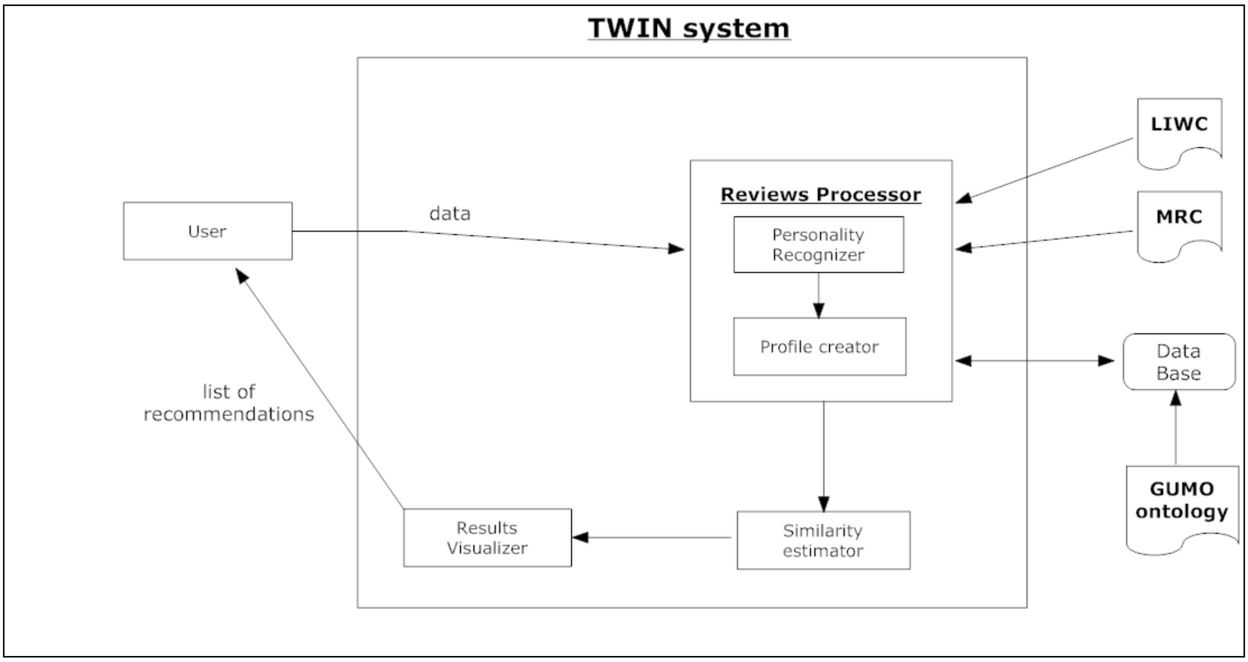

Figure 2. TWIN system architecture

amount of "outside world" information in order to describe the user from different points of view. For example, Ghosh and Dekhil [3] propose to describe the user through gathering the data from a number of sources: Google Calendar[1], FOAF[2] (the Friend Of a Friend project), etc. and to cover the diversity of formats in the data through the construction of a Retail User Profile Ontology to store semantically enriched profile entries. The user profile thus is described using RDF triples and is accessible via SPARQL queries, creating an easily extensible approach to effectively deal with semi-structured and sparse data. Diederich and Iofciu [2] describe the user based on the keywords (viewed as tags) crawled from the DBLP[3] data set publications. But recently there appeared an interest in the personality-based user profile. Nunes [10] proposes to construct it by means of personality questionnaires to explicitly retrieve the information about the user.

[1] https://www.google.com/calendar

[2] http://www.foaf-project.org/

[3] http://www.informatik.uni-trier.de/~ley/db/

With all the gathered data the system is able to provide recommendations of people with similar (using various existing similarity measures, e.g. cosine similarity) profiles or artifacts produced by them (content, etc.) to a specific person (*collaborative filtering recommender systems*). Initially the activity of other people was not taken into account and the user profile information was used for purely filtering purposes to find items with matching characteristics (*content-based recommender systems*). Nowadays those two approaches are often combined to create systems that utilize both of the above-mentioned approaches (*hybrid recommender systems*) [9].

3. TWIN SYSTEM

In [13] we have proposed the prototype of the TWIN Personality-based Recommender System. We make an assumption that the "similarity" between people can be established by analyzing 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 of particular places to stay (hotels reviewed in TripAdvisor, in our case) could be quite similar and thus could be recommended to each other. This approach provides recommendations that rely on the factors independent in many ways from the user's preexisting attitudes in the hotels market and also avoids the subjective step of specifying explicit preferences.

3.1 Architecture

The diagram in Figure 2 represents the main components of the proposed TWIN recommender system described below.

Reviews Processor

The Reviews Processor component retrieves the textual data from the user (plain text written by person) and does the text preprocessing step (dealing with special characters, etc.). It consists of the *Personality Recognizer* (a tool for the estimation of personality scores described in more detail in Section 3.2) and a *Profile creator* that stores the general information about the user (login, age group, etc.) as well as personality scores in the user profile that follows the GUMO ontology (General User Model Ontology) [4]. This ontology provides a way of extensive description of the user and is a part of the framework that realizes the concept of ubiquitous user modelling. It includes demographic information, psychological state and a lot of other aspects. It has appropriate classes to represent the Big Five model personality parameters as well as general user data (age, gender, etc.).

Similarity Estimator

The Similarity Estimator component utilizes the Weka[1] clustering model built using the K-Means algorithm [18]. During the recommendation process the above-mentioned model is assigning the person to the appropriate cluster based on her profile information. Recommendations are calculated considering the items liked by people in this estimated cluster.

Results Visualizer

The Results Visualizer is constructed as a web-based Flash application to represent the results of the recommendation for the user, i.e. the list of hotels. The resulting list of hotels is depicted on the Google Map[2].

3.2 Personality from the Text

With the appearance of social data on the Internet (live journals, blog entries, etc.) linguists and psychologists have gained access to large corpora of texts reflecting the way people talk naturally. This data combined with the one collected through thoroughly organized laboratory studies provides a broader view of the interconnection between language and human behaviour.

Most of the research on personality recognition from the text has been done through utilizing the Big Five model. It has been found that there exist correlations between the linguistic features and the Big Five personality types of the authors [5][15].

3.2.1 Personality Recognizer algorithms

One of the tools available for the personality from the text construction is the Personality Recognizer [5]. It is based on the dictionary created for the Linguistic Inquiry and Word Count (LIWC[3]) program (text analysis tool) [15] and MRC[4] (Medical Research Council) Psycholinguistic database dictionary.

At the moment LIWC contains more than 80 predefined categories including:

1. General descriptor categories (word count, words per sentence, etc.)
2. Standard linguistic dimensions (pronouns, adverbs, prepositions, etc.)
3. Words reflecting psychological processes (social processes, cognitive processes, etc.)
4. Personal concern categories (work, leisure, money, etc.)
5. Paralinguistic dimensions (assents, fillers, etc.)
6. Punctuational categories

The Personality Recognizer processes the text word by word getting the category of each word and calculating the overall percentage of each of the discovered categories. In order to establish the personality of the author the Personality Recognizer applies Weka models trained on the psychology essays corpora (comprising texts and personality scores of the authors collected through the Big Five questionnaire [11]). We considered 4 data mining algorithms available in the program: linear regression, M5' model tree, M5' regression tree and support vector machines for regression.

Linear Regression

One of the basic algorithms for mining the data with numeric attributes is linear regression. It provides the linear combination of all the attribute values with their weights estimated from the training data in order to represent a particular class [18]:

$$x = w_0 + w_1a_1 + w_2a_2 + ... + w_ka_k$$

where w_k is a weight, a_k is the value of the attribute and x is a class.

Below is the example of the fragment of the openness to experience Weka model with linear regression algorithm:

*ZOPEN = 0.0008 * BROWN-FREQ + -0.009 * CONC + -0.0212 * FAM + 0 * K-F-FREQ + 0.1653 * K-F-NCATS + 0.0076 * MEANC + 0.2425 * NLET + 0.9659 * NPHON + -0.917 * NSYL + 0 * T-L-FREQ + 0.0007 * WPS + 0.1557 * Qmarks + 0.0127 * Unique + -0.0377 * Dic + -0.0168 * Sixltr + 0.0546 * Pronoun + 0.1225 * We + -0.0434 * Self + 0.0792 * You + 0.0747 * Other + 0.0668 * Article + 0.0219 * Preps + -0.0636 * Number + 0.1443 * Affect + -0.1315 * Posemo + 0.1275 * Posfeel + -0.1593 * Optim + -0.1151 * Negemo + -0.0821 * Anger + -0.0365 * Cogmech + 0.0606 * Insight + 0.0664 * Discrep + 0.1525 * Inhib + 0.0343 * Tentat + 0.0741 * Certain + 0.0564 * Senses + 0.0671 * Hear + -0.0928 * Feel + -0.066 * Social + 0.0739 * Humans + -0.0291 * Time + -0.0177 * Present + 0.0319 * Excl + -0.0701 * Achieve + 0.2034 * Swear + 0.0541 * Period + 0.0557 * Comma + 0.2532 * Colon + 0.1957 * SemiC + 0.0475 * Dash + 0.0599 * Apostro + 0.3377 * Parenth + -0.0443 * AllPct + 13.0094*

[1] http://www.cs.waikato.ac.nz/ml/weka/

[2] http://code.google.com/apis/maps/documentation/flash/

[3] http://www.liwc.net

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

where attribute values (*BROWN-FREQ*, *CONC*, etc.) are a percentage of words found under each of the LIWC and MRC categories. The choice of the specific categories for modelling of each of the Big Five dimensions is learned from the training data.

Regression Tree and Model Tree

Regression and model trees are types of general decision trees [18] that deal with numeric prediction (rather than categories prediction). A decision tree is a structure learned from the training data. Each node of the tree represents a function to compare specific attributes of the unseen instance that needs to be classified. Each leaf assigns a class (or the probability of the specific class) to the instance after all the comparisons had been made while traversing through all the nodes of the tree down to the particular leaf. Regression trees differ from ordinary decision trees in the fact that their leaves contain the average value assigned to the instances that reach that leaf while model trees have linear regression models in their leaves.

Below is the example of the fragment of the Weka consciousness model using the M5' Model tree algorithm (only the first of the leaves with linear regression model - *LM1* - is shown):

Swear <= 0.925 :

| *Pronoun <= 16.71 : LM1 (22/49.343%)*

| *Pronoun > 16.71 : LM2 (46/77.164%)*

Swear > 0.925 :

| *Sexual <= 0.615 : LM3 (14/68.303%)*

| *Sexual > 0.615 :*

| | *NSYL <= 1.137 : LM4 (4/34.268%)*

| | *NSYL > 1.137 : LM5 (10/52.965%)*

LM num: 1

CON.J =

*0.0133 * CONC*

*- 0.0222 * IMAG*

*- 0.0572 * Pronoun*

*+ 0.0658 * Posemo*

*- 0.5933 * Sad*

*+ 0.0386 * Insight*

*- 0.0387 * Senses*

*+ 0.0437 * Comm*

*- 0.1082 * Swear*

+ 8.2099

where attribute values (Pronoun, Swear, etc.) are a percentage of words found under each of the LIWC and MRC categories.

Support Vector Machines for Regression

As with the linear regression the main aim of the support vector machines algorithm is to find a function to approximate the training data points and minimize the prediction error [18]. The difference lies in the fact that the user can specify the ε parameter in order to draw a tube around the regression line (see Figure 3).

The ε parameter represents how close the line will fit the training set. All the errors (the deviations from the regression line) within the tube are ignored and only support vectors (those lying on the tube border or outside) will have an influence on the prediction error. The task of the algorithm is to find a balanced solution between a minimum prediction error and the maximum flatness of the tube.

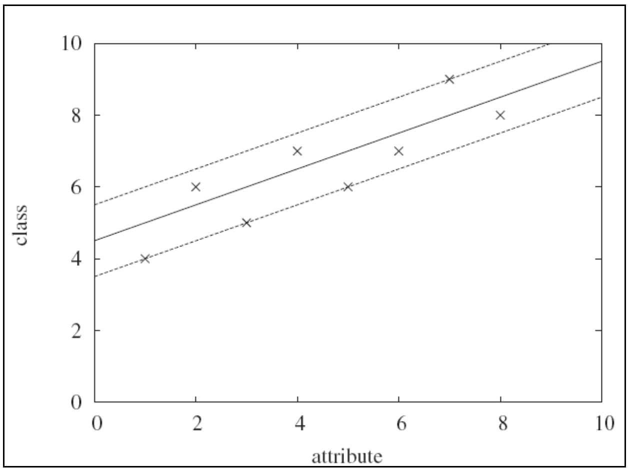

Figure 3. Support vector regression. ε = 1. Regression line for 8 data points with only one attribute considered

4. EVALUATION

The primary goal of our current work is the creation of the user profile for the TWIN system to provide personalized recommendations based on psychological aspects of the person. In [13] we have shown that it is possible to use the mean personality score (per each of the Big Five dimensions) of the texts written by the same author to construct the 5-dimentional vector to store in her user profile.

For the purposes of our experiment we have collected the reviews dataset from the TripAdvisor site [13] that consists of hotels, reviews and reviews authors. To evaluate the performance of the 4 algorithms available in the Personality Recognizer we have selected 15 people who contributed more than 30 reviews from the above mentioned dataset. We have calculated the scores of each of the Big Five parameters for each review separately applying all the 4 algorithms. To compare the performance of the algorithms we analyze standard deviations of scores of reviews written by the same person. The hypothesis we consider is that the algorithm producing scores that differ the least for the reviews of the same author will be the best to apply for the personality recognition task in the TWIN system.

We have studied the performance of the algorithms separately for each dimension as the Big Five scores are estimated independently from the text of the review. For each personality trait we have constructed 4 samples (representing each of the 4 algorithms) of standard deviations of reviews scores of 15 people. The ANOVA test [8] results have shown significant differences between the samples for almost all the Big Five parameters. Therefore, the performance of the algorithms differs for all of the traits (see Table 1).

Table 1. ANOVA test for the algorithms comparison

	Algorithms results
Openness to experience	Differ at $p < 0.05$
Consciousness	Differ at $p < 0.001$
Extraversion	Differ at $p < 0.001$
Agreeableness	Differ at $p < 0.001$
Neuroticism	Differ at $p < 0.001$

It could be conjectured that openness to experience is the easiest trait to model as the difference between the algorithms results is less significant than in the rest of the traits. This result is in agreement with the previous research [5]. On the other hand consciousness trait (see Figure 4) is likely to be the hardest to model (this was also shown in [5]) as all the algorithms show highly significant difference in the results variation.

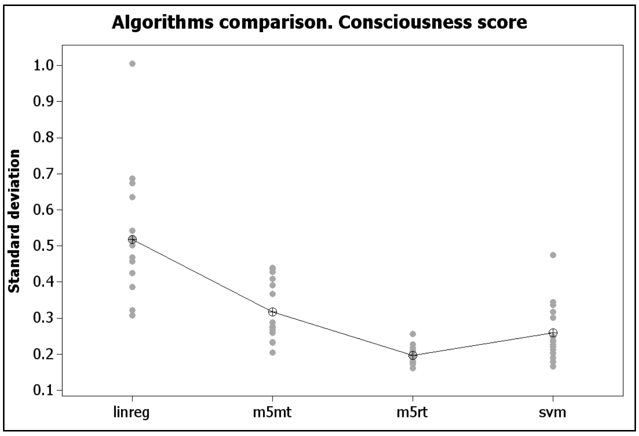

Figure 4. Algorithms results comparison. Consciousness score. (linreg - linear regression , m5mt - M5' model tree, m5rt- M5' regression tree, svm - support vector machines)

In order to find the algorithm that has shown the best results for each of the traits we have performed Fisher's Least Significant Difference (Fisher's LSD) test to carry out all pairwise t-tests [17]. The results are shown in Table 2.

Table 2. Algorithms that differ significantly

	Algorithm that differs significantly
Openness to experience	M5' regression tree (slightly significantly better)
Consciousness	Linear regression (significantly worse)
Extraversion	M5' regression tree (significantly better)
Agreeableness	M5' regression tree (slightly significantly better)
Neuroticism	M5' regression tree (significantly better)

It can be concluded that the M5' regression tree algorithm outperforms others for most of the Big Five traits on the TripAdvisor reviews dataset. Therefore, we have chosen the above mentioned algorithm for the construction of the TWIN system personality profile.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented an overview of our ongoing work on the construction of the TWIN Personality-based Recommender System. In particular we have performed the comparison of the algorithms available in the Personality Recognizer tool utilized by the TWIN. In order to compare the results we have applied those algorithms to the personality estimation of 15 randomly selected people who had contributed more than 30 hotel reviews to the TripAdvisor site.

We have hypothesized that the algorithm to be considered the best would produce scores that differ the least for the reviews of the same author, and compared results separately for each of the Big Five traits as the estimation of the personality scores from the review text is performed independently for each dimension.

Therefore, we have found that the M5' regression tree algorithm of the Personality Recognizer tool performs better on the reviews data than the other 3 algorithms available. We have decided to utilize it for the personality profile construction task in the TWIN system.

Our future work will include the application of clustering algorithms in order to estimate the percentage of people rightly assigned to a specific cluster based on the personality type constructed from the texts they had written.

6. ACKNOWLEDGEMENTS

We would like to thank James Reilly for his valuable feedback on this paper. The work of the third author has been done in the framework of the VLC/CAMPUS Microcluster on Multimodal Interaction in Intelligent Systems and it has been partially funded by the European Commission as part of the WIQEI IRSES project (grant no. 269180) within the FP 7 Marie Curie People Framework, by MICINN as part of the Text-Enterprise 2.0 project (TIN2009-13391-C04-03) within the Plan I+D+i.

7. REFERENCES

[1] Coltheart, M. 1981. *The MRC Psycholinguistic Database*. Quarterly Journal of Experimental Psychology, 33A(4):497-505.

[2] Diederich J., Iofciu T. 2006. *Finding Communities of Practice from User Profiles Based On Folksonomies.* Practice, 288-297.

[3] Ghosh R, Dekhil M. 2008. *Mashups for semantic user profiles*. Proceeding of the 17th international conference on World Wide Web 08, 1229.

[4] Heckmann D. 2005. *Ubiquitous User Modeling*. IOS Press.

[5] Mairesse F., Walker M. A., Mehl M., Moore R. 2007. *Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text*. Journal of Artificial Intelligence Research, 457-500.

[6] Matthews G., Deary I. J., Whiteman M. C. 2009. *Personality Traits*. Cambridge University Press, Cambridge, UK.

[7] McCrae, R. R., Costa, P. 1996. *Toward a new generation of personality theories: Theoretical contexts for the five-factor model*. The Fivefactor Model of Personality, 51-87. Guilford, New York.

[8] Meloun M., Militky J. 2011. *Statistical data analysis: A practical guide.* Woodhead Publishing India.

[9] Nageswara Rao K., Talwar V. G. 2008. *Application Domain and Functional Classification of Recommender Systems – A Servey*. DESIDOC Journal of Library & Information Technology, 28(3):17-35.

[10] Nunes M. A. S. N. 2008. *Recommender Systems based on Personality Traits*, PhD thesis. Université Montpellier 2.

[11] Pennebaker J. W., King L. A. 1999. *Linguistic styles: Language use as an individual difference*. Journal of Personality and Social Psychology, 77:1296–1312.

[12] Ricci F., Rikach L., Shapira B., Kantor P. 2010. *Recommender Systems Handbook.* Springer, US.

[13] Roshchina A., Cardiff J., Rosso P. 2011. *User Profile Construction in the TWIN Personality-based Recommender System.* Submitted for publication.

[14] Surowiecki J. 2005. *Wisdom of the Crowds*. Doubleday, New York.

[15] Tausczik Y. R., Pennebaker J. W. 2009. *The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods*. Journal of Language and Social Psychology, *29*(1):24-54.

[16] Terveen L., Hill W. 2001. *Beyond Recommender Systems: Helping People Help Each Other*.

[17] Westfall P. H., Tobias R. D., Rom D., Wolfinger R. D., Hochberg Y., 1999. *Multiple Comparisons and Multiple Tests Using the SAS System*. SAS Institute.

[18] Witten I. H., Frank E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques. Machine Learning*. Morgan Kaufmann.

[19] Xiang Z., Gretzel U. 2010. *Role of social media in online travel information search.* Tourism Management, 31(2):179-188.