EVALUATING THE BIG FIVE PERSONALITY TEST AS A FOUNDATION FOR AN ONLINE STUDENT COURSE RECOMMENDATION SYSTEM

Queen Esther Booker
Minnesota State University, Mankato. 150 Morris Hall, Mankato, MN 56001,
queen.booker@mnsu.edu, 507-389-2445

Fred L. Kitchens
Ball State University, Miller College of Business, WB 203, Muncie, IN 47306,
fkitchens@bsu.edu, 765-285-5305

Carl Rebman
University of San Diego, Olin Hall 208, 5998 Alcalá Park, San Diego, CA 92110,
carlr@sandiego.edu, 619-260-4135

ABSTRACT

Recommendation systems have emerged as a useful e-commerce tool to assist customers in making purchases based on similarities and preferences of others. They require some basis for making recommendations ranging simply from past behaviors to complicated algorithms based on demographic and personality information. This study compares and contrasts the Big Five personality tests against simple past behaviors as part of a learning algorithm for understanding which traits are important in the selection of courses students find appealing.

Keywords: Education, Interactive systems, Recommender Systems

Introduction

Online educational programs are on the rise as confirmed by the 2005 Sloan Consortium, a distinct difference of reporting than those in 2001 when the Chronicle of Higher Education reported that online learning had gone "bust". The growth in the market can be attributed to improvements in course management systems and course technology. Yet, little has been done with one aspect of student support services – sharing information about which courses to take to fulfill requirements, a role traditionally performed using "word of mouth" between students and advisors, alike. The word of mouth approach is not easily transferable to the online student because the student's interaction with other students is oftentimes limited to course discussions and emails, limiting interaction with a broader range of the student body. With the barrier of face-to-face meetings removed, students will potentially face an explosion of courses available. Alba et al. (1997) argued that this is of little benefit to consumers without an efficient way to "screen products" effectively, where products in this case are courses. Without such screening, students may suffer from information overload (Jacoby, Speller, & Berning, 1974) or stop searching long before exhausting the set of relevant courses (Diehl, Kornish, & Lynch, 2003).

The search stop could potentially increase the likelihood of the student not completing the program, increasing the already statistically high attrition rates that currently plaque online enrollments at most colleges and universities.

To overcome the inherent difficulty of product selection in the e-commerce market, smart recommendation agents have emerged (Ansari, Essegaier, & Kohli, 2000; Iacobucci, Arabie, & Bodapati, 2000; Schafer, Konstan, & Riedl, 1999; Shardanand & Maes, 1995; West et al., 1999). Smart agents counteract the information explosion problem by "considering" a large number of alternatives and recommending only a small number to the consumer, similar to getting word of mouth recommendations from hundreds of people who share similar tastes rather than one or two people who's opinions a person trusts but may not share in similarity in tastes. An effective screening system that is highly correlated with consumers' overall preferences will have utility almost as high as if every alternative in the universe were inspected— but with dramatically reduced effort on their part (Häubl & Trifts, 2000). A search algorithm that provides such a service to students could assist in reducing attrition rates as students select courses (and perhaps a course of study) that are consistent with their interests. Several existing recommendation systems use some type of preference and/or personality test as a basis for making a recommendation. The two most popular are the online match-making companies and online college selection programs. Most recommendation systems simply rely on past behaviors of customers such as the recommendation systems for the purchase of multimedia and books such as Amazon.com. Given the importance of selecting courses to improve student retention, it seemed appropriate to evaluate personality tests

This study evaluates the use a Big Five personality test against with a simple demographic test to determine if either is effective in selecting courses that students liked. The Big Five tests were selected over the Jung and Myers-Briggs because the Big Five appear to be more accepted in the research community as a more reliable predictor than Jung/Myers-Briggs http://www.centacs.com/quickstart.htm). The study uses nearest neighbor as the recommendation algorithm and evaluates the usefulness of each test on the recommendation process.

Big Five Personality Test

Personality research, like any science, relies on quantifiable concrete data which can be used to examine what people are like and why people behave as they do. There have been several major personality tests that are used to help people identify potential careers or to understand management styles include the Big Five, Jung, and Myers-Briggs.

The Big Five was originally derived in the 1970's by two independent research teams who took slightly different routes but arrived at similar results: most human personality traits can be boiled down to five broad dimensions of personality, regardless of language or culture. These five dimensions were derived by asking thousands of people hundreds of questions, then analyzing the data using factor analysis. The Big Five is now the most widely accepted and used model of personality (http://www.centacs.com/quickstart.htm).

The bulk of academic research points to only five purely independent personality elements. "Only five" in the sense that every other personality trait will have some correlation to one or more of these five key traits. The Big Five personality system is based on the five proven independent elements: **Extroversion, Emotional Stability, Orderliness, Accommodation, and**

Intellect. These elements make up the primary colors of personality; the interaction of elements in each person yields their overall personality profile.
Each element has two oppositional type extremes:
Extroversion - Social and Reserved type
--Social types feel at ease interacting with to others
--Reserved types are uncomfortable and/or disinterested with social interaction
Emotional Stability - Limbic and Calm type
--Limbic types are prone to moodiness
--Calm types maintain level emotions
Orderliness - Organized and Unstructured type
--Organized types are focused
--Unstructured types are scattered
Accommodation - Accommodating and Egocentric type
--Accommodating types live for others
--Egocentric types live for themselves
Intellect - Non-curious and Inquisitive type
--Non-curious types are less intellectually driven
--Inquisitive types are insatiable in their quest to know more

The weakness of the Big Five theory is that there is some debate among researchers as to what makes up the core of each element. To employ a geographic metaphor, there is agreement about what region each of the five elements are located in, but different researchers might disagree on what precise city in that region is the core or keystone of some elements. Extroversion and Emotional Stability descriptions are fairly consistent. Orderliness, Accommodation, and Intellect are less consistently described. Nonetheless, as different as two researcher's labels or descriptions of something like Accommodation might be (some label it Agreeableness), their descriptions are still more similar than dissimilar. The differences, in their most extreme, are in the order of orange vs. red, not blue vs. yellow. (http://www.centacs.com/quickstart.htm). A copy of the Big Five test is available upon request.

Collaborative Filtering Algorithms

Recommender systems apply data analysis techniques to the problem of helping users and the items they would like to purchase by producing a predicted likeliness score or a list of top N recommended items for a given user. Item recommendations can be made using different methods. Recommendations can be based on demographics of the users, overall top selling items, or past buying habit of users as a predictor of future items.

Collaborative Filtering (CF) (Resnick, et al, 1994) is the most successful recommendation technique to date. The basic idea of CF-based algorithms is to provide item recommendations or predictions based on the opinions of other like-minded users. The opinions of users can be obtained explicitly from the users or by using some implicit measures.

The goal of a collaborative filtering algorithm is to suggest new items or to predict the utility of a certain item for a particular user based on the user's previous likings and the opinions of other like-minded users. In a typical CF scenario, there is a list of *m* users U = {u1, u2, …um} a list of *n* items I = {i1, i2; …, in}. Each user has a list of items within the list, about which the user has expressed his/her opinions. Opinions can be explicitly given by the user as a rating score, generally within a certain numerical scale. There exists a distinguished user u_a called the active user for whom the task of a collaborative filtering algorithm is to find an item likeliness that can be of two forms: prediction or recommendation.

Prediction is a numerical value expressing the predicted likeliness of item for the active user. This predicted value is within the same scale as the opinion values provided by active user.

Recommendation is a list of items that the active user will like the most. Note that the recommended list must be on items not already purchased by the active user.

There are many social filtering algorithms. Four such algorithms are described below.

- *The Mean Squared Differences Algorithm*. This algorithm measures the degree of dissimilarity between two user profiles, *Ux* (U subscript x) and *Uy* (U subscript y) by the mean squared difference between the two profiles: Predictions or recommendations can then be made by considering all users with a dissimilarity to the user which is less than a certain threshold *L* and computing a weighted average of the ratings provided by these most similar users, where the weights are inverse proportional to the dissimilarity.
- *Nearest Neighbor Algorithm.* Nearest neighbor is a special case of k-nearest neighbor method for classifying items based on closest training examples in the feature space. The training examples are mapped into multidimensional feature space. The space is partitioned into regions by class labels of the training samples. A point in the space is assigned to the class *c* if it is the most frequent class label among the *k* nearest training samples. Usually Euclidean distance is used. The best choice of *k* depends upon the data; generally, larger values of *k* reduce the effect of noise on the classification, but make boundaries between classes less distinct. A good *k* can be selected by parameter optimization using, for example, cross-validation. The special case where the class is predicted to be the class of the closest training sample (i.e. when *k* = 1) is called the nearest neighbor algorithm.
- *Clustering.* Clustering is a common technique for data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. Clustering is the classification of similar objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait - often proximity according to some defined distance measure.
- *The Pearson Algorithm*. An alternative approach is to use the standard *Pearson r* correlation coefficient to measure similarity between user profiles. This coefficient ranges from -1, indicating a negative correlation, via O, indicating no correlation, to +1 indicating a positive correlation between two users. Again, predictions can be made by computing a weighted average of other user's ratings, where the Pearson *r* coefficients are used as the weights. In contrast with the previous algorithm, this algorithm makes use of negative correlations as well as positive correlations to make predictions. (Booker, Kitchens, Rebman 2006)

The researchers chose to use the Pearson algorithm because it is used in many highly publicized recommendation systems such as Grouplens (Resnick et al, 1994)

The Research Method

The research study included 517 existing students. Students completed the following approval form and then completed a survey about their demographics and the Big Five personality test. After completing the demographic and Big Five personality test questions, students were given a list of elective courses. They were asked to rate the courses they had taken using the ratings found in Table 1: Scale for rating classes.

1 Best class I ever had.
2 Good course, worth the effort.
3 The course isn’t bad; it’s just OK.
4 I am neutral on the course
5 Barely interesting
6 I could hardly keep my eyes open or my mind from wandering
7 Worst class possible!

Table 1. Scale for rating classes.

The scale for ratings varies from 1 ‘‘best class I ever had!’’’ to 7 ‘‘worst class possible’’ (Table 1). A seven point scale was selected since studies have shown that the reliability of data collected in surveys does not increase substantially if the number of choices is increased beyond seven (Sharadan and Maes, 1995). Because users rate courses in different ways, the ratings are not normalized. For example, some users only gave ratings to courses they liked while others gave rating whether they liked the course or not. An absolute scale was employed and descriptions for each rating point were provided to make it clear what each number means.

Students rated an average of five elective classes out of a population of sixty possible courses. There are many advantages to using collected data rather than running the system real-time. Using such an approach one can systematically manipulate each of the factors in isolation, learning about its unique contribution; a simulation can be used on many different scenarios with high speed and efficiency to provide a general understanding of the system, parts of which can later be supported by behavioral experiments; finally, in a simulation, the right answer is known and it is easier to assess the accuracy of recommendations. Like any scientific method, simulation studies have disadvantages. In particular, a simulation can create a specific marketplace with a specific structure and decision rules. The results, therefore, cannot be generalized to all marketplaces and should be treated as examples and not proofs.

The researchers developed ten models from the collected data by randomly selecting 400 profiles for training the algorithm and 100 for testing the algorithm for both the simple demographic model and the Big Five test model. A random test was conducted on the 100 test items as well to measure the effectiveness of the demographic and Big Five models.

Quantitative Results

The study used the mean squared difference to compare the two three models. To evaluate the algorithm using each model using the demographics model and the Big Five personality test, a predicted value was established for each course in the course list. The Pearson correlation coefficient was defined as the basis for the weights (Resnick et al., 1994). The correlation between users *a* and *i* is:

$$w(a,i)=\frac{\sum_j (v_{a,j}-\overline{v_a})(v_{i,j}-\overline{v_i})}{\sqrt{\sum_j (v_{a,j}-\overline{v_j})^2 (v_{a,j}-\overline{v_i})^2}} \qquad (1)$$

where the summations over *j* are over the items for both users have evaluated a course.

Ten such target sets and data sets were randomly created and tested, to check for consistency in the results. In the source set, each person rated on average six courses of the 60 possible. The median number of ratings was 3. The *L* value was set at .5. The predictive value observed was between 1 and 7, reflective of how well a student would enjoy a particular course. Fractional components were rounded to the nearest integer. For example, if the predictor was 3.2 then the course was predicted as "just ok" where as a 3.5 was predicted as neutral.

The mean squared error and the standard deviation of errors of each predicted rating must be minimized was the criteria used to evaluate the models as well as the percentage of target values for which the algorithm was able to compute accurate predictions.

A summary of the predictive power of the three models in terms of accurate predictions is presented in Table 2.

Model	*Test Group Number*										Average
	1	2	3	4	5	6	7	8	9	10	
Random	50	49	30	45	60	55	50	57	43	50	48.9
Demographic Model	84	70	87	82	83	88	83	87	82	85	83.1
Big Five Personality Test Model	81	88	89	82	89	85	87	86	84	87	85.8
Big Five Plus Demographic Model	70	78	77	72	80	76	73	75	77	79	75.7

Table 2: Accuracy of the four models for predicting how much a student would like a course.

As the table shows, the most successful model was the Big Five Personality test but it's average was not significantly different from that of the demographic model (t-Test, P(T<=t) one-tail=0.1, P(T<=t) two-tail=0.2).

Conclusions and future work

The course recommendation system has potential. Beyond the recommendations, there are other ways in which OSRS can be appealing based on the online student's needs. In addition to grouping students based on similar interests and courses, online students can find online clubs, learning communities, that are oftentimes missing from the social aspects of the online learning experience.

The next steps are to examine more algorithms, make the system assessable to an online environment where more students from the University can have access to the system, and increase the type of uses to support the ever growing online student community. However, one of the challenges facing such a system is the difference in professorial approaches to the class. Another step would be to analyze the courses by sections and professors to determine the effect of the professor on student recommendations in a digital environment.

REFERENCES

Alba, J., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A., et al. (1997). Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces. *Journal of Marketing, 61,* 38–53.

Ansari, A., Essegaier, S., & Kohli, R. (2000). Internet Recommender Systems. *Journal of Marketing Research,* 363–375.

Booker, Q., Kitchens, F. L., Rebman, C., & Sharma, S. (2006). A Recommendation System for Student Program Planning. In *Proceedings of the 2006 Decision Sciences Conference.*

Diehl, K. R., Kornish, L. J., & Lynch, J. G., Jr. (2003). Smart Agents: When Lower Search Costs for Quality Information Increase Price Sensitivity. *Journal of Consumer Research, 30,* 56–71.

Häubl, G., & Trifts, V. (2000). Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science, 19*(1), 4–21.

Iacobucci, D., Arabie, P., & Bodapati, A. (2000). Recommendation Agents on the Internet. *Journal of Interactive Marketing, 14*(3), 2–11.

Jacoby, J., Speller, D. E., & Berning, C. K. (1974). Brand Choice Behavior as a Function of Information Load: Replication and Extension. *Journal of Consumer Research, 1,* 33–42.

Resnick, P., Iacovou, N., Sushak, M., Bergstrom, P., & Riedl, J. ‘‘GroupLens: An Open Architecture for Collaborative Filtering of Netnews’’, Proceedings of the CSCW 1994 conference, Chapel Hill, N C October 1994.

Schafer, J. B., Konstan, J., & Riedl, J. (1999, November). Recommender Systems in E-commerce. *Proceedings of the ACM Conference on Electronic Commerce,* 158–166.

Shardanand, U., ‘‘Social Information Filtering for Music Recommendation’’, MIT EECS M. Eng. thesis, also TR-94-04, Learning and Common Sense Group, MIT Media Laboratory, 1994.

Shardanand, U.,&Maes, P. (1995). Social information filtering: Algorithms for Automating “Word of Mouth.” In *Proceedings of the Conference on Human Factors in Computing Systems,* I. Katz, R. Mack, L. Marks, M.B. Rosson, and J. Nielsen, Eds., CHI ‘95, (pp. 210–217). New York: ACM

Press.

West, P. M., Ariely, D., Bellman, S., Bradlow, E., Huber, J., Johnson, E., et al. (1999). Agents to the Rescue? *Marketing Letters, 10,* 285–300.