Identifying Personality Traits among Peers in Collaborative Learning Environment

Shri Vindhya and T.Mala

Article history:
Received 12 October 2014
Received in revised form 26 December 2014
Accepted 1 January 2015
Available online 25 February 2015

Keywords:
Collaborative learning, Personality traits, Naive Bayes Classifier, MRC.

Abstract
Collaborative learning is group learning in which students are grouped together for learning purpose and will have objective in each group to learn towards a common goal. Heterogeneous grouping in such an environment will improve the performance of the students considerably. The Personality traits of a student have been shown to be an important predictor of performance in a group. The probability of a student’s personality types are detected automatically from his/her text and behavior with the system respectively. The Naive bayes classifier is used to classify the students personality based on the big five personality traits. MRC psycho linguistic database is used for preparing the training set and testing data set preparation.

Introduction

Collaborative learning is a situation in which two or more people learn or attempt to learn something together. Unlike individual learning, people engaged in collaborative learning capitalize on one another’s resources and skills (asking one another for information, evaluating one another’s ideas, monitoring one another’s work, etc.). It is based on the model that knowledge can be created within a population where members actively interact by sharing experiences and take on asymmetry roles. These include both face-to-face conversations and computer discussions (online forums, chat rooms, etc.).

Heterogeneous group formation is said to play a critical role in terms of enhancing the success of collaborative learning and therefore increasing the learning progress of students. Further, the students who ranked low in learning achievement have better learning behavior in heterogeneous grouping, James W. Pennebaker (1999).

The personality composition of groups of people working collaboratively on shared tasks has been shown to be an important predictor of performance. Rutherford (2001), conducted a study with groups of people to determine whether those with a heterogeneous personality composition were more productive and enjoyed working together more than those with a homogenous personality composition. Results showed that the homogenous control groups experienced more problems on a personal level, rather than technical problems. The heterogeneous experimental groups conveyed a broader and more varied skill of problem-solving and interacted more.

Student’s individual personality and his learning skill are the main predicted parameters that affect their performance in the group. Automatic detection not only reduces the work of the teacher but helps to provide significant results. Continuous monitoring of the student in a group is done to notice the dynamic change in his/her learning skill and Personality traits. Based on learning skill and Personality traits, grouping the students heterogeneously so that people within a group share differences and learns from each other. This kind of effective group formation can also be applied to non-academic activities.

The Big Five Personality traits, Demetrio A. Ovalle (2012) are proposed to be identifiable from a student’s interaction with the computer system. A Naive Bayes classifier is trained and tested for personality traits identification.

It is a simple probabilistic classifier based on applying Bayes theorem with strong independence assumptions. In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based...
on applying Bayes’ theorem with strong (naive) independence assumptions between the features. Naive Bayes models are also known under a variety of names in the literature, including simple Bayes and independence Bayes.

Related Work:
The collaborative learning environment is one in which students learn, interact with each other and improve their skills. Collaborative learning is a learning method that uses social interaction as a means of knowledge building. This concept has been around a long time. It is a learner-centered approach that requires learners working together to accomplish shared learning goal and to maximize their own and their group member’s achievements.

Several researches have been carried out to identify the personality traits of humans. The research has been done based on factors-analyzing the self-report and questionnaire data, peer reviews, and objective measures from experimental procedure in order to find the underlying factors of personality. In this paper the naïve bayes system approach predicts the human personality traits and academic characters. Each of the Big Five personality traits contains Uncertainty and retreat for Neuroticism; Interest and Confidence for Extraversion; Intelligent and Freeness for Openness/Intellect; Energetic and pattern for Conscientiousness; and Understanding and Courtesy for Agreeableness.

Identification of human nature from his/her traits has long been proposed by researchers. Human traits are often embedded in one's writings. Although some work has been done on identification of traits from essays, very little work can be found on extracting personality traits from written texts. Psychological studies suggest that extraction and prediction of rules from a data has been long pursued, and several methods have been proposed. In the present work we used Rough sets to extract the rules for prediction of personality traits. Rough Set is a comparatively recent method that has been effective in various fields such as medical, geological and other fields where intelligent decision making is required. Our experiments with rough sets in predicting personality traits produced encouraging results.

Mobashar Rehman (2012), proposed the personality traits for software engineers. Software development is performed by humans which emphasizes the importance of research on human of software engineering. This paper mainly focuses on the human aspect personality. Author mapped the skills required by various developers and then linked them to personality traits using OCEAN property.

Shlomo argamon(2005), The ways individuals use words can reflect basic psychological processes, including clues to their thoughts, feelings, perceptions, and personality. This paper focuses on determining “personality type” of the author from casual written text. It focuses on two key dimensions of personality: Neuroticism and Extraversion. It integrates two existing strands of research in language psychology and computational stylistics.

Each individual has unique personality attributes like behavioral, unreliable, expressive and psychological factor. This paper produces the results for all Big Five personality traits, in both talk and text, utilizing both self and observer ratings of personality, Digman J M (1990)

System Design:
The Learning Management System (LMS) is an environment which supports Collaborative Learning with all the essential features like a discussion forum, chat facility, course materials, self-assessments etc. The students can freely access all features of the LMS with a unique Login ID. The various comments made in the discussion forums, chats and their behaviour with the system such as accessing course materials, participation in activities are stored in the database.

The Personality Identifier process, as the name suggests aims to find the probability of a person belonging to a particular personality type such as Open, Conscientious, Extravert, Agreeable and Neurotic (OCEAN Model/Big 5 Personality Model). With the help of a dataset containing the text and corresponding personality (obtained from study conducted on college students by PenneBaker & King 1999), the extracted text of the students participating in the LMS are classified using the Naïve Bayes Classifier. The following Figure.1 depicts the detailed system architecture of personality traits identification process.

The essays and conversations of the student are extracted from the discussion forum. This text is given as input to the module. It is subjected to cleaning where stop words and tokens are removed using Lucene Standard Analyzer (Wikipedia stop word list is used). The output is a cleaned text void of stop words and tokens.

The Naïve Bayes Classifier which is used in the system for personality classification needs to be trained. The dataset used: Pennebaker and King Essays dataset. A group of college students were asked to mention their thoughts randomly and also were subjected to questionnaires to determine their personality according to the Big Five Traits. The dataset is subjected to TextCleaner module also. The cleaned text from TextCleaner Module is given as input to this module. Each word is categorized as per MRC dictionary and the average of the values of the fields is fixed for the conversation. The output contains the value of all fields of MRC for the dataset.
Naive Bayes Classifier is used for classification of personality (personality traits). It is trained using the File obtained from the MRC Field Value Calculator Module. The input conversations subjected to text cleaning and MRC Field Value Calculator is given as input to this module. Values of each field are labelled as high, medium, low.

![System Architecture](image)

**Fig. 1: System Architecture.**

The Personality Identifier system consists of a Text Pre-processor which removes the stop words (and, is, was, etc.) using the Wikipedia stop word list. (Available in the Internet) and also removes the tokens from the text (., .., …, “ ”, ”” etc.).

The cleaned text is now subjected to the MRC Field value Calculator which finds the average value for the text for each of the 14 fields of MRC Database. This is an input to the Naïve Bayes Classifier (testing data).

Similarly, the essay dataset of Pennebaker & King is processed by subjecting it to cleaning and MRC Field Value Calculator. The dataset contains conversations of more than 2300 students and their personalities detected by questionnaire. This is now used as the training dataset for the Naïve Bayes Classifier. It consists of average value of each conversation in the dataset for various fields of the MRC Database along with their corresponding personality.

The Naive Bayes classifier now classifies the text of the students participating in the learning environment to one of the personality classes using the Bayes Algorithm.

**Algorithms and implementation:**

This system has three parts:

1) Training set preparation
2) Testing Set Preparation
3) Classification of personality using Naïve-Bayes Classifier

**Training set preparation:**

Penne-Baker and King Essay Dataset is taken as input. The dataset contains the conversation of various students and their personalities obtained through questionnaire. It has about 2300 entries. Now the text is extracted, cleaned and tokenized. The MRC Field values are calculated and the average value for each field is found out. The values are found out from the MRC database which can be accessed from java code using jmrc interface. This is done for each conversation in the dataset and their corresponding personalities also noted. This is the training dataset for the classifier.

**Algorithm 1: Training set Preparation algorithm**

Input: Cleaned Penne Baker and King Essay Dataset.

```
mrc2.dct (MRC Database)
```

For each Text entry, Repeat
For each word
Get net, conc, aoa, nphon, nsyl, fam, kffreq, kfncats, kfnsamp, tffreq, bfreq, meanp, meanc, imag
Calculate avg value of each field for the conversation

Output:
File Containing MRC Field values for all entries in dataset

**Testing Set Preparation:**

The texts of the particular student extracted from the discussion forum in chats, Self Essays etc. is taken and the MRC field values for each student are calculated. Now this is the other input for the classifier.
Algorithm 2: Testing set Preparation algorithm

Input: Cleaned Student Conversation
mrc2.dct (MRC Database)
Algorithm:
For each Text entry,
Repeat:
For each word
Get nlet, conc, aoa, nphon, nsyl, fam, kffreq, kfncats,kfnsamp, tlfreq, brfreq, meanp, meanc, imag
Calculate avg value of each field for the conversation
Output:
File Containing MRC Field values for all entries in dataset.

Algorithm 3: Naive bayes classifier to determine personality:

The classifier used here is Naïve-Bayes Classifier. From the training dataset the value of each MRC field is categorized as high, med and low. Now the testing set is compared and the probability of belonging to a particular personality class is calculated using the Bayesian formula.

The output from the previous two sub modules is given as input to this module. The classifier used here is Naïve-Bayes Classifier. From the training dataset the value of each MRC field is categorized as high, med and low. Now the testing set is compared and the probability of belonging to a particular personality class is calculated using the Bayesian formula.

Algorithm 3: Naive bayes classifier algorithm

Input:
Training Set:
File Containing MRC Field values along with personality for all entries in PenneBaker & King dataset
Testing Set:
File Containing MRC Field values for students’ input text from discussion forum
Algorithm:
Initialize:
Extravert Probability=1, Agreeable Probability=1, Open probability=1, Conscientious probability=1, Neurotic probability=1
Calculate total people, no of extravert, no of agreeable,
no of open, no of conscientious people from training Set.
Define range for high, medium ,low values for each field in MRC.
For each student
End
Output:
P (Extravert), P (Agreeable), P (Open), P(Conscientious), P (Neurotic).

Conclusion:
In recent times, collaborative learning has emerged as the new means of e-learning. Studies have shown that, it is an effective learning methodology as it helps the students to share their knowledge and views among them thus improving their skills. Generally students perform better when they are in a group than as individuals. The grouping method proposed in this work will further enhance the experience of any collaborative learning environment. Since it creates an optimal heterogeneous group considering the learning skill and personality traits of the students, it will improve the individual performance of a student in a group. Further being an automatic process, it eliminates the need for the instructor to group students manually and also the need for questionnaires. The intra heterogeneous and inter homogeneous group formation which is obtained, will help the students to gain more from the group than when he/she is in a group which is formed in a random manner. Thus, the work of the instructor while forming student groups is minimized since the detection of personality and learning skills and the group formation is automated. This can be integrated with any Learning Management System (LMS) as it only requires features which are commonly found in most of the systems found today.

REFERENCES


