Nitte Meenakshi Institute of Technology

Department of Computer Science and Engineering

18CSE751 Introduction to Machine Learning

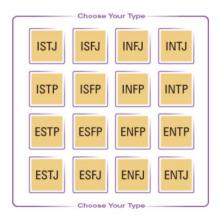# Learning Activity Proposal

Personality Type Prediction
Gagan R, Monish K and A S Prithvi Raj

Abstract

Myers Briggs Type Indicator (MBTI) is a personality type system that divides everyone into 16 distinct personality types across 4 axes:

- Introversion (I) – Extroversion (E)
- Intuition (N) – Sensing (S)
- Thinking (T) – Feeling (F)
- Judging (J) – Perceiving (P)

For example, someone who prefers introversion, intuition, thinking and perceiving would be labelled as INTP in the MBTI system. Based on the posts and searches made by a user, their personality type can be determined. It is used in businesses, social sites, for fun, for research and lots more. MBTI sorts for preferences or behavior and does not measure trait, ability, or character. The 16 personality types of MTBI are listed below in the form of a table.

Introduction

Machine learning is a subfield of artificial intelligence which is broadly defined as the capability of a machine to imitate intelligent human behavior. It allows machines to become more accurate at predicting outcomes without being explicitly programmed to do so.

Theory of MTBI: Random variation in behavior is quite orderly and consistent, being due to basic differences in the ways individuals prefer to use their perception and judgment. If people differ systematically in what they perceive and how they reach conclusions, then it is reasonable for them to differ correspondingly in their interests, reactions, values, motivations, and skills.

Data Set

The dataset contains over 8600 rows of data where each row contains:

- type - Personality type (4 letter MBTI code)
- posts - Last 50 things a user has posted/searched (Each entry separated by '|||')

The dataset needs to be preprocessed by separating each post/search entry. It also contains links to YouTube videos and webpages whose content can be taken to predict the personality type.

Link to dataset: https://www.kaggle.com/datasnaek/mbti-type

Machine Learning Methods

The Machine Learning methods that will be applied on the above dataset are as follows:

1. C4.5
2. Support Vector Machine (SVM)

C4.5: It is a decision tree classifier and an extension of ID3 algorithm. It generates decision trees by choosing the attribute with the highest normalized information gain.

Its advantages over other decision tree classifiers are:

- Single pass pruning – efficient and prevent overfitting
- Handles both discrete and continuous attributes
- Handles incomplete values and missing data

SVM: It is a linear model for classification and regression problems based on the margin maximization principle. It constructs a hyperplane or a set of hyperplanes in a high dimensional space to classify datapoints.

It can also perform non-linear classification by using kernel trick which implicitly maps the inputs into high-dimensional feature spaces.

Its advantages are:

- Effective in higher dimensional spaces
- Can perform non-linear classification
- Relatively memory efficient

Assessment

Accuracy: It is the fraction of predictions that our model has got right.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

Precision: It is the quality of positive predictions made by our model.

$$Precision = \frac{Number\ of\ true\ positives}{Total\ number\ of\ positives}$$

F1 Score: Weighted average of precision and recall. Used to compare two different classifiers.

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Presentation and Visualization

- Determined personality type for each test sample (MBTI code)
- Percentage of top MBTI codes matching with the user
- Overall accuracy, precision and F1 score of each model

Roles

- Gagan R - Proposal documentation, some modules of SVM, part of report documentation
- Monish K - Preprocessing, some modules of C4.5 and SVM, part of report documentation
- A S Prithvi Raj – Preprocessing, some modules of C4.5, part of report documentation

Schedule

Date	Tasks to be Completed
20/12/21	Proposal documentation
02/01/22	Preprocessing of dataset
09/01/22	Implementation of models
16/01/22	Assessments, results, and report

Bibliography

[1] https://www.myersbriggs.org/my-mbti-personality-type/
[2] https://en.wikipedia.org/wiki/C4.5_algorithm
[3] https://en.wikipedia.org/wiki/Support-vector_machine