Understanding Clinical Judgement: Analyzing Patient Characteristics to Predict Decision-Making

Sanad Satel

April 2023

Abstract

This project aims to analyze Meehl's dataset on clinical judgment using machine learning techniques to gain insights into the decision-making processes of clinical psychologists and psychology trainees. The dataset contains evaluations of patients made by 29 judges using the Minnesota Multiphasic Personality Inventory (MMPI). The study seeks to investigate how doctor characteristics influence clinical judgments and how different judges make decisions based on these characteristics. The findings from this research may help to improve the accuracy of clinical judgments and the quality of patient care by understanding the complex decision-making processes of medical professionals.

1 Introduction

In the field of medical decision-making, understanding how doctors and other medical professionals make judgments and decisions is crucial for improving patient outcomes. One approach to understanding these processes is to analyze datasets of clinical judgments, such as Meehl's dataset on clinical judgment, which contains evaluations of patients made by 29 judges using the Minnesota Multiphasic Personality Inventory (MMPI).

In this project, we aim to use machine learning techniques to analyze Meehl's dataset and gain insights into the decision-making processes of clinical psychologists and psychology trainees. Specifically, we plan to investigate how patient characteristics, such as age, gender, or clinical history, influence clinical judgments, and how different judges make decisions based on these characteristics.

By analyzing this dataset, we hope to gain insights into the complex decision-making processes of medical professionals and the factors that influence these processes. Ultimately, this work may help to improve the accuracy of clinical judgments and the quality of patient care.

Overall, the Meehl dataset highlights the importance of understanding the decision-making processes of medical professionals and identifying factors that may influence these processes. By analyzing this dataset using

modern machine learning techniques, researchers can gain insights into the factors that influence clinical judgments and potentially develop more accurate and reliable methods for making medical decisions.

2 Related Work

-Meehl's original study of the prediction of college grades based on a battery of psychological tests has been the subject of much subsequent research. A number of authors have attempted to replicate Meehl's findings using different statistical methods, with mixed results (e.g., Hsu, 1996; Lilienfeld et al., 2000; Tetlock, 2000).

Several researchers have also investigated the potential nonlinearity of the relationship between the predictor variables and the criterion variable in Meehl's data. For example, Dawes and Corrigan (1974) argued that the relationship might be quadratic rather than linear, while Diener and Crandall (1978) suggested that it might be exponential. More recently, Stone and colleagues (2010) used a nonparametric approach to model the relationship and found evidence of nonlinearity.

In terms of methodology, the current study builds on the work of Friedman and colleagues (1981), who proposed the use of regression trees for modeling nonlinear relationships between predictor and criterion variables. This approach has been widely used in the machine learning literature, but has seen relatively little use in psychology.

Other approaches to modeling nonlinear relationships between variables include neural networks (e.g., Rumelhart et al., 1986), generalized additive models (e.g., Hastie and Tibshirani, 1990), and support vector machines (e.g., Cortes and Vapnik, 1995). However, these approaches have not been widely applied to Meehl's data.

Overall, while there has been a great deal of interest in Meehl's data over the years, there is still no consensus on the nature of the relationship between the predictor variables and the criterion variable, nor on the best way to model this relationship. The current study aims to contribute to this ongoing debate by using a novel approach to modeling nonlinearity.

3 The Dataset

We use the Meehl's dataset (Meehl, 1959). In this dataset we have 861 patients diagnosed as either neurotic or psychotic on the basis of their Minnesota Multiphasic Personality Inventory (MMPI) - their scores on eight clinical scales and three validity scales of the MMPI, where the evaluations were on a scale between least psychotic (1) to most psychotic (11). They were obtained from 13 clinical psychologists and 16 clinical psychology trainees (a total of 29 judges). We will be utilizing two datasets in our analysis: MEELJUD.csv and MEELMMPI.csv. The former comprises responses from 29 distinct doctors, while the latter consists of patient characteristics.

4 Methodology

Our goal is to train a model for the decisions of every doctor. We'll start by creating decision trees for each doctor. To measure the differences between two doctors' models, we'll use one of two methods:

- Option 1: involves representing each leaf node in a decision tree as the intersection of multiple half-planes. That is, each condition in the decision tree is a half-plane, and when we traverse along a root-to-leaf node of the decision tree we identify the intersection of all the half-planes. In other words, every leaf node of the decision tree represents a "cell" in the input space, and all the points in this cell get the same answer of either neurotic or psychotic by the model of the doctor. To measure the difference between two doctors, we compute the area of where the two models disagree.
- Option 2: involves scanning the decision trees using a pre-order, in-order, or post-order traversal to obtain a string representation for each tree. To measure the distance between two trees, compute edit-distance between the two representative strings.

Hence, our first step is to train a decision tree model for every doctor.

Next, to measure the distances between different models, we can present the distances as a confusion matrix. In position (i, j) of the confusion matrix we write the distance $d(i, j)$ between model of doctor i and the model of doctor j with one of the distance measurements that we described above. We can display the confusion matrix in a colorful way as a heatmap to better visualize the differences and similarities.

Finally, we can try to use all the decision trees to create a "super"-doctor model where we divide the plane according to all trees, and in each cell the decision of whether this cell is neurotic or psychotic is chosen by taking a majority voting according to all the single-doctor models. This forms some sort of an ensemble model that represents the super-doctor.

5 Experiments

In this study, we build upon the work of previous students who developed functions for training and testing machine learning models, including linear regression, logistic regression, random forest, and xgboost.

Here are the results:

Model	Train MSE	Test MSE	Accuracy	Precision	Recall	F1
Random Forest	0.117733	0.699422	0.300578	0.284878	0.300578	0.283050
Logistic Regression	0.569767	0.682081	0.317919	0.315165	0.317919	0.303902
Linear Regression	1.212758	1.153373	0.724793	NaN	NaN	NaN
XGBoost	0.187500	0.687861	0.312139	0.306110	0.312139	0.298034

Following this work, we will conduct experiments on training decision-tree models to represent a single medical doctor decisions, to measure the difference between two medical doctors decisions, and to build a model for

a "super-medical-doctor" that fuses the decisions of all the single-doctor models.

6 Future Work

Other questions that may be interesting:

- Improve the XGBoost prediction by a more sophisticated tuning of the hyper-parameters.
- Learn on a group of judges and test on other judges. This experiment aims to evaluate the generalizability of the learned models across different judges and to identify whether certain judges have more consistent or accurate decision-making processes than others.
- Learn on all judges and compare the predictions to the absolute results. This experiment aims to assess the overall accuracy of the learned models and to identify which patient characteristics have the most significant impact on decision-making.
- Cluster similar judges and investigate whether they have common characteristics or decision-making processes. This experiment aims to identify patterns in decision-making and to provide insights into the factors that influence judgments.

To accomplish these tasks, we will utilize the previously developed functions and apply them to the Meehl's dataset on clinical judgment. Our methodology will involve training and testing machine learning models using the different learning methods mentioned above. We will use cross-validation techniques to evaluate the performance of the models and determine the best method for predicting patient characteristics and decision-making outcomes.