Term Project Report

Project: The Homedepot

Team Member: Tao Li (github:taolee24 ;TAL88)

Yue Su(github:LexieSu; YUES55)

Dan Sun(github: danielsun510;DAS225)

Yumeng Lu(github:babylu; YUL139)

Zhaoxuan Ren(gitbub:helenren1994;ZHR5)

Table of Contents

Introduction

Home Depot is one of the world's largest home improvement retailers and the second largest retailer in the United States, it provides a platform for customers to find and get their productions efficiently and accurately to meet their home improvement needs.

By clicking the mouse or tapping the screen, customers will get the correct information. In the term project, our purpose is to build an appropriate model using certain methods and tools to predict the relevance of search results and improve the customers' shopping experience.

In this project, our task is to complete the analysis of data set to generate the results, to illustrate the relevance relation. In particular, we see how we can apply machine-learning techniques to this project.

Data Analysis and Visualizations

The data we use is the training set of Homedepot Search Relevance from Kaggle.com. The datasets include 5 csv files which are :

train.csv - the training set, contains products, searches, and relevance scores

test.csv - the test set, contains products and searches. You must predict the relevance for these pairs.

product_descriptions.csv - contains a text description of each product. You may join this table to the training or test set via the product_uid.
attributes.csv - provides extended information about a subset of the products (typically representing detailed technical specifications). Not every product will have attributes.

sample_submission.csv - a file showing the correct submission format
relevance_instructions.docx - the instructions provided to human raters
product_

Tools we used

SPSS

Matlab

Time Schedule

3.5 First Meeting with team members and make division

3.10 Complete the report together

3.15 Second meeting

3.20 Finish coding part

3.25 Revise