**Feature Engineering for Search Advertising Recognition**

Feature engineering is one of the key technologies in the research of search advertising recognition. Most of the existing search advertising methods are selected according to the prior knowledge, which is too subjective to be popularized. Taking the advertisement of Ali search advertising as the research object, a feature processing method based on the pre-analysis of store and user data is put forward, and then the conversion rate is predicted with XGBoost (eXtreme Gradient Boosting). Experiments show that compared with other priori Feature Engineering, the proposed method can significantly improve the prediction results.

**EXISTING SYSTEM:**

In Feature Engineering, traditional feature processing methods are linear combination of original features, one-hot coding and so on. It is difficult to improve the recognition rate in traditional ways Businesses purchase specific keywords according to the characteristics of the goods. When users enter these keywords, the corresponding advertising goods will be displayed in the pages that the user sees.

**DISADVANTAGES OF EXISTING SYSTEM:**

* The conversion rate of search advertisements is used as an index to measure the effect of advertising transformation, that is, the probability of advertising products being bought by users after clicking.
* Data is the carrier of information, but the original data contains a lot of noise, and the expression of information is not concise enough.

**PROPOSED SYSTEM:**

This paper, taking Ali search advertising as the research object, proposes a feature processing method based on store and user data pre-analysis, which aims to pre-analyze the features, that is, the first prediction processing of the features of users and stores, and as a new feature. Calculate the correlation between each feature and response variable. The commonly used methods in engineering include Pearson coefficient and mutual information coefficient, Pearson coefficient can only measure linear correlation, and mutual information coefficient can measure various correlations well.

**ADVANTAGES OF PROPOSED SYSTEM:**

* The model of a single feature is constructed, and the feature is selected by the accuracy of the model, and then the final model is trained when the target features are selected.
* After feature selection, features are selected again if the user id and user characteristics are combined to obtain a larger feature set and then select a feature, this practice is more common in recommendation system sand advertisement systems.

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium Dual Core.
* Hard Disk : 500 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 1GB.

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 7.
* Coding Language : Python
* Tool : PyCharm, Visual Studio Code
* Database : MYSQL

**REFERENCE:**

Yue Sun School of electronic information, Qingdao University, Guowei Yang College of information engineering, Nanjing Audit University, “**Feature Engineering for Search Advertising Recognition**”, 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) IEEE Xplore: 06 June 2019, DOI: 10.1109/ITNEC.2019.8729471