

A Generalized and Automated Data Mining Approach for Sentiment Analysis & Recommendation

Introduction: In today's technically advanced world where product and services increased way more than we imagine. With a lots of available choices users get a lot of option to choose. While a lot of same type of products available, customers get confused with the choice he/she should make. In such situation user try to collect information about available product through social media to find the best product. So, ass a recommendation system developer, it's our job to understand their need and recommend them the best product they been searching for, suggest them a product which satisfies them. This process of SANR (sentiment Analysis & Recommendation) works in two steps, SANR for new users whom data is available for the 1st time, and enhancement model that enhances the recommendation for already registered customers. The core of this idea lies in parallel working of these two steps, that is, "Recommend & Enhance".

Objective: Based on available data, this model has 3 main objectives. 1st is to collect data for new users and recommend products, this a kind of generalization, where we recommending to all based of the definition they satisfy. 2nd is to wait for user response to recommendation, when we have sufficient data, we will use this data to train a new recommendation model to these users (moving from general to specific), this will be used in two ways. 1st is to enhance our 1st model (recommend to new users) and 2nd to create new model for already registered user. 3rd step involves prediction, since we have a lots of data, we would like to have a prediction model in background (since it will help us when user response not available). We will use this model to predict whether user will respond or not. It will helps us in two ways, 1st will helps predicting response (a way of filling missing data), 2nd will helps us to improve our 2nd model of recommendation (for old users).

Steps Involved:

1. This step includes SANR for new users as described above. In order to do that we will create a universal feature matrix which will we created from the data we collected. Data corresponding to all users will be mapped into this frame. Further, it will be used to clustering and based on results we will recommend products to users. Here we are using a generalize clustering, a huge number of customer will recommended the same products, so the error in confusion matrix are going to be in great amount. A fix frame of feature matrix will be implemented, modification (adding/removing of feature) can be done based on the requirement we have (mostly in second step).
2. This step is an enhancement step, it will helps us to implement specific and effective recommendations, after releasing the recommendations, we will wait for user to act on those recommendations, this will collect us the response data. This response data along with data present in feature data will help us to create a classification model (in our case multiclass-classification). Some time it might happen that user might not respond or not action is recorded, in such cases 3rd step described below will help us to fill missing data. Here we will use 10+ ML (machine learning) algorithms to create a model of multi-class classification. Here we will look for specific groups and recommend them specific products. Here if be about to find a new feature that has more weightage in classification, we will add that to our feature matrix. On the other hand, if the feature seems not to affect our recommendations then we will remove that.
3. This step will include predictions, whether user will choose to respond to the recommendation or not. So, this will work like a background model, where we predict the outcomes of recommendations, we will predict future outcomes, future trends to our model. Each time, this model will keep improving, since we will also be having true data (coming from user response). This model will also fill up missing response data (as mentioned above).

Technology used:

Mostly, will work with python, R. Everything will be automated through python scripts. MLs will used frequently.

Benefits:

1. In 1st step we are using feature matrix for recommendation, which is optimized in later steps, and will be more efficient and fast. This feature matrix can be modified according to needs.
2. 2nd step uses multiclass classification which will reduces error to bare minimum. It quite effective and this model can also be used in our 3rd step. And 3rd steps will provide us backbone and really good strategies. This will helps us in preventing annoying recommendations and will make our model superefficient.
3. For reference, link to one of my own [project\(on multiclass- classification\)](#) and [case study\(on prediction\)](#)