What’s Good – Item Recommendation Engine

GA Data Science - Final Project Outline

Tom Harel

November 19, 2014

1. Problem to be solved

The problem to be solved is a prediction problem. We want to predict items a particular consumer will want to purchase in order to increase overall conversion rate and company profits.

2. Description of dataset

* What dataset will you be using for your project?
  + Internal company data. Specifically purchase data for 1,000 top clients in San Francisco in YTD 2014.
* How will you access this data? API? Scraping? Other?
  + Direct access to company database.
* What sort of information is in the dataset? In other words, what features are available?
  + The dataset is composed of the items ordered by the 1,000 top clients above.
* How will you turn that data into a training set? (If using a supervised approach)
  + By nature of past data, it depicts user preferences and will be naturally used as a training set.
* How do you anticipate processing that data to get it into a form to use for your modeling?
  + Using Python’s libraries like Pandas and Numpy.

3. Hypothesis

* What is your hypothesis? In other words, what do you hope to predict or otherwise learn as the outcome of your project?
  + I hope to be able to learn that past purchase behavior is predictive of future choices and preferences.
* What are some of the features you might use?
  + Past user purchase records.

4. Statistical methods I plan to use and why

Think back to our 2x2 matrix slide: Is your problem a classification problem or a regression problem? Will your approach use supervised approaches or unsupervised approaches?

* This is a supervised regression problem (I will be using ALS).

Which of the machine learning algorithms that we have learned do you plan to use for your final project and why? Which do you explicitly NOT plan to use and why?

* I plan to implement a model-based collaborative filtering (CF) recommendation engine. By the nature of the data available to me (order history), I think this CF model fits my data best.

5. Applications the finding may have

Once you have completed your project, what are some of the applications of your findings? In other words, how might those findings be applied? What is the “practical” value of the model you will have built?

* This model will ultimately be scaled and used on production on Eat24’s *What’s Good* mobile app feature (on iOS and Android).

Also, what will your deliverable be in addition to your code and data? Will you write a report in the style of CS229? Will you create a visualization? (NOTE: Do NOT attempt to learn D3 on top of everything else unless you are already a javascript ninja! Seriously.)

* Neither applies to me. My deliverable is a useable product that will be (hopefully) implemented.