Optimizing Conversion Rate

This document describes an approach to optimize the ranking (sorting) of a list of products for conversion rate.

Conversion Rate

Conversion rate of a single product is defined as the total orders (purchases) per unique page view of the product. A higher conversion rate for a product implies that a higher percentage of views are getting converted to purchases.

Objective

The goal of this task is to “optimize” the ranking of a list of products for conversion rate. One approach to achieve this is to calculate the conversion rate for each product and assign higher ranks to the products with higher conversion rates.

However, two main challenges are faced in achieving this goal. First challenge is that only a small percentage of the listed products are purchased atleast once. Therefore, for a large number of products, conversion rates equal zero. In order to rank the non-purchased products, a different scoring mechanism needs to be devised. In this approach, total activity (of any kind) divided by the age of the product is used to rank the non-purchased products.

Secondly, page views for a product is tracked by events such as ad click, email click, view from a mobile application and finally a webpage view. Often, these events are out of sync with the purchase events or entirely unavailable. Hence, unique page views for a product can only be approximated.

Code Usage

Open the command line and type

“python optimize_conversion_rate.py <path/to/event_data.txt>”

Algorithm

The algorithm takes event information as input (as provided in event_data.txt). This is in the form of a table with each tuple (record) containing event information such as product id, event name, event date, total events and unique customers participating in the event.

1. The first step of the algorithm is to convert the event data to product (also referred as item) data. Indexing by product id speeds up calculation of conversion rates and activity score for each product. This returns a dictionary with product id as the keys and all the events corresponding to that product (e.g. order, mobile view, date select etc.) as values. This also identifies the items that have been purchased atleast once and separates them from the

products that have zero purchases.

2. The second step is to calculate the conversion rate for the purchased products. For a given product, conversion rate is calculated by dividing the total number of order events (order) by the total number of unique customers in the page view event (pv) corresponding to the product. However, since pv in itself can be unreliable, unique customers in other events such as ad click, email click and mobile view are also augmented in order to calculated the total views of the product.
3. The third step is to calculate the total activity around a non-purchased product since it was uploaded, divided by how long it was available online. In this method, the total activity is calculated by summing the unique customers in all the events corresponding to a product. The age of a product is approximated by the date of the first event. This gives an approximation of how popular the product is, normalized by its age.
4. The final step is to sort the products. The purchased products are sorted according to their conversion rates (in descending order). For the non-purchased products, sorting is performed based on the unique activity normalized by the age. The two sorted list are concatenated to produce the final ranking of the list of products. The products ranked higher in the list are the ones that have been purchased atleast once and have a high conversion rate. These are followed by the products that have not yet been purchased but display high overall activity. The lowest in the list are the products that have not yet been purchased and display low overall activity.

Performance

Steps 1, 2 and 3 are linear in the number of events, i.e. as the number of events increases, the time taken for steps 1, 2 and 3 increase linearly. Step 4 has a time complexity of O(N log N), where N is the number of events. This is due to the sorting sub-step involved in the Step 4.

Other Possible Ideas

The non-purchased items can be interleaved with the already purchased ones. Placing a non-purchased item between two items with high conversion rates may increase the chances of conversion of the non-purchased item.

Improving in a different direction, a second ranking may be created that places much higher importance on the most recent activity (say last one week). This would capture the notion of current trend thus making it more sensitive to factors such as weather, holiday season etc.