Kruegger
64th place

Next stage: from Bronze to Silver in 3 days...

posted in Instacart Market Basket Analysis 13 days ago

70

Resurrected post. The title should be read as **"Brief summary of next steps..."**

As previous, this post is for novice kagglers who are want to learn something new about methods, technics and approaches..

To understand and implement these approaches you have to spent your time and efforts, so there are no "Silver bullet" or "Just click and get top-10 LB".

And I highly appreciate and want to say "Thank you!" to all people who shared their thoughts and who first go on this road to make this way more easy to all.

OK. Let's begin...

This competition (from my unexperienced point of view) divide your work into two main stage:

1) Build the model to get probabilities for each product to be reordered

2) Convert these probabilities to product list by order to submit

1'st stage we discussed in my previous post, you can implement any model you want, stacking, blending, ... any other method to get probabilities. At 2'nd stage you have to convert it to product list.

The simplest way to do it (also implemented in all public kernels) is to use fixed treshold (0.19-0.21) and put all product with probabilities > treshold to predicted basket, or 'None' if there are no such products.

The simplest, but not optimal.

Great thanks to kaggler's who first find this approach:

@SVJ24 >Think of a better strategy to get the number of products per order: If I select the number of items actually present in the order to select the top products based on my binary classifier, I get a mean F1Score of 0.403 on my CV which should translate to 4.05 on LB. This means if I have a model which can predict the number of items in each order with 100% accuracy then this is the score I should get. So there is this much potential to improve my strategy to get the

order size Think about how to handle None: I got a 0.003 percentage point lift, when I figured out a better way to handle this, I am sure there is an even better way

@zaq1xsw2tktk

> According to my experiment and calculation, different order should have different threshold. For example, case1: reorder_prob(A) = 0.9, reorder_prob(B) = 0.3 expectation(F1(A)) = 0.81 expectation(F1(A B))=0.71 case2 reorder_prob(A) = 0.3, reorder_prob(B) = 0.2 expectation(F1(A)) = 0.28 expectation(F1(A B))=0.313333333

@SVJ24

> As far as Nones are concerned there are two ways to handle it:
>
> - Method 1: As already mentioned in this post https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/35716#198992, add None as a product to any order which has zero reordered product
> - Method 2: Alternatively, P(None) = P(not item 1)P(not item 2).....*P(not item k)

To summarize: If you want to convert your probabilities obtained from model to product list in optimal way, you have to implement at leas two things:

- Use variable treshold for order in different ways
- Use more smart 'None' handling

Here are the links to two interested papers about first one:

- https://arxiv.org/abs/1206.4625
- https://arxiv.org/abs/1402.1892

And links to most interesting topics (IMHO) about this problem in the messageboard:

- https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/36544
- https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/33276

- https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/36134
- https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/36239
- https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/35048

If you prefer "Click and win" solution, it also exist:

- https://www.kaggle.com/c/instacart-market-basket-analysis/discussion/35048
- https://www.kaggle.com/mmueller/f1-score-expectation-maximization-in-o-n

P.S. Thank for all your feedback about this (deleted) post. I have learnt a lot about community.

@kruegger

Options