



Featured Prediction Competition

Corporación Favorita Grocery Sales Forecasting

\$30,000

Prize Money

Can you accurately predict sales for a large grocery chain?



Corporación Favorita · 1,707 teams · 15 hours ago

Overview

Data

Kernels

Discussion

Leaderboard

Rules

Team

My Submissions

New Topic

**slonoslon**posted in [Corporación Favorita Grocery Sales Forecasting](#) 5 hours ago

Congratulations to the members of #1 and #2 teams, you did a great job here. And my deepest condolences to the teams suffered in this shake-up: I feel for you, guys.

My teammate, [Ahmet Erdem](#), will hopefully describe his approach in more detail later, when the time will permit, but as a short summary : He used gradient boosting (LGBM) with an original set of features, and tried to capture the monthly dynamics by using sales on the same days of different months as the train targets.

My part of the final solution consists mostly of neural networks, although I also used a LGBM model with a low weight in the final ensemble.

When it comes to time series, I strongly believe that the two most important things are validation and bagging. By correct validation I mean simulating train/test (train/unknown future) split to avoid any kind of future information leaks. Thus, the validation set should include a time period that is not present in train in any way, even with different timeseries, because these time series could correlate with each other.

So I used the following train/validation split: for the training set, I took a history of 80 consecutive sales days for each (item, store pair), and used the next 16 sales days as a train set target. The split between training history and training target was always from Tuesday to Wednesday, as in the original train/test split, in order to capture the weekly dynamics. After training my models on this set and estimating an optimal number of epochs to run, I retrained them on the latest

data available (i.e. finals days of the original training set) in order to use the most up-to-date information.

About bagging : our final result includes models, trained on 10 runs each. The weights initializations are different each time, so results are different as well, and averaging them helps to improve the solution, particularly when dealing with the uncertain future.

Another part of my bagging was a bit unusual. After each training epoch (including the very first) I predicted the targets, and then averaged these predictions. I've read about averaging a few last epochs, but never heard about averaging everything, even the first very underfitted epoch. Yet my validation consistently showed that it does improve the results.

As for the model's achitectures, I belive they are relatively less important here. I had an ensemble of 3 NNs : CNN, LSTM and GRU with almost equal weights. As a single model, GRU shows a bit better results than the rest, but an ensemble is even better, as usual.

All of them have 3 top (convolutional or recurrent) layers, 2 or 3 bottom fully-connected layers and embedding layers for categorical features. I tried different tweaks like skip-connections, concatenating CNN layers with different convolution kernel sizes and so on, but the results were negligible. The only achitectural tweak that worked better was using dilated convolutions: my CNN uses convolutions with dilation of 2 on the first layer, 8 on the second and 32 on the third (and no pooling). But an original plain CNN without dilated convolutions worked just a bit worse.

What did not work: The main disappointment was the poor results of a version with updated promotion information for the train set. I've used a relatively sophisticated method to fill onpromotion status for the dates of the training period without sales, and it showed very promising results on the public leaderboard on the last day of this competition, boosting our score from .501 to .499. And it failed miserably on the private, so we remained in the prize area only because our other submit was more conservative and used less weight for these adjusted promotions. So next time I'll try to be more careful with the last day "improvements".

Options

Comments (1)

Sort by Hotness



Click here to enter a comment...

CPMP • 5 hours ago • Options • Reply

1



Thanks for sharing.

capture the monthly dynamics by using sales on the same days of different months as the train targets.

Interesting, I'm not sure we managed to capture the monthly dynamics (payday) that well.

After each training epoch (including the very first) I predicted the targets, and then averaged these predictions

Also something I wanted to try and did not. We definitely do it next time.

Congrats again on your prize, well deserved.