Machine Learning Design

Duration: 20 Minutes
Difficulty: Medium
Domains: Product / Forecasting

Problem

An ecommerce platform like Amazon wants to forecast sales. How would you design a forecasting system?

Solution

[Candidate] To address this problem, I need to frame it first. Can I assume that the forecasting is performed at the item level? Or, is it performed at the department level?

[Interviewer] Let’s presume that the design is performed at the item level.

[Candidate] I see. Right off the bat, I could foresee a potential issues that need to be addressed. In forecasting, the more granular the time series, the higher the sparsity (i.e. lack of data sales history at the item level compared to the department roll-up).

[Interviewer] Good point. We can address the cleaning issue later, but I’d like to see you frame the problem more. Can you think of other considerations?

[Candidate] Certainly. The granularity and horizon need to be defined. Are we talking about forecasting sales of an item at the weekly level or monthly level? Is the forecast horizon one week ahead or one month ahead? These are a few items that need to be defined. They are usually defined based on the business needs.

[Interviewer] Okay, how would you guide what they should be to the business?

[Candidate] I suppose, if the forecasting is performed at the department level, I would imagine this is usually for reporting purpose and projecting potential sales. So, monthly granularity with a long forecast horizon, let’s say 1 year ahead, should suffice. If it’s performed at the item level, I believe the application for this is more for planning and inventory fulfillment. So, it makes most sense to perform a 1 week granularity and 1 to 3 months of forecast horizon.

[Interviewer] That makes sense. Let’s resume the data quality issue. You mentioned that the more granular the data, the more sparse it is. How would you address this in time-series?

[Candidate] So, based on my experience, one common issue is the lack of data history. Some items might be new. Hence, there’s not enough transaction history to train a model to forecast the sales. In a sense, it’s a “cold-start” problem commonly found in recommender system. *How do you predict a behavior on a new user without their activity data?*

[Interviewer] So, how would you address such situation in time-series forecasting?

[Candidate] Hmm… I believe there’s two ways to handle this. One is to forecast items with at least 1 year of history.

[Interviewer] Gotcha. Let’s proceed. Can you talk about preprocessing steps? What kind of modeling would you use?

[Candidate] I would remove outliers using median based smoothing first and apply time series decomposition to see if there are trends, seasonality and holiday effects. One model that can handle trends and seasonality is SARIMAX. So, I would start with that model as the baseline.

[Interviewer] Great. Can you talk about how you would evaluate the model? And, what does a success look like to you?

[Candidate] So, I normally use mean absolute percentage error (MAPE) for evaluating a forecast model. But, in this particular case, some items may have $0 sales on some weeks. This will pose an issue for metric given that MAPE assumes that the actual is never 0. The formula is MEAN(|expected - actual| / |actual|). To address the 0 issue, the MAPE can be translated into the following: MEAN(|expected - actual| / |actual + 1|).

In terms of the model success, I would say anywhere between 10 to 20% MAPE achieved should be considered a successful roll-out. If there's already a benchmark forecasting and, it beats the business by 5 to 10%, then that should considered as a success as well.

Interviewer Solution

A naive response will say “use univariate forecasting models such as Autoregressive Integrated Moving Average (ARIMA) to project sales for the next several weeks.”

Applying a univariate forecast model is just a part of a solution. You need to engage with the interviewer to design a system that addresses key points:

1. Determine the granularity of the forecast system
2. Handle non-stationarity
3. Determine the training window and forecast horizon
4. Algorithm selection and evaluation
5. Production modeling

Determine the forecast granularity

Let’s discuss each of the points above. The first step is to determine the forecast granularity. Are you forecasting at each item, item category, device, region or country?

You can aggregate and forecast sales by each item, or apply the same procedure on other granularity. You could also apply the procedure within each combination of the factors. For instance, forecast the sales of an item X in country Y.

Depending on what your choice of aggregation, the number of time-series and sparsity will change. For instance, if your dimensionality is two such that you are predicting an item X within device X and country Z, the number of your time-series will be far greater than if you were to merely predict for each item, summed across devices and countries. Sparsity is far greater as you increase granularity because the sales quantity will be lower and, some cases, just 0 on periods.

Be clear with your forecast objective. You can explain to the interviewer the following: “As a starting point, I’m going to presume that the forecast is performed at each item category level because time series trends in category are less sparse. This should serve as a starting point of the forecast system.”

One other vital decision is determining your period window. Suppose you have sales measured at each hourly level. Are you training and forecasting hourly observations? Or, will you aggregate at a daily, weekly, or monthly to generate forecast?

These decision points need to be accomodated based on data availability, business need and problem objective. Generally, sales projection is applied for inventory tracking and financial reporting purpose. Monthly forecast should suffice.

Handle non-stationarity

You don’t need to spend too much explaning this step. Many econometric models presume stationarity of a time-series to train and forecast. You can evaluate the stationarity of a time series based on auto-correlation, partial auto-correlation and/or Dickey Fuller statistical test. Correct the non-stationarity using differencing or de-trend using a linear or polynomial model.

Determine the training window and forecast horizon

The length of your training data size is not necessarily the same as your training window. For instance, you may have ten years worth of daily observation data. But, you may not need to train on the entire ten years worth of data to forecast. You could just use the most recent two years which are sufficient for a model such as the Seasonal Auto-Regressive Integrated Moving Average (SARIMA) to incorporate seasonal patterns in forecasting.

You also need to determine your forecast horizon. Using the same daily observation data, are you forecasting just one day ahead (one observation point)? Or are you forecasting three months out (about 90 observation points)?

Similiar to determining the granularity of the model forecast, you need to define your horizon based business objective and model performance. Engage with the interviewer to refine the business objective.

Algorithm selection and evaluation

You have a myraid of forecasting algorithms to choose:

1. Exponential Smoothing
2. Linear Model
3. ARIMA
4. SARIMA
5. VARMA
6. SARMA
7. Bayesian forecasting
8. LSTM
9. Much more!

Note that the focus of the problem is not compare and contrast among algorithm selections, but to design a forecasting solution. A novice interviewer will invest too much time discussing the algorithm approach but, not system design. Just choose one. Starting with SARIMA is generally an acceptable option for sales forecasting as many items such as clothing and utility goods follow seasonal patterns.

For your evaluation criteria, you can mention using mean absolute error (MAE), mean absolute percentage error (MAPE), or root mean squared error (RMSE). Unless the interviewer asks you to justify your choice, just choose one and move to the next talking point.

Production modeling

The last step differentiates an experienced data scientist or ML engineer from a novice. Many modeling problems use online data, not offline data frozen in time. You need to explain how your model will forecast in real-time, meaning determine the frequency of a model update.

Suppose you forecast on 30 days ahead on daily observations of sales. Will you train your model and generate a new 30-day forecast at the end of each day? Or, will you wait until the 30-days of the current forecast has lapsed, then forecast the next 30-days yet to be forecasted? Again, the forecast decision is based on the business problem.

Some projects will require that you update your forecast as often as possible. Basically, suppose in day 0, you forecast sales from day 1 through 30. The following day, you generate new forecast from day 2 through 31. The day 2 through 30 forecasts already exists from the previous forecast so your new forecast will replace the old.

If the frequency update is not necessary, wait until the current forecasting horizon lapsed then, forecast the next horizon.

Interviewer Assessment

In the statistics section, a candidate is assessed based on correctness and soundness of statistical methodology, product sense and communication. For each dimension the candidate is rated in the following scale: (5) superior, (4) good, (3) adequate, (2) marginal, (1) not competent.

Assessments	Rating	Comments
ML Methodology	5	The candidate possesses a strong grasp in forecasting as he was familiar with how to frame a forecasting problem (i.e. granularity, forecasting horizon) and aware of potential challenges (i.e. data sparsity). This suggests that he is experienced in forecasting problems which is a plus for the role. He also understood the trade-offs on the dimension to forecast (i.e. department vs item). Finally, his initial model design and evaluation criteria were sound.
Product Sense	5	The candidate demonstrated potential challenges in forecasting performed at item level vs department level. He understands how the dimensional granularity reflects the time series patterns. This suggests that he knows how to connect the dot from business/product issues to crucial ML steps involving data cleaning and modeling.
Communication	5	The candidate’s explanations were clear and comprehensive. He explored how he would walkthrough a conversation on item vs. department comparison to a business stakeholder. This suggest that he’s knows how to drive a conversation in data science project scope and goals.