

Demand Forecasting with Matrix Factorization

Increase margins and lower costs through cutting-edge predictive methods.

BENEDEK TOTH - DATA SCIENTIST

Forecasting demand is essential for most businesses, whether in the service or retail industry: from setting prices to determining stock or staffing levels, decision makers must understand how to supply products and services to match customer demand. Often, this analysis must be carried out amidst dynamic conditions. Getting demand planning right is essential for profitability, customer satisfaction and retention. This white paper will introduce latent matrix factorization to model demand curves and discuss how it can be used to achieve these outcomes.

Introduction

Demand forecasting is a quantitative problem. Companies often use simple linear or quadratic models to estimate a demand curve for a single product, looking at historic averages to predict seasonal changes in sales. Simple solutions like these may be sufficient for stable market conditions, but in dynamic situations, they may easily become unreliable. They also frequently fail to meaningfully account for complex relationships between products that are often bought together (the market basket problem). For decades, insufficient data for complex modeling led to the misguided belief that decision makers couldn't do more to improve forecasting. Better methods, such as multinomial logistic regression, have been described for at least twenty years, but they remain limited and require a relatively large amount of specific information about individual products to build a forecast.

Today, companies collect, store and process more data than ever. This has widely created incentives within organizations to use that data, and enticed industry players to build tools and services to enable them to do so. This in turn has led to new methods of analysis. Retailers in particular who can match purchases to customers now enjoy a whole host of modeling possibilities. In the last five years, academics in the field of machine learning, management studies and economics have joined forces to develop new models of customer preferences based on latent matrix factorization.

This method — also referred to as collaborative filtering — is widely used in recommender systems on shopping and entertainment streaming sites. These models account for the intricate relationships between products that are complementary to, or substitutable for, each other, and are able to forecast how customers react to changes in price or the available range of products. Additionally, these new methods explicitly model customer heterogeneity, which makes them suitable to adapt pricing (e.g. by targeted discounts) to specific customers based on demographics or behavior. Such models can also reliably forecast demand changes for items with very little historical data, by relating them to other similar products based on previous customer choices.

This white paper outlines how these models work and why are they able to provide better results than other alternatives. To illustrate this, the results from two academic papers (Athey et al. (2017) and Athey et al. (2019)) are presented, and the kind of data necessary to use these models — and how they can contribute to more efficient pricing and inventory decisions — are

Latent Matrix Factorization-based models and what you can achieve with them

In practice, every single customer has specific preferences and specific price elasticities across the space of all products, as well as across all combinations of products — for instance, few people buy three different sets of napkins at the same time, but most people do buy hot dogs and buns together. Modeling these connections has not been possible previously, both because the scale of the computational problem involved was prohibitive and because data of the required granularity was often unavailable. Even when purchases were tracked and prices were stable, the complexity of customer choices was hard to capture as customers usually buy only a small subset of all available products, exhausting only a small corner of the product permutation space each time.

Enter matrix factorization, a method that identifies latent or unobserved characteristics of both products and customers and combines them to quantify both preferences and price elasticities for every customer-item pair. The model takes advantage of the fact that consumers' price sensitivities can be inferred across items, and that some product attributes span multiple categories — think of sporty, elegant, or organic, for example. Similarly, this method is appropriate to measure the interdependencies of product pairs or product sets. While it may feel like a daunting task, clearly there's a lot of information to be gained from considering all products in a shared space of properties and preferences instead of looking at them one-by-one. This advantage is especially significant in the case of price changes, as these models account for the various price elasticities different consumers have for different items.

Latent matrix factorization-based models also have the ability to predict the ripple effect that price changes have on the demand of related items. With the ability to identify relationships between products, the models forecast the effect of a single price change for all related items. For example, if hot dogs become more expensive, a latent matrix factorization-based model would predict the resulting decrease in both hot dog and bun sales, while predicting an increase in the sales of alternatives like hamburgers.

In many settings, especially in e-commerce, there are thousands or even millions of products, many of which are never purchased by any specific customer. Even if a customer never bought anything from a product category, these models can make personalized predictions about what they might like, based on purchases from an entirely different category. Latent matrix factorization resolves this prediction problem by representing every product by a fixed number of attributes, called latent factors. These factors are shared across all products, do not have a fixed definition and do not need to be determined in advance. Their values are computed based on purchasing patterns (behavior) and capture attributes that may span product categories. This enables us to infer a customer's shampoo preference based on their snack purchases, by examining what brand of shampoo was associated with customers who had the same snack purchase preference.

Models that use matrix factorization also enable retailers to understand what makes their customers come to their store or site in the first place. Some products might be crucially important for specific consumers, and retailers can respond by ensuring that these important items are always available and are priced competitively. This responsiveness is essential for retaining existing customers and attracting new ones. With customer-specific predictions, retailers can choose discounts and offers optimally for any segment, addressing a small demographic or behavioral segment, or even a 'segment of one' at a much lower cost than applying discounts widely.

Modeling can make a huge difference, even if the task is as simple as deciding between two prices. In their 2019 paper, Athey et al. compared the two most common prices for a variety of products. They calculated the average profits earned per shopping trip for each product and price, then ran the model to predict which of the two prices would be better. The model consistently picked the price leading to higher profits, and in some cases the resulting differences were very large. As Figure 1 (below) shows, the profit gain per product regularly exceeded 50%, sometimes even 200%, while profit losses due to the model were both rare and much smaller.

FIGURE 1

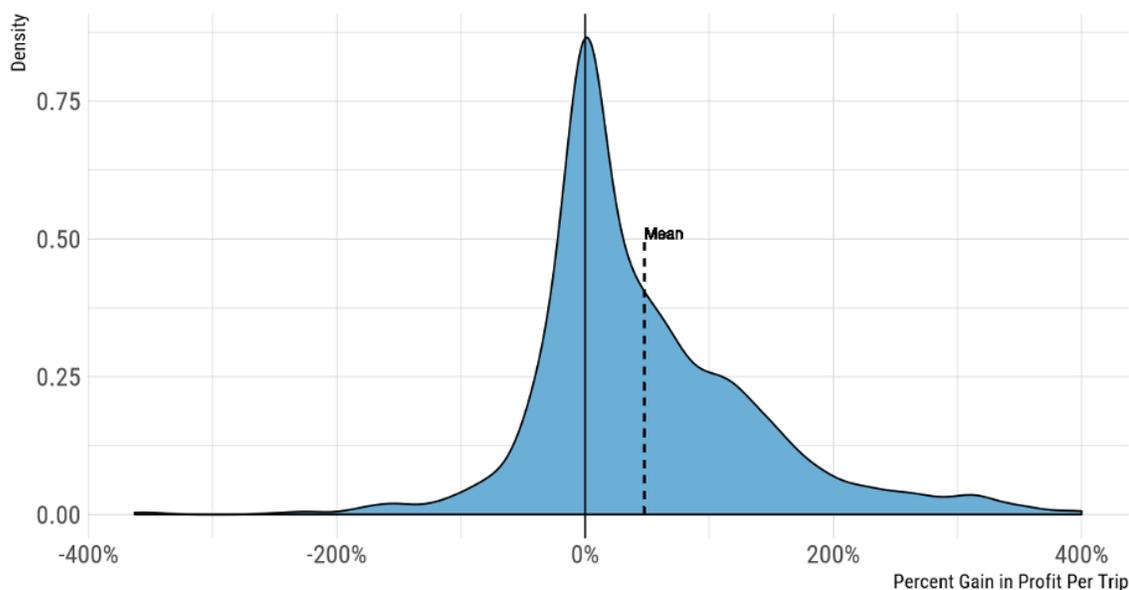


Figure 1. Distribution of percent gains in profit (From Athey et al., 2019)

What are the data requirements?

The key for building such models is the ability to match purchases to individual customers. In the case of brick-and-mortar retailers, this is typically accomplished using loyalty card data, while online stores can directly use customer baskets. In addition, credit card companies may provide useful data of this type. Once we know who bought what, we can leverage these methods and models to the retailer's advantage. It is important that such models require a Single Customer View, as it is the customer's preferences that affect future purchasing decisions.

Thus, the underlying data structure must be able to join multiple transactions to the same customer in order to meaningfully predict future purchasing decisions on an individual level and obtain the outstanding forecasting precision these models exhibit.

How do Latent Matrix Factorization models work?

Understanding the heterogeneity of consumption choices starts from the premise that every customer has specific preferences and specific price elasticities across all products, and virtually all products have complex complementary or substitution dependencies. These relationships can be mathematically formulated in a predictive model of consumer choice. This model would require so many parameters that they could not be reliably estimated, even based on the largest datasets. The solution to this problem is to understand these parameters as combinations of a limited number of latent factors. This way, we do not capture single relationships in individual parameters but rather represent every product as a point in a latent factor space, which has as many dimensions as we have latent factors. Relationships are then estimated over these latent factors.

Purchases need to be stored on the item level and historical price data is needed for demand forecasting. To better understand the relationships between products, it's advantageous to track the periods when each item was out of stock — as these occasions provide natural experiments to observe substitution patterns —, but this information is not essential.

In the case of customer preferences, for example, it's impossible to estimate every consumer's specific attitude toward every product. However, it's possible to define latent product characteristics and estimate those characteristics for every product when given a sufficient amount of historical purchase data. The number of these characteristics is predefined, yet their content is not. The values of these characteristics are learned from the data and captured in complex product properties in a limited number of variables. Similarly, consumer preferences, relationships between products and factors determining price elasticities — both on the product and consumer level — can all be modeled this way.

This matrix factorization method is analogous to that of collaborative filtering algorithms used for recommender systems. This method was made famous by the Netflix Prize challenge and relies on decomposing the user-item interaction matrix into the product of two lower dimensionality matrices, drastically decreasing the number of parameters in the model.

In a retail application, Athey et al. (2017) show that a two-dimensional t-SNE projection of the product features learned by the model retain intuitively meaningful properties with items that belong to the same general category located close to each other on the projection (see Figure 2).

FIGURE 2

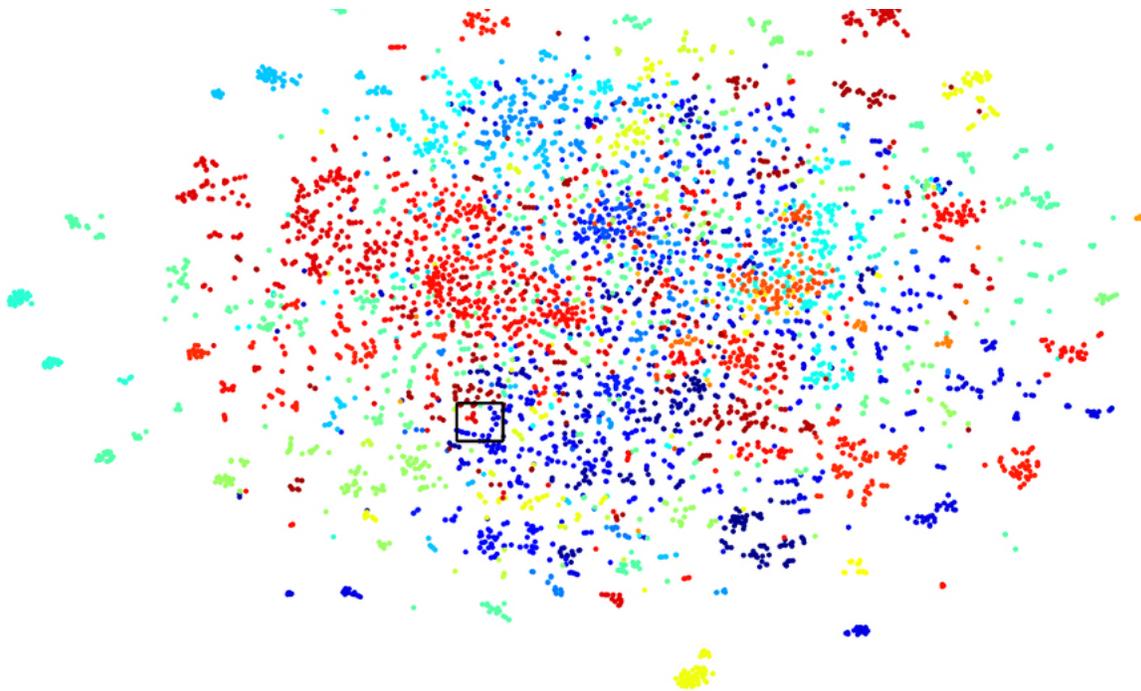


Figure 2. Products in a two-dimensional projection of their latent factors (From Athey et al., 2017)

Conclusion

Latent matrix factorization models provide more precise forecasts and lead to optimal pricing and inventory decisions. This enables firms to find 'sweet spots' in their market and increase their profits, while decreasing the cost of discounts by limiting those to small segments that are most likely to be susceptible to them. Accounting for the intricate relationships between the demand for different items, companies can avoid poor choices in pricing and the selection of available products. These methods can even find the most profitable targeted discounts for consumer groups or individuals. An academic study (Athey et al., 2019) found that even the two most common prices of a product can lead to drastically different profits on that item: choosing a better price with an advanced model led to an average profit increase of over 50%. Companies that choose to use matrix factorization-based models can expect to increase profits and customer satisfaction at the same time.

References

Francisco J. R. Ruiz, Susan Athey, David M. Blei, 2017. "SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements," Papers 1711.03560, arXiv.org, revised Jun 2019.

Rob Donnelly, Francisco R. Ruiz, David Ble, Susan Athey, 2019. "Counterfactual Inference for Consumer Choice Across Many Product Categories," Papers 1906.02635, arXiv.org.

ABOUT THE AUTHOR



Benedek Tóth

Benedek Tóth, data scientist at Starschema, specializes in time series and geolocation data analytics for the telecommunications and manufacturing sectors. He holds an MA degree in Economics from Central European University. He is also an economics instructor at Corvinus University and Eötvös Loránd University in Budapest. Before joining Starschema, Benedek worked at a management consulting firm focusing on retail banking and financial innovations.

COMPANY

At Starschema we believe that data has the power to change the world and data-driven organizations are leading the way. We help organizations use data to make better business decisions, build smarter products, and deliver more value for their customers, employees and investors. We dig into our customers toughest business problems, design solutions and build the technology needed to compete and profit in a data-driven world.

GET IN TOUCH

- info@starschema.com
- starschema.com