RESEARCH ARTICLE OPEN ACCESS

Refining Predictions on Customer's basket with Cross Elasticity

Ananya Srivastava [1], BagyaLakshmi V [2], Ram Rajagopalan [3], Balaji Ramachandramurthy [4], Saloni Kamboj [5]

[1][2][3][4][5] Tata Consultancy Services Pvt. Ltd, Chennai - India

ABSTRACT

Cross-price elasticity measures how sensitive the demand of a product is over a shift of a corresponding product price. Often, in the market, some goods can relate to one another. This may mean a product's price increase or decrease can positively or negatively affect the other product's demand. In this paper we concentrate on cross-elasticity of substitutes and complements to study their impact on Quantity Demand of the product. Our analysis includes price and cross elasticity on both substitutes and complements in demand and in supply to cover the entire market.

Keywords: - Cross Elasticity, Price Elasticity, Recommendation System, Association rule mining, Regression techniques.

I. INTRODUCTION

In economic terms, cross elasticity of demand is the responsiveness of demand for a product in relation to the change in the price of another related product. The relevant word here is "related" product. Unrelated products have zero elasticity of demand. The most important concept to understand in terms of cross elasticity is the type of related product. The cross elasticity of demand depends on whether the related product is a substitute product or a complementary product.

Substitute products are goods that are in direct competition. An increase in the price of one product will lead to an increase in demand for the competing product. For instance, an increase in the price of petrol will force consumers to go for diesel and increase the demand for diesel. Now, the cross elasticity value for two substitute goods is always positive.

Complementary goods, on the other hand, are products that are in demand together. An ideal example would be coffee beans and coffee paper filters. If the price of coffee increases, then the demand for filters would reduce because the demand for coffee will reduce. The cross elasticity of demand for two complementary products is always negative.

In this Paper, We will see the quantity impact on related product if there is increase/ decrease in one product.

II. RELATED WORK

It is important to know how consumers will react when prices change. The Cross-Price Elasticity of Demand (XED) measures this relationship between two goods. It determines if they are more or less sensitive, providing valuable insight for companies optimizing their pricing strategy in different markets with varying economic conditions.

In this study, we have defined the methodology for obtaining the cross/price Elasticity of Demand. So, initially we have divided this into PART A and PART, where PART A is "Finding the complementary products" and PART B is "Finding the substitute products". If the cross-elasticity result comes out as positive(+ve) then it is substitute and if it is negative(-ve) it will be complementary item. Let's discuss about PART A "Finding the Complement". So below are the steps performed:

- We required the customer transactional Dataset.
- Finding the complementary item using the association rule mining techniques - Apriori/fp-growth/fpmax/eclat.
- Sorting and considering top 10 complement products.
- Applying regression techniques such as Linear/Ridge/Lasso to first check the Price elasticity and then getting the Cross Elasticity. (Price Elasticity will give the output if the price of Product A increases what will be its demand but in cross elasticity we will consider two products and it will check which if product A's price is increased/ decreased what will be the quantity impact on Product B)

PART B - "Finding the Substitute using Recommendation system":

- We required dataset having the product content i.e, Product, category, ingredient.
- Using the recommendation system techniques like collaborative filtering/ content-based filtering/ lightfm to find Substitute products.
- Sorting and considering top 10 complement products.

- Applying regression techniques such as Linear/Ridge/Lasso to first check the Price elasticity and then getting the Cross Elasticity. (Price Elasticity will give the output if the price of Product A increases what will be it's demand but in cross elasticity we will consider two products and it will check which if product A's price is increased/ decreased what will be the quantity impact on Product B).

III. METHODOLOGY

In this study, we had the unlabelled dataset which was consists of attributes such as Invoice ID, Date, order ID, Product ID, Customer ID, Quantity, Product ID, Category, Sub-Category and Price. It is very to important to have Category if we need to find the frequently bought items. Based on the Frequently bought item, our aim is to find the complementary products and their Price/Cross-Elasticity.

Secondly, We required to "Find the Substitutes", which can be done by using Recommendation systems. Below steps will provide the deep insight about the process and how it is possible to identify the Quantity impact in E-commerce domain.

A. Dataset:

In this, the customer provided the Unlabelled Dataset and for implementing the techniques we required the major attributes such as Product ID, invoice ID, Product Name, Category and Price. It was basically a superstore dataset and customer requirement to firstly find the complementary items and their quantity impact on complement products. Next, they required to find the substitute itemset and their respective quantity impact.

B. Pre-Processing:

Pre-processing is the first step to convert unstructured/ unlabelled data into structured form. Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations. So here, we performed some Data Visualization based on the Dataset to check if we have any outliers or null values. Also, Data Visualization based on Sales vs Qty, Price Vs Quantity, Sales per hour/day/month etc. We Counted the Quantity sale per category and so on..

C. Method:

Part A - It was exactly to "find the complementary itemset and their quantity impact":

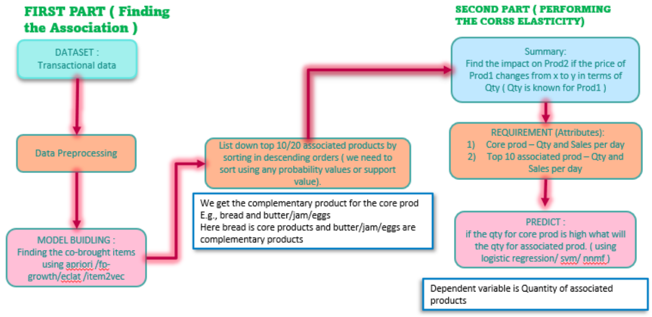

Fig. 1 Technical Architecture for Part A – "Finding the Complementary Items"

In Fig 1, the architecture explains the roadmap which was followed to perform cross elasticity by finding the complementary items.

Association Rule is one of the very important concepts of machine learning being used in market basket analysis. Market Basket Analysis is the study of customer transaction databases to determine dependencies between the various items they purchase at different times. Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It identifies frequent if-then associations called association rules which consists of an antecedent (if) and a consequent (then). For example: "If tea and milk, then sugar" ("If tea and milk are purchased, then sugar would also be bought by the customer")

- Antecedent: Tea and Milk
- Consequent: Sugar.

Apriori Algorithm has three parts:

- Support
- Confidence
- Lift

Support (I)= (Number of transactions containing item I) / (Total number of transactions)

Confidence (I1 -> I2) = (Number of transactions containing I1 and I2) / (Number of transactions containing I1)

Lift (I1 -> I2) = (Confidence (I1 -> I2) / (Support(I2))

Support: This says how popular an itemset is, as measured by the proportion of transactions in which an itemset appears.

Confidence: This says how likely item Y is purchased when item X is purchased, expressed as {X -> Y}.

Lift. This says how likely item Y is purchased when item X is purchased, while controlling for how popular item Y is.

Once the above output is generated, we sorted the based-on confidence value so that we can item having most of the complements items purchased together. With most goods, an increase in price leads to a decrease in demand – and a decrease in price leads to an increase in demand. When there is a large change in demand after a price change, that good is considered to have 'elastic demand.'

On the other hand, if there is only a small change in demand, that good is considered to have relatively inelastic demand. So, we tried by increasing and decreasing the price of a product and saw the drastic impact on quantity. When the price was increasing the demand was decreasing and when the price was decreasing the demand was increasing.

Similarly, we implemented the cross elasticity by sorting the mostly purchased itemset and then by applying the Regression technique which achieved the desire outputs. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. It performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output).

The simplest form of a simple linear regression equation with one dependent and one independent variable is represented by:

$$y = mx + C$$

where, we assumed the m is our slope and C is the intercept.

So if the slope values comes out a negative (-ve), it truly states that the Product is Complement. So we applied multiple Regression models here for Part A "Finding the complements Items and their Price/Cross Elasticities". Such as Linear Regression, Ridge Regression and Lasso Regression.

Price Elasticity of Demand = Percentage Change in Quantity ($\Delta q/q$) / Percentage Change in Price ($\Delta p/p$).

It is of paramount importance for a business to understand the concept and relevance of price elasticity of demand to understand the relationship between the price of a good and the corresponding demand at that price. Price elasticity of demand can decide the pricing policy for different markets and various products or services.

Part B - "Finding the Substitute and Their Cross/Price Elasticity"

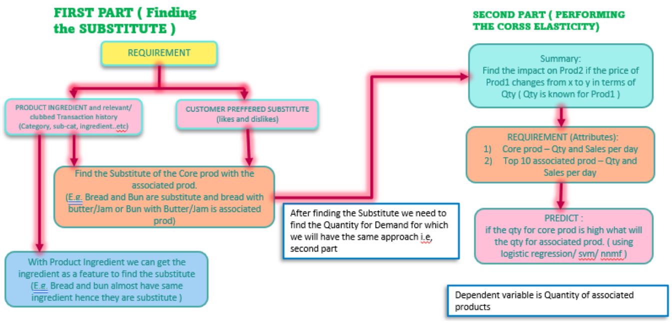

Fig. 1 Technical Architecture for Part B – "Finding the Substitute Items"

Fig 2 provides the overall structure on how we achieved the Price/ cross elasticity on substitute items.

Substitute goods or substitutes are at least two products that could be used for the same purpose by the same consumers. If the price of one of the products rises or falls, then demand for the substitute goods or substitute good (if there is just one other) is likely to increase or decline. So, to find out the substitute items we approached recommendation systems using Content-Based Filtering which tries to guess the features or behaviour of a user given the item's features, he/she reacts positively to. An excellent example of how content filtering works is identifying specific patterns in content, such as strings of text or objects in images, that signify the presence of undesirable content.

In our case the content was ingredients and we made a use of category, so the model predicts better outcomes. In this method, the algorithm is trained to understand the context of the content and find similarities in other content to recommend the same class of content to a particular user.

1. It begins by identifying the keywords to understand the context of the content. In this step, it avoids unnecessary words such as stop words.

2. Then it finds the same kind of context in other content to find the similarities. To determine the similarities between two or more contents, the content-based method uses cosine similarities.

3. It finds similarities by analysing the correlation between two or more users.

4. Then finally it generates recommendations by calculating the weighted average of all user ratings for active users.

So, in this part, we applied TF-IDF for finding similar posts for users based on their Purchases. Here TfidfVectorizer is used to create raw documents to a matrix of TF-IDF features.

TF*IDF is an information retrieval technique that weighs a term's frequency (TF) and its inverse document frequency (IDF). Each word or term that occurs in the text has its respective TF and IDF score. The product of the TF and IDF scores of a term is called the TF*IDF weight of that term. Put simply, the higher the TF*IDF score (weight), the rarer the term is in a given document and vice versa.

The next step will be to use the sigmoid kernel function. The function sigmoid_ kernel computes the sigmoid kernel between two vectors. The sigmoid kernel is also known as a hyperbolic tangent, or Multilayer Perceptron (because, in the neural network field. Sigmoid kernel works similarly the way Logistic Regression works. The sigmoid function is a mathematical function having a characteristic "S" — shaped curve, which transforms the values between the range 0 and 1. The sigmoid function also called the sigmoidal curve or logistic function. It is one of the most widely used non- linear activation function.

Now, the next comes defining the function for recommending products so we are taking product id and sigmoid values. Next, line of code will give the indexes of corresponding product name and then we got the pairwise similarity scores and after this we are sorting the products in descending order and took the score of top 10 products (we can set the threshold as per the requirements).This is how our user-defined function was generated for top 10 recommendations products.

IV. RESULT AND DISCUSSION

When it comes to the evaluation section, we have RMSE (Root mean squared error) The Root Mean Square Error (RMSE) (also called the root mean square deviation, RMSD) is a frequently used measure of the difference between values predicted by a model and the values observed from the environment that is being modelled. These individual differences are also called residuals, and the RMSE serves to aggregate them into a single measure of predictive power. To phrase it another way, RMSE is a good way to answer the question: "How far off should we expect our model to be on its next prediction?"

If the noise is small, as estimated by RMSE, this generally means our model is good at predicting our observed data, and if RMSE is large, this generally means our model is failing to account for important features underlying our data. So, when we evaluated the results using the rmse score, Lasso outperformed with minimum RMSE Score in PART A and for cross-elasticity in PART B, BayesianRidge Model provided the minimum RMSE score. There are few advantages of RMSE:

- RM-SEs avoid the use of absolute value, which is highly undesirable in many mathematical calculations.
- The underlying assumption when presenting the RMSE is that the errors are unbiased and follow a normal distribution.
- Giving higher weighting to the unfavourable conditions, the RMSE usually is better at revealing model performance differences.
- It also provides the loss function of the second degree.
- In addition, we demonstrate that the RMSE satisfies the triangle inequality required for a distance function metric.

V. CONCLUSION

In the present world of digitally connected world every shopping mall desires to know the customer demands in advance to avoid the shortfall of sale items in all seasons. Day to day the companies or the malls are predicting more accurately the demand of product sales. Extensive research in this area at enterprise level is happening for accurate Demand Forecasting. As the profit of a company is directly proportional to the accurate predictions of Demand.

To deal with demand prediction, various techniques of regression analysis and data mining are used under the predictive methods. The purpose of this work is to make history-based demand prediction of sales by using generalized linear models.

Regression Analysis forecasting is meant for those companies that need in-depth, granular, or quantitative knowledge of what might be impacting sales. It easily helped us to achieve our goals by simply using the regression technique and getting the Quantity impact due to cross elasticity as well as Price elasticity. Next, we will implementing more Advance techniques such as Random-Forest Regression, K-Neighbors Regressor, Neural Networks etc.

- Regression Techniques performs well when the dataset is linearly separable. We can use it to find the nature of the relationship among the variables.
- Regression Techniques is easier to implement, interpret and very efficient to train.
- Regression Techniques is prone to over-fitting but it can be easily avoided using some dimensionality reduction techniques, regularization (L1 and L2) techniques and cross-validation.

REFERENCES

[1] Richa Richa, Market Equilibrium: A Cross Elasticity approach , 2019.

[2] Philip Graves, Robert L. Sexton, Cross-price Elasticity and Income Elasticity of Demand: Are your students Confused?, The American economist, Oct. 2009.

[3] Lorenzo Sabatelli, PhD, Relationship between the Uncompensated Price Elasticity and the Income Elasticity of Demand under Conditions of Additive Preferences, GLOBMOD Health, Market Analysis Unit, Barcelona, Spain, March 21, 2016.

[4] Chengcheng Liu, Mátyás A. Sustik , Walmart Labs, San Bruno, CA, Elasticity Based Demand Forecasting and Price Optimization for Online Retail, June 2021.

[5] Mr. Yash Ketan Bhanushali, Mr. Yash Shankarbhai Patel, Movie Recommendation System, IRJET, Volume: 08 Issue: 04, Apr 2021.

[6] Niya N J, Jasmine Jose, Sale prediction using Linear Regression Model, IJCRT , Volume 9, Issue 3 March 2021.

[7] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research 12, arXiv, January 2012.

[8] Santosh Kumar Ray, Seba Susan, Performance Evaluation using Online Machine Learning Packages for Streaming Data, International Conference on Computer Communication and Informatics (ICCCI), January 2022.