Lifting customer basket value using market basket and product affinity analysis

Tomaž Kaštrun

7.11.2013

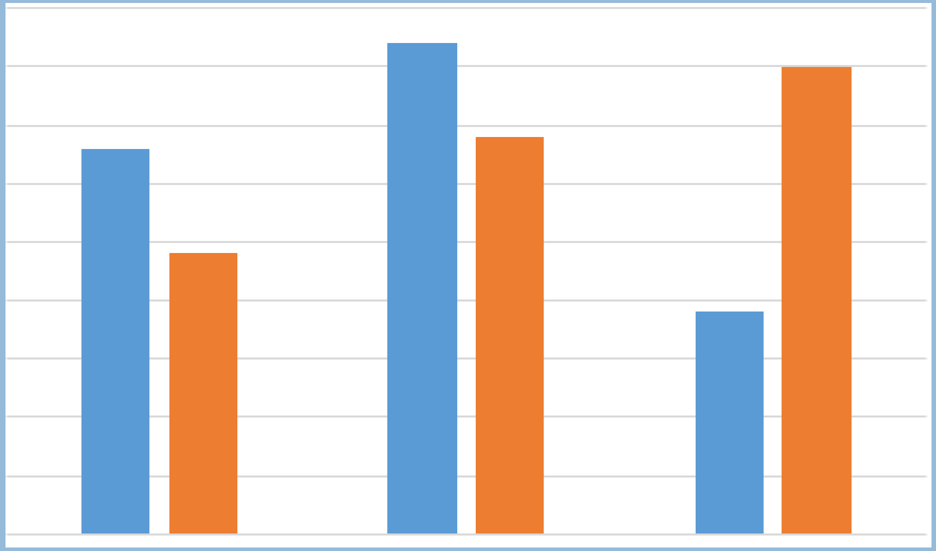
Budapest BI Forum

Introduction

Tomaž Kaštrun is BI developer at Studio Moderna and main focus are data mining (R,SPSS), SQL and .NET programming. He has been working with SQL server since version 2000 with focus on Analysis services.

Studio Moderna is leading multi-channel and e-commerce company with direct to consumer strategy situated in Central & Eastern Europe, with a vertically-integrated network reaching more than 400m consumers across 21 countries.

Common Solution to old problem



OR

	Q-2012	Q-2013
Lettuce	29	28
Milk	42	34
Coffee	19	40

Introduction to Customer Basket Value

- Market Basket Anaylsis (MBA)
- Analysis of „items being purchased together"
- Identifying purchase patterns
  - What items are purchased together
  - What items are purchased sequentially
  - What items are purchased frequently
  - What items are purchased periodically (seasons, special events,…)

Customer Categorization

- Categorizing customer behaviour
  - One-time buyer
  - More-time buyer – Rebuyer (within 1-2 years)
  - Loyal/Frequent Buyer
  - Returning Buyer (Reactivated buyer – with some special offers or campaign; after longer period)

Selling Items

- Identifying what sells?
  - Top quality products,
  - New products,
  - good service,
  - coupons,
  - free shipping,
  - gifts,
  - etc.
- Finding the perceived value of items
- Calculating profitability of each selling item (COGS, Gross Margin)

Identifying what drives the sale

- Profiling your customers
- Profiling your sales
- Customer interact with products; products interact with customers.

Know your marketing strategies and actions

- Use customer behaviour patters when creating actions
- Use business rules and company strategy when preparing marketing actions
- Analyse your
  - Print campaigns
  - Television ads
  - Jumbo ads
  - Internet visitors
  - Telemarketing specifics and agent climate

Customer centrism and Omnichannel

- Customer purchase experience through one or more channels
- Campaigns should have omnichannel customer view
- Possible Market baskets are different depending on channel
- Using omnichannel data yield better overview of customer potential needs
- Use business rules and company strategy when preparing marketing actions

EXAMPLE

Customer Basket items

- Five Products: Matress, Sleepers, Frying Pan, Walking Shoes, Camping Tent
- Six Customers:
  - Customer #1: Sleepers, Frying Pan
  - Customer #2: Walking Shoes, Camping Tent, Sleepers (CrossSell)
  - Customer #3: Camping Tent, Frying Pen
  - Customer #4: Matress, Sleepers
  - Customer #5: Walking Shoes, Sleepers, Frying Pan (CrossSell)
  - Customer #6: Frying Pan, Matress

Customer Basket items

- Five Products: Mattress, Sleepers, Frying Pan, Walking Shoes, Camping Tent
- Six Customers:
  - Customer #1: Sleepers, Frying Pan
  - Customer #2: Walking Shoes, Camping Tent, Sleepers (CrossSell)
  - Customer #3: Camping Tent, Frying Pen
  - Customer #4: Mattress, Sleepers
  - Customer #5: Walking Shoes, Sleepers, Frying Pan (CrossSell)
  - Customer #6: Frying Pan, Mattress

Product Frequency Matrix

	Mattress	Sleepers	Frying Pan	Walking Shoes	Camping Tent
Mattress	2	2	1	0	0
Sleepers	2	4	2	1	0
Frying Pan	1	2	4	1	1
Walking Shoes	0	1	1	2	1
Camping Tent	0	0	1	1	2

Generating Rules (SUM OF: Product *{support}*):

1) Mattress (4x) + Sleepers (2x) -> Profile: *Sleeping kings*
2) Frying Pan (4x) + Sleepers (2x) -> Profile: *Home cosy lovers*
3) Walking Shoes (2x) + Frying Pan (1x) -> Profile: *Weekend out-goers*
4) Camping Tent (2x) + Walking Shoes(1x) -> Profile: *Nature campers*
5) Walking Shoes (2x) + Camping Tent (1x) + Frying pan (1x) + additional rules-> Profile: *Summer camping vacationers*
6) etc...

Extending purchase products with profile

<u>Profile</u>: *Sleeping Kings*

Core Products: Mattress, Sleepers

Adding more products to this profile (ordered by importance):

- Bed covers
- Bed Sheets
- Pillows
- Bathrobe
- Towels

Caution: Aware of over-prunning the profile

Extending purchase habits with profile

- Profile: *Sleeping Kings*

  Setting profile with demographic data:
  - Average on age, income, number of children
  - Modus on gender, education, marital status
  - Highly likely regions(cities)

  Setting profile with economic data:
  - Most likely RFM region
  - Loyalty program

  Setting profile with activity data:
  - Internet browsing history
  - Telephony data
  - Television preferences
  - Catalog activities

Market Basket and profiles

- Combining profiles with economical data gives value of Market basket for each profile
- Easier setting customer and their profiles to specific campaigns
- Targeting right people for the right campaign
- Three way comparison
  - Using customer proximity (similarity)
  - Using product proximity (similarity)
  - Using rules proximity (similarity)

Building Models

- Profiling customers based on Clustering techniques and analysis of variance
- Product affinity based using HOMALS, Factorial analysis and Naive Bayes for product proximity and similarity
- Extending cross-selling products based analysis of Association Rules

Data Overview

- Transactional and customer data (from CRM system)
- Single observation can be customer or purchase (invoice)
- Data resides in database (MSSQL Server)
- Additional data usage presents additional load of extracting and loading data

DEMO SQL Server and R using Association rules

Using Association Rules in SQL Server 2012

- Two relational tables that reside in SQL database as a source of data with nested tables:

  [dbo].[vAssocSeqOrders_BIFORUM],
  [dbo].[vAssocSeqLineItems_BIFORUM]
- Defining key values, Input values and predict values
  - Key value is a primary key defining these values
  - Predict values are key that the model will be predicting
- Defining data type and content type (discrete, continous, key values)
- Defining Algorithm parameters (min_probability, min_support)

Using DMX in SQL Server (1)

```
-- Query for Predicting Associated Items for Bed Cover
SELECT
        PredictAssociation([Association].[vAssocSeqLineItems],INCLUDE_STATISTICS, 3)
FROM [Association]
                NATURAL PREDICTION JOIN
(
        SELECT
                (SELECT 'Bed Cover' as [Model])
        AS [v Assoc Seq Line Items]
)
AS t
```

Using MDX in SQL Server (2)

```
-- Returning the confidence for Related ItemsSets for Items Sleepers
SELECT TOP 50
FROM
        (  SELECT
                 FLATTENED NODE_CAPTION
                ,NODE_SUPPORT
                ,(SELECT ATTRIBUTE_NAME
                        FROM NODE_DISTRIBUTION
                        WHERE ATTRIBUTE_NAME = 'v_Assoc Seq Line Items(Sleepers)'
                ) AS D
         FROM Association.CONTENT
         WHERE
                NODE_TYPE = 7
) AS Items
WHERE [D.ATTRIBUTE_NAME] <> NULL
ORDER BY NODE_SUPPORT DESC
```

Using Association Rules in R

- One flattened tables that reside in SQL database as a source of data
- Defining key values, Input values and predict values
  - Key value is a primary key defining these values
  - Predict values are key that the model will be predicting
- Defining Algorithm parameters (min_probability, min_support)

Using R

```
# Creating Association Rules
AR_rules <- apriori(fo)
inspect(AR_rules)
AR_rules <- apriori(fo, parameter = list(minlen=1, supp=0.015, conf=0.85),
                    + appearance = list(rhs=c("Product=Mattress"), default="lhs"),control = list(verbose=F))
rules.sorted <- sort(AR_rules, by="lift")
inspect(AR_rules.sorted)
```

Comparing SQL Server and R

Practical differences between both environemnts

SQL Server	R
Relational and transactional data	Data serialization / Vectorization
Massive in-memory data processing	In memory (with limitation)
Can handle wast database	Will have problems with wast database
No external connections	OLEDB connection to database
Limited statistical approaches	Unlimited statistical approaches
Internal SQL database	Usable with other environments
Limited data modelling (DMX)	No limitation on data modelling
Limited data visualization	No limitation on data visualization

Conclusion

- Quick method for finding patterns
- Robust models
- Models prunning.
- Overcomplexity will happen quickly
- Using simple computations
- No hypotesis setting / testing needed
- Prerequisits: Meaninfull product classification, knowledge on regressional modelling and Naive Bayes

THANK YOU!

Tomaz Kastrun

Tomaz.kastrun@studio-moderna.com
@tomaz_tsql

www.studio-moderna.com