Implementing Social Group Shopping using Support Vector Machines

Prasanna Kothalkar
Nexsus Techno Solutions Pvt. Ltd.
Mumbai, India

Priyanka Desai
Thakur College of Engineering and Technology
Mumbai, India

ABSTRACT:

Social media is currently the center of technological innovations and research. The plethora of actionable data made available by modern social networks has brought forth the need for use of intelligent algorithms that can process such volumes of data.

Group shopping websites are one of the innovations of such social existence of the web. Many websites such as groupon, livingsocial, dealster, buywithme etc. currently offer some form or the other of group shopping.

The paper presents a model of applying SVM algorithm to our case of group shopping with two aims:

i) Predicting potential customers for a given product which shall enable us to launch group shopping campaigns more effectively or consider whether they should be launched at all, in the first place.

ii) Rather than having open ended campaigns, implementing targeted marketing.

An application that implements the above is also presented. If the admin of the website realizes the potential of a product sold by the site or a vendor selling products and services complementary to the social shopping site, he or she may choose to launch the campaign. On doing so, the deal shall be available to all and the potential customers will be specially notified.

General Terms:

Machine Learning, Support Vector Machines, Advanced Computer Application, Artificial Intelligence, Intelligent Systems

Keywords:

Social Shopping, Group Shopping, Artificial Intelligence, Intelligent Systems, Machine Learning, Collective Intelligence, Support Vector Machines, Intelligent Systems

1. INTRODUCTION:

1.1 Group Shopping

The concept behind group shopping can be explained as follows:

1.1.1 Premise

A product or service is offered at a highly discounted rate for a limited time, under the condition of minimum number of purchases.

1.1.2 Action

A user interested in the product purchases it by logging into the website and entering his personal information related to the purchase. (E.g. entering credit card information)

1.1.3 Condition

However, this is a one way purchase as the deal goes through only when the minimum number of users required for the deal purchase the product. If the minimum number of users is not met the deal is cancelled and the purchase amount is not deducted from the user's account.

Knowledge of users and their preferences is of prime importance in formulating such deals. Knowing which users to target for specific products and then pursuing them aggressively rather than only having open deals bombarded at all users is the motivation behind the analysis of user data.

Although Retail specific customer classification models based on Support Vector Machines have been presented in [1] and [2], we are studying a problem that is a sub domain of the same. The size of data available in such situations is limited as this is a fairly new field unlike the larger domain of retail where historical transactional data is available in data warehouses. The classification in our model is largely based on user-product association which results in our derivation of potential customers.

Due to the lack of shopping related user data in the format required for our study, we have created a sample application that helps us populate the data and also acts as a demonstration of how applications in group shopping domain can be modeled and developed along the lines of intelligent applications found in [3], [4] and [5]. However, this has limited the scale and quantity of the data available. We refer to the application as 'Social Shopping'. Also, we have no restriction on the number of deals unlike many group shopping sites, which have one deal per day.

2. DEVELOPING THE MODEL

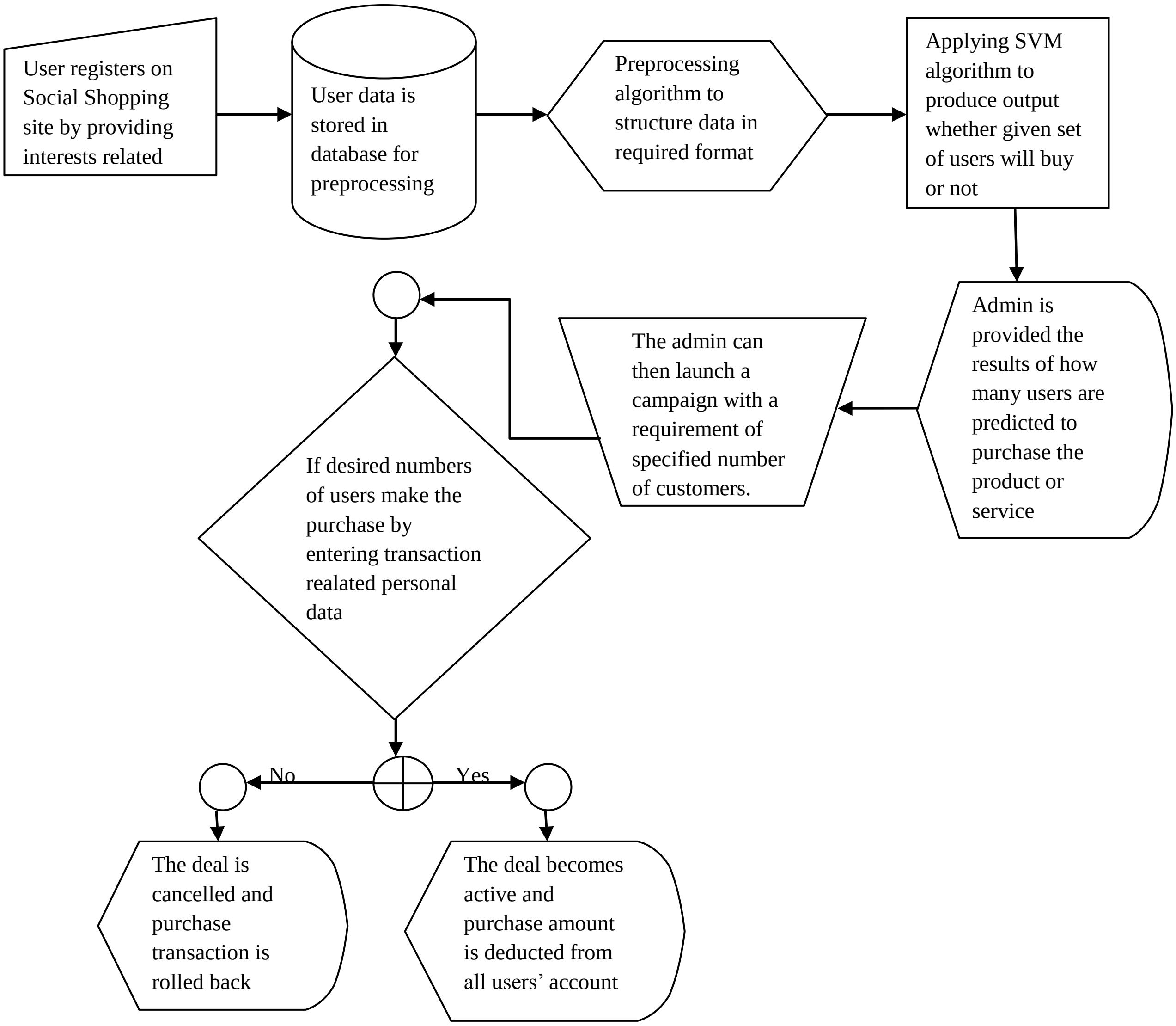

Figure 1: SVM Model for Group Shopping

The format of data available which shall be used for our analysis is as follows:

2.1 For user:

2.1.1 Categories purchased in

2.1.2 Categories interested in

2.1.3 Wish list products

2.1.4 Age

2.1.5 Sex

2.1.6 Annual Income

2.1.7 Location (For location sensitive deals)

2.2 For products:

2.2.1 Product category

2.2.2 Age Category

2.2.3 Gender related product classification (such as men's and women's clothing and personal care products etc.)

2.2.4 Product price category

2.2.5 Product or service location (if location sensitive)

2.3 Rules for generating training data

2.3.1 Wish list trumps all (W)

If a product is in the user's wish list we consider the user to be a potential customer irrespective of the other parameters.

2.3.2 Deal breakers- Gender (d_g), Age (d_a), Income Range (d_i), and Location (d_l) (for location sensitive deals)

Comparing user and product attributes on above parameters we predict if a user is a potential customer

2.3.3 Other parameters

Categories interested in Categories interested in **(C_i)**, Categories purchased in **(C_p)**. A positive in the above against a given product makes the user a potential customer albeit the deal breaker conditions.

The rules for generating output class in our training data can be represented mathematically as follows:

$$\mathbf{Y = S \cdot W - w_g \cdot d_g - w_a \cdot d_a - w_i \cdot d_i - w_l \cdot d_l + C_i + C_p}$$

where $\mathbf{w_g}$, $\mathbf{w_a}$, $\mathbf{w_i}$, $\mathbf{w_l}$ are the weights assigned to Wish list & each of the deal breaker parameters ($\mathbf{d_g}$, $\mathbf{d_a}$, $\mathbf{d_i}$, $\mathbf{d_l}$) respectively.

Based on conditions, the training data is generated in format to work with libsvm. For applying SVM to our group shopping concept related data we do not perform scaling as it is done for most datasets for which SVM is applied. In fact, even though the data can be represented in binary format, we provide appropriate weights to each feature. This enables us to provide relative importance to different features. For example, in our case we provide largest weight age to wish list, because if a product is in a person's wish list, we assume highest probability of the person purchasing it.

Support vector Classification provides an excellent off the shelf machine learning algorithm, which can be used to generate predictions for our testing data.

2.4 Support Vector Machine Algorithm:

Support vector machines are used for classifying huge amounts of high dimensional data into different classes.

We are given some training data $\mathcal{D}$, a set of n points of the form $\mathcal{D} = \{(x_i, y_i) \mid x_i \in \mathbb{R}^p, y_i \in \{-1,1\}\}_{i=1}^{n}$ where the y_i is either 1 or −1, indicating the class to which the point X_i belongs. Each X_i is a p-dimensional real vector. We want to find the maximum-margin hyperplane that divides the points having $y_i = 1$ from those having $y_i = -1$. Any hyperplane can be written as the set of points **X** satisfying $w \cdot x - b = 0$ where $\cdot$ denotes the dot product and **w** the normal vector to the hyperplane. The parameter $\frac{b}{\|w\|}$ determines the offset of the hyperplane from the origin along the normal vector **w**. We want to choose the **w** and b to maximize the margin, or distance between the parallel hyperplanes that are as far apart as possible while still separating the data. These hyperplanes can be described by the equations $w \cdot x - b = 1$ and $w \cdot x - b = -1$

2.4.1 Primal form

Minimize (in w, b)

$$\frac{1}{2}\|w\|^2$$

subject to (for any $i = 1, \ldots, n$)

$$y_i(w \cdot x - b) \geq 1$$

Expressing the previous problem in Langragian form by means of non-negative Lagrange multipliers α_i as

$$\min_{w,b} \max_{\alpha} \left\{ \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i [y_i (w \cdot x_i - b) - 1] \right\}$$

2.4.2 Dual form

Writing the classification rule in its unconstrained dual form reveals that the maximum margin hyperplane and therefore the classification task is only a function of the support vectors, the training data that lie on the margin. Using the fact, that $\|w\|^2 = w \cdot w$ and substituting $w = \sum_i \alpha_i y_i x_i$ one can show that the dual of the SVM reduces to the following optimization problem:

Maximize (in α_i)

$$\tilde{L}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j}^{n} \alpha_i \alpha_j y_i y_j k(x_i, x_j)$$

subject to (for $i = 1, \ldots, n$)

$\alpha_i \geq 0$ and to the constraint from the minimization in b

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

Here the kernel is defined by

$k(x_i \cdot x_j) = x_i \cdot x_j$ and $w = \sum_i \alpha_i y_i x_i$

For non linear classifiers we apply kernel trick to the maximum margin hyperplane. Every dot product is replaced by non linear kernel function. Some kernel functions are:

Polynomial:

$$k(x_i \cdot x_j) = (x_i \cdot x_j)^d$$

Gaussian radial basis:

$$k(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2)$$

Hyperbolic tangent:

$$k(x_i, x_j) = \tanh(kx_i \cdot x_j + c)$$

First step in non linear classification is choosing the kernel and hence the mapping $x \rightarrow \phi(x)$

Create H where

$$H_{ij} = y_i y_j \phi(x_i) \cdot \phi(x_j)$$

Maximize

$$\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \alpha^T H \alpha$$

subject to constraints

$0 < \alpha_i \leq C \;\; \forall_i$ and

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

A more detailed explanation of the algorithm can be found in [6], [7] , [8] ,[9], [10] and [11].

3. EXPERIMENTAL RESULTS:

We have applied the problem for the product 'Elegant Short Kurti Embroidered Indian Ethnic Top Large' belonging to 'Apparels' category and we assume the location of the shop[1] to be in the city

[1] Shop is a generalized term for malls, restaurant, theaters etc.

of Mumbai having Pin code 400093. All the customers are from different parts of Mumbai. As the campaign is location sensitive we apply appropriate distance-feature selection for SVM model using Yahoo Geocoding Web Services API. Website data collected in format for preprocessing.

Apparels	Apparels:Books & Magazines:Mobile Phones
	Books & Magazines:Mobile Phones:Video & Computer Games
Consumer Electronics	Consumer Electronics
	Apparels:Books & Magazines:Movies & Music
	Apparels:Books & Magazines:Mobile Phones
Books & Magazines:Consumer Electronics	Mobile Phones
Mobile Phones	Video & Computer Games
	Consumer Electronics:Apparels:Books & Magazines
	Mobile Phones
Books & Magazines:Consumer Electronics	Consumer Electronics:Books & Magazines:Movies & Music
	Mobile Phones
Toys	Consumer Electronics:Books & Magazines:Movies & Music:Toys
	Apparels:Books & Magazines:Movies & Music
Mobile Phones:Memory Cards Pen Drives & HDD	Books & Magazines:Memory Cards Pen Drives & HDD
Books & Magazines	Consumer Electronics:Books & Magazines
Movies & Music	Apparels:Books & Magazines:Movies & Music
	Consumer Electronics:Books & Magazines
Consumer Electronics	Mobile Phones

Figure 2.1: Sample training data part i) (Customer Related)

The Da Vinci Code by Dan Brown:Apple Ip	F	400020	5
Samsung Galaxy S3:Apple Iphone 3GS:Fif	F	400063	3
Apple iPod	F	400006	
Apple Iphone 3GS:Aliens:Backstreet Boys	F	400063	
Apple Iphone 3GS:LG Optmus-1:Harry Pot	M	400006	5
Samsung Galaxy S3	M	400020	
Fifa 2012:Deus Ex Human Revolution	F	400092	5
Apple Iphone 3GS:Harry Potter SET OF 5	M	400006	
Apple Iphone 3GS:Sony Erricson X8	M	400005	5
Titanic:Speed 2:Backstreet Boys:Apple iP	F	400006	5
Samsung Galaxy S3:Backstreet Boys	M	400063	2
Linkin Park:Apple iPod:LG Plasma	M	400020	5
Speed 2:Linkin Park:Apple iPod	F	400020	5
Sandisk Cruzer Blade 8 GB:The Rediscov	M	400017	
The Rediscovery of India By Meghnad De	M	400063	5
Western Digital 320 gb:Speed 2:Aliens:El	F	400006	5
Backstreet Boys	M	400064	2
Apple Iphone 3GS:LG Plasma	M	400065	

Figure 2.2: Sample training data ii) (Customer Related)

Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels
Elegant Short Kurti Embroided Indian Ethnic Top Large	Apparels

Figure 2.3: Sample training data part iii) (Product Related)

F	0	3	0
F	0	3	0
F	0	3	0
F	0	3	1
F	0	3	0
F	0	3	0
F	0	3	0
F	0	3	0
F	0	3	0

Figure 2.4: Sample training data iv) (Product Related)

The data is then preprocessed to be in format suitable for applying SVM algorithm. For applying SVM algorithm we use library function provided by [12] referring [13]. Radial Basis Function kernel is a reasonable choice for small datasets. It non-linearly maps samples into higher dimensional space. A detailed explanation of kernel methods is available in [14] and [15].

We perform support vector classification and the results are as follows:
Kernel type: Radial basis function
Gamma: 0.2
Total Support Vectors: 20
Rho: -0.176018
Classification Accuracy: 95.9732%

The admin of the website can then launch a campaign if he feels that desired numbers of users are available. The 'deal' shall be available for all the users of the website. However, the potential customers predicted by SVM algorithm can be notified of the same by email and Short Messaging Service (SMS) and have the deal posted on their homepage.

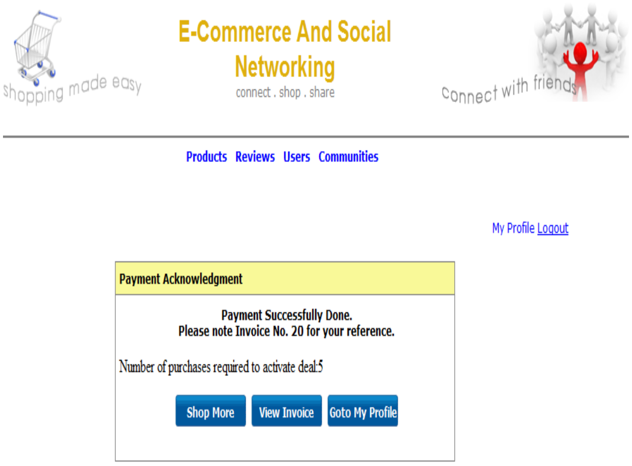

Figure 3 Group Shopping Payment Confirmation stating the number of users required for deal to become active.

4. FUTURE SCOPE:

The project captures certain parameters which act as features in the SVM algorithm. However, many more features can be captured that can lead to better analysis. Other features such as ratings, tags, reviews read, blogs, comments on reviews and community discussions etc. will capture information that is not directly available. Technologies in the field of Natural Language Processing and Semantic Web will have to be employed for the same.

At the application level, there are plans for taking the architecture to Model-View-Controller. Making the user interface interactive using AJAX and other RIA technologies is also another area of future effort.

5. REFERENCES

[1] Yangfeng Qian and Chunhua Ju, 2009, Study on Retail Customer Classification Based on Support Vector Machine, IEEE, 978-1-4244-3894-5

[2] Liu Yue, Liao Zhenjiang, Yin Yafeng, Teng Zaixia, Gao Junjun, Zhang Bofeng, 2010, Selective and Heterogeneous SVM ensemble for Demand forecasting,10th IEEE International Conference on Computer and Information Technology

[3] Toby Segaran, Programming Collective Intelligence, 2007, O'Reilly Media, Inc.

[4] Haralamos Marmanis and Dmitry Babenko, 2009, Algorithms of the Intelligent Web, Manning Publications co.

[5] Satnam Alag, 2009, Collective Intelligence in Action, Manning Publications co.

[6] C. Cortes, V. Vapnik, 1995, Support Vector Networks in Machine Learning, pp. 273-297

[7] Christopher J.C. Burges, 1998, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, 2, 121-167, Kluwer Academic Publishers

[8] Tristan Fletcher, 2009, Support Vector Machines Explained

[9] Dustin Boswell, 2002, Introduction to Support Vector Machines,

[10] J.P. Lewis, 2004, A Short Support Vector Machine Tutorial

[11] John C. Platt, Fast Training of Support Vector Machines using Sequential Minimal Optimization

[12] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines, 2001.Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[13] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, 2010, A Practical Guide to Support Vector Classification

[14] Thomas Hofmann, Bernhard Sch¨olkopf and Alexander J. Smola, 2008, Kernel Methods in Machine Learning, The Annals of Statistics, Vol. 36, No. 3, 1171–1220

[15] Frank J¨akel, Bernhard Sch¨olkopf, Felix A. Wichmann, 2007, A tutorial on Kernel Methods for Categorization