Research Article

# Bayesian modelling average on mixed attributed data

S Prabakaran*, M R Srinivasan**

*Research Scholar, **Professor, Department of Statistics, University of Madras, Chennai, Tamil Nadu, INDIA.
**Email:** prabakarans745@hotmail.com, mrsvasan8@hotmail.com

## Abstract

Ranking the mixed information as categorical, nominal and score card values to stream and identify average performance of the product under different values which help to identify the top priority from customer performance. This paper widely concentrate on averaging the score which obtain from textual information, dichotomous and continuous which leads us to mixed attribute data values from feature of the product, rating and opinion or sentimental score. Organization will get benefited on understand about their service, product requirement expectations from customer prudence and opinion. Thus, ranking or scoring is required to alter such service in online recommender system

**Keywords:** Average Score card, Bayesian, Text Mining and Mixed attributed data

***Address for Correspondence:**
Dr. S Prabakaran, Research Scholar, Department of Statistics, University of Madras, Chennai, Tamil Nadu, INDIA.
**Email:** prabakarans745@hotmail.com

Received Date: 12/10/2014 Accepted Date: 24/10/2014

| Access this article online | |
|---|---|
| Quick Response Code: | Website: www.statperson.com |
| | DOI: 05 November 2014 |

## INTRODUCTION

The rate of internet growth has resulted an enormous data about customer requirements and user utilization on different levels such information which captured as data in multiple platforms and its vast array of choice for consumers. Hence recommendation to consumer will be helpful to make them to purchase or utilize our services based on their preferences. Identifying their preference and requirements will be a key.

**Example:** Few consumers will be gained based the previous rating provided by existing customer and few might dig in deep to the opinion and feedback provided by the existing customers. Thus rating system playing important role on the product recommendation. Existing rating method tends either on the product or just an open opinion feedback by customer. When we considering both in analyzing to make an weighted ranking system, we need an common averaging method, here we are going to utilize Bayesian model to average the product ratings, sentiment or opinion scores by customers.

How to site this article: S Prabakaran, M R Srinivasan. Bayesian modelling average on mixed attributed data. *International Journal of Statistika and Mathemtika* Nov 2014 to Jan 2015; 12(1): 65-68. http://www.statperson.com (accessed 06 November 2014).

## REVIEW OF LITERATURE

### Why Bayesian on Recommendation System?

On marketing product through online to increase sales and leveraging customers prior to rating of product generating subsequent recommendation and suggestions is key feature in online market. Extant models takes customer opinion and rating are non ignorable on recommendations. Recommendation quality will improve substantially by join modeling "Selection" ,"ratings" and "Opinion Scoring" whether and how an item is rated among customers. Recommendations systems became popular study in data mining under machine learning techniques. Recently online marketing activities much utilizing machine learning methods as collaborative filtering methods (Mild, 2003), Collaborative Systems aggregate ratings or recommendations of objects, recognize commonalities between users on the basis of their ratings, generate new recommendations based on inter-user comparisons and possibly, use time-based discounting of ratings. Demographic method categorizes users based on personal attributes and make recommendations based on demographic classes. Content-based method objects defined by their associated features, learn profile of the user's interests based on the features present in objects the user has rated and long-term models, updated as more evidence about user preferences is observed. Utility-based make suggestions based on a computation of the utility of each object for the user, employ constraint satisfaction techniques to locate the best match and no long-term generalizations about users. Knowledge-based as functional knowledge: how a particular item meets a particular need ,can reason about the relationship between a need and a possible recommendation and no long-term models these type of methods widely derived on Clustering approach k-nn methods. Set recommendation (Linden, 2003) and match making to predicting the user preference. For example, Netflix data, and Movie review Imdb and Amazon recommendation request customers to rate a product or review as a scaling system say 1 to 5 as in "stars" and the Movie data also collective reviews similarly. Thus these methods can be categorized into two different classes: heuristic methods and model-based methods. Heuritics Methods often utilize the clustering type algorithms as nearest neighbor methods or k-nn methods. The benefits on these methods easy on implementation and helps to achieve results easily. But these are often adhoc and have been shown to be broadly inferior to model-based methods. (Breese, 1998) Model based methods invoke a probability distribution for customer responses and therefore explicitly hypothesize a data generation process. Model based methods that have been used to generate product recommendations include the mixture model (chien,1999) the hierarchical bayes model (Ansari, 2000), factor analysis (canny,2002) and Bayesian network model (Breese, 1998). An powerful recommendation system based on Bayesian mixture model (Chein, 1999) adapted and implemented on EachMovie data, which outperform the nearest neighbour methods (sarwar et al. 2000). In Bayesian model averaging weighted average methodology on recommender systems, help to wrap the space for lack of data availability on user preferences which are uncertain to identify through normal approaches. Under strict uncertainty (Braziunas, 2006), recommender systems incorporate knowledge from different sources by assuming a set of hypotheses, with no belief on the strength of these hypotheses

## COMBINATION OF METHODS

The first method on combining model is presented through airline passenger data (Barnard, 1963),Stimulated a flurry of article in economic literature (Clemen,1989) this type of works not availed in statistical journals. Combining two different opinions or models implemented (Robert, 1965).Thus distribution, essentially a weighted averaged of posterior distribution models, is similar to Bayesian averaging model. The fundamental idea of BMA carted in uncertainty (Leamer, 1978). The ignorance on model uncertainty has recognized by many authors over the decade (Dijkstra, 1988). However implementing BMA through computational machine learning power help to overcome such issue. In Bayesian setting unknown variables are described using probability distributions and observing data allow knowing about their distribution to get updated on bayes theorem. Bayesian view point uncertainty using probabilities based on each new observation collected which set as prior distribution; from Bayesian theorem allow developing posterior distribution to evaluate the effect of new observation. Such methods always help to build a model specifically in online system. Where new data always in flow from various user on same time. The powerful feature of Bayesian framework is the ease to build hierarchical models. It also significantly predicts perfectly to overcome from the over-fitting where parameters are turned into noise data and missing values.

## BAYESIAN METHOD ON AVERAGE SCORE ON DIFFERENT FEATURES OF DATA

Consider the individual customer had rated the product and expressed about the opinion, Conditional rating contributed (Ying, 2006) using ordered probit model. We assuming the rating as i driven by evaluated as $\alpha_{ij}$

$$\alpha_{ij} = \mu_{i0} + \delta_j$$

Where $\mu_{i0}$ for multiple level of baseline ratings like positive or negative for different customer and $\delta_j$ for variation of products features and quality such that $\delta_j \sim N(0, \sigma_\delta^2)$. The adjustment effects for various rating and point scaling environment,
we have including adjust effect

$$\alpha_{ijk}^{*} = \alpha_{ij} + \delta_{i,1:N} X_{j,k(i)}$$

where $X_j$ is $N \times 1$ vector of covariates, describes the rating for product $j$time of rating and session $k$ by individual $i$. The $1 \times N$ vector $\delta_{i,1:N}$ helps to capture the impact on rating.

For Rating a product on scaling method, we considering as follows

$$P(y_{ijk} = r | Z_{ijk} = 1) = \Pr(\mu_{i,r-1} < \alpha_{ijk}^{*} + \varepsilon_{ijk} < \mu_{i,r})\ ;\ r = 1, \ldots, 5$$

Where $y_{ijk}$is the rating contribution by $i$ product, $j$ rating and $k$ session. $Z_{ijk}$ indicates a rating contribution and $\varepsilon_{ijk}$ is the idiosyncratic error with mean zero.
Under the assumption that $\varepsilon_{ijk}$ follows a normal distribution, the probability with which an $r$-star rating is contributed is represented by the following ordered probit specification:

$$P(y_{ijk} = r | Z_{ijk} = 1) = \begin{cases} \varphi(-\alpha_{ijk}^{*}), r = 1 \\ \varphi(\mu_{i1} - \alpha_{ijk}^{*}) - \varphi(-\alpha_{ijk}^{*}), r = 2 \\ \varphi(\mu_{i2} - \alpha_{ijk}^{*}) - \varphi(\mu_{i1} - \alpha_{ijk}^{*}), r = 3 \\ \varphi(\mu_{i3} - \alpha_{ijk}^{*}) - \varphi(\mu_{i2} - \alpha_{ijk}^{*}), r = 4 \\ 1 - \varphi(\mu_{i3} - \alpha_{ijk}^{*}), r = 5 \end{cases}$$

Where $\mu_{i.}$individual specific cutpoints for order is probit model and $\varphi(.)$ is standard normal cumulative distribution function. This for 5 star rating scale evaluations. For continuous scaling assuemed as follow an normal distribution, in the case reported rating $y_{ijk}$can be model directly using linear model (Ansari, 2000) with mean $\alpha_{ijk}^{*}$.
The full model of BMA is represents as $Y = X\beta + \in$ where $X$ is $n \times p$ and $\epsilon \sim N(0, \sigma^2 \boldsymbol{I})$. The space of all possible model configurations is $M = [M_1, M_2 \ldots, M_q]$ where$q = 2^p$.

We setup a hierarchical mixture model:

$$M_k \sim \pi(M_k)$$
$$\sigma | M_k \sim \pi(\sigma | M_k)$$
$$\beta_\omega | M_k \sim \pi(\beta_\omega | M_k, \sigma^2)$$

Where $\Omega = \omega_1, \omega_2, \ldots, \omega_p$ a vector of 0's and 1 is's representing the inclusion of variable in $M_k$
We can analyze the conditional model:

$$Y | \beta, \sigma, M_k \sim N(X_\omega \beta_\omega, \sigma^2 I)$$

The posterior distribution model $M_k$is $p(M_k|Y) = \frac{p(Y|M_k)\pi(M_k)}{\sum_{k=0}^{q} p(Y|M_k)\pi(M_k)}$
The expected value of $\beta_k$ is $E(\beta_k|Y) = \sum_{k=0}^{q} p(M_k|Y) E(\beta_k|M_k, Y)$
And the posterior probability distribution of $\beta_k$ is $p(\beta_k|Y) = \sum_{k=0}^{q} p(M_k|Y) p(\beta_k|M_k, Y)$

## CONCLUSION

The above method help to combining both rating methods and to develop scoring and providing benchmark work for creating rating. In future work we are going to expand the method by considering the experience on customers and evaluate the competence of averaging.

## REFERENCES


1. Amabile, Teresa M. (1983), "Brilliant but Cruel: Perceptions of Negative Evaluators," *Journal of Experimental Social Psychology*, 19 (March), 146–56.
2. Anderson, Eugene W., (1998), "Customer Satisfaction and Word-of-Mouth," *Journal of Service Research*, 1(1), 1-14.
3. Anderson, Eugene W. and Mary W. Sullivan (1993), "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science*, 12 (2), 125-143.
4. Ansari, Asim, Skander Essegaier, and Rajeev Kohli (2000), "Internet Recommendation Systems," *Journal of Marketing Research*, 37 (August), 363–75.

5. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (B. Petrox and F. Caski, eds.) 267.
6. Barnard, G. A. (1963). New methods of quality control. *J. Roy. Statist. Soc. Ser. A* 126 255.
7. Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statist. Sci.* 2 317–352.
8. Berger, J. O. and Sellke, T. (1987). Testing a point null hypothesis (withdiscussion). *J. Amer. Statist. Assoc.* 82 112–122.
9. Bernardo, J. and Smith, A. (1994). *Bayesian Theory*. Wiley, Chichester.
10. Besag, J. E., Green, P., Higdon, D. and Mengerson, K. (1995). Bayesian computation and stochastic systems. *Statist. Sci.* 10 3–66.
11. Berinsky, Adam J. (2004). *Silent Voices.* New Jersey: Princeton University Press.
12. Breiman, L. (1996). Bagging predictors. *Machine Learning* 26 123–140.
13. Buckland, S. T., Burnham, K. P. and Augustin, N. H. (1997). Model selection: an integral part of inference. *Biometrics* 53 275–290.
14. Carlin, B. P. and Chib, S. (1993). Bayesian model choice via Markov chain Monte Carlo. *J. Roy. Statist. Soc. Ser. B* 55 473–484.
15. Carneiro, Herman Anthony and Eleftherios Mylonakis (2009), "Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks," *Clinical Infectious Diseases*, 49 (10), 1557-1564.
16. Chevalier, Judith and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 48 (August), 345-354.
17. Dellarocas, Chrysanthos and Ritu Narayan (2006), "A Statistical Measure of a Population's Propensity to Engage in Post-experience Online Word-of-Mouth," *Statistical Science*, 21 (2), 277-285.
18. Eskin, Gerald J. (1973), "Dynamic Forecasts of New Product Demand Using a Depth of Repeat Model," *Journal of Marketing Research*, 10 (2), 115-129.
19. Fleming, John H., John M. Darley, James L. Hilton, and Brian A. Kojetin (1990), "Multiple Audience Problem: A Strategic Communication Perspective on Social Perception," *Journal of Personality and Social Psychology*, 58 (April), 593–609.
20. Fourt, Louis A. and Joseph W. Woodlock (1960), "Early Prediction of Market Success for New Grocery Products," *Journal of Marketing*, 25 (2), 31-38.
21. Gartner, M. (1976), "Endogenous Bandwagon and Underdog Effects," *Public Choice*, 25, 83-139.
22. Hardie, Bruce G.S., Peter S. Fader and Michael Wisniewski (1998), "An Empirical Comparison of New Product Trial Forecasting Models," *Journal of Forecasting*, 17 (June/July), 209- 229.
23. Heckman, James J. (1979), "Sample Selection Bias and Specification Error," *Econometrica*, 47 (1), 153-161.
24. Jackson, John E. (1983), "Election Night Reporting and Voter Turnout," *American Journal of Political Science*, 27 (4), 615-635.
25. Kamakura, Wagner A., Michel Wedel, Fernando de Rosa and Jose Afonso Mazzon (2003), "Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction," *International Journal of Research in Marketing*, 20 (1), 45-65.
26. Kuksov, Dmitri and Ying Xie (2010), "Pricing, Frills, and Customer Ratings," *Marketing Science*, 29 (5), 925-943.
27. Ying, Yuanping, Fred Feinberg and Michel Wedel (2006), "Leveraging Missing Ratings to Improve Online Recommendation Systems," Journal of Marketing Research, 43 (3), 355- 365.

Source of Support: None Declared
Conflict of Interest: None Declared