International Journal on Applications in Engineering and Technology
Volume 1: Issue 6: June 2015, pp 1-.3 www.aetsjournal.com *ISSN (Online) : 2455-0523*

Author Spamicity Model For Opinion Spammers

Karpagam R, Dr P.Ezhilarasu,Manoj Prabhakar J

Abstract— In e-commerce application reviews play a very important role. Opinionated social media such as product reviews are now widely used by individuals and organizations for their decision making. Since these reviews contribute for success or failure in sales of a product, reviews are being manipulated for positive or negative opinions (e.g., writing fake reviews). In recent years, fake review detection has attracted significant attention from both the business and research communities. This work proposes a novel angle to the problem by modeling spamicity as latent. An unsupervised model, called Author Spamicity Model (ASM) is proposed. It works to exploit various observed behavioral footprints of reviewers. . Several extensions of ASM are also considered leveraging from different priors.

Keywords— e-commerce apps; reviews; Author Spamicity Model.

I. INTRODUCTION

Online reviews of products and services are used extensively by consumers and businesses to make critical purchase, product design, and customer service decisions. However, due to the financial incentives associated with positive reviews, imposters try to game the system by posting fake reviews and giving unfair ratings to promote or demote target products and services. Such individuals are called opinion spammers and their activities are called opinion spamming.

The growth of e-commerce is rapid in the whole world. In India e-commerce market was worth about $3.8 billion in 2009, in 2013 it went up to $12.6 billion. According to Google India there were 35 million online shoppers in India in 2014 and it is expected to cross 100 million by the end of the year 2016. A research says that electronics and apparels are the biggest categories in terms of sales.

Reviews and ratings play a major role in the products and services that are available on the e-commerce sites. Product reviews are one of the resources shoppers trust the most when they are researching on new products and services. Many customers read reviews of products or stores before making the decision of what or from where to buy and whether to buy or not. In order to attract more number of customers and to increase sales in the market, these reviews are manipulated. Manipulated reviews are fake reviews or untruthful reviews. As writing fake reviews comes with monetary gain, there has been a huge increase in deceptive opinion spam on online review websites. Basically fake review or fraudulent review or opinion spam is an untruthful review. Positive reviews of a target object may attract more customers and increase sales as well as negative review of a target object may lead to lesser demand and decrease in sales. Both are comes under fake review category only. As a result it is a very difficult task for an ordinary customer to differentiate between fraudulent reviews from genuine ones, by just looking at each review.

Role of online reviews in current scenario:

- 70% of customers consult reviews or ratings before making a final purchase.
- 63% of consumers are more likely to purchase from a site only if it has product ratings and reviews.
- 67% of consumers read six reviews or less before they feel they can trust a business enough to make a purchase.
- As many as 79% of consumers trust product reviews as much as a personal recommendation.
- 80% of consumers have changed their mind about purchases based on negative information they have found online.
- 71 % of consumers agreed that reviews make them more comfortable that they are purchasing the right product or service and also from the right place.

II. MOTTO

It proposes a novel and principled method to exploit observed behavioral footprints to detect spammers (fake reviewers) in an unsupervised Bayesian framework precluding the need of any manual labels for learning which is both hard and noisy. A key advantage of employing Bayesian inference is that the model facilities characterization of various spamming activities using estimated latent variables and the posterior. It facilitates both detection and analysis in a single framework rendering a deep insight into the opinion spam problem. This cannot be done using existing methods. To our knowledge, this is the first principled model for solving this problem.

It proposes a novel technique to evaluate the results without using any labeled data. This method uses reviews of the top ranked and bottom ranked authors produced by the model as two classes of data to build a supervised classifier. The key idea is that the classification uses a complete different set of features than those used in modeling. Thus, if this classifier can classify accurately, it gives a good confidence that the unsupervised spamicity model is effective.

Karpagam R, PG Scholar, Department Of Computer Science And Engineering, Hindusthan College Of Engineering And Technology, Coimbatore, Tamilnadu, India (Email; rkarpagam23@gmail.com)

Dr.P Ezhilarasu ,Professor, Department Of Computer Science And Engineering, Hindusthan College Of Engineering And Technology, Coimbatore, Tamilnadu, India. (Email: prof.p.ezhilarasu@gmail.com)

Manoj Prabhakar J, PG Scholar, Department Of Computer Science And Engineering, Hindusthan College Of Engineering And Technology, Coimbatore, Tamilnadu, India (Email; manojprabhakarj92@gmail.com)

III. MODEL OVERVIEW

The existing system only deals with prediction and assumption charts, here the charts will be in the normal format to understand the data.

In classification, one is concerned with assigning objects to classes on the basis of measurements made on these objects. There are two main aspects to classification: discrimination and clustering, or supervised and unsupervised learning. In unsupervised learning (also known as cluster analysis, class discovery and unsupervised pattern recognition), the classes are unknown a prior and need to be discovered from the data. In contrast, in supervised learning (also known as discriminate analysis, class prediction, and supervised pattern recognition), the classes are predefined and the task is to understand the basis for the classification from a set of labelled objects (training or learning set). This information is then used to classify future observations. The present article focuses on the unsupervised problem, that is, on cluster analysis, but draws on notions from supervised learning to address the problem.

In cluster analysis, the data are assumed to be sampled from a mixture distribution with K components corresponding to the K clusters to be recovered. Let (X1, ..., Xp) denote a random 1 ×p vector of explanatory variables or features, and let Y∈ {1, ..., K} denote the unknown component or cluster label. Given a sample of X values, the goal is to estimate the number of clusters K and to estimate, for each observation, its cluster label Y. Suppose we have data X = (xij) on p explanatory variables (for example, genes) for nobservations (for example, tumor mRNA samples), where xij denotes the realization of variable Xjfor observation i and xi = (xi1,...,.xip) denotes the data vector for observation i, i = 1,...,n, j = 1,...,p. We consider clustering procedures that partition the learning set $\mathcal{L}$ = {x1,...,xn} into Kclusters of observations that are 'similar' to each other, where K is a user-prespecified integer. More specifically, the clustering $\mathcal{P}(\cdot;\mathcal{L})$ assigns class labels $\mathcal{P}$ (Xi;$\mathcal{L}$) = $\hat{y}_i$ to each observation, where $\hat{y}_i \in \{1,\ldots,K\}$. Clustering procedures generally operate on a matrix of pair wise dissimilarities (or similarities) between the observations to be clustered, such as the Euclidean or Manhattan distance matrices. A partitioning of the learning set can be produced directly by partitioning clustering methods (for example, k-means, partitioning around medoid (PAM), self-organizing maps (SOM)) or by hierarchical clustering methods, by 'cutting' the dendrogram to obtain K 'branches' or clusters. Important issues, which will only be addressed briefly in this article, include: the selection of observational units, the selection of variables for defining the groupings, the transformation and standardization of variables, the choice of a similarity or dissimilarity measure, and the choice of a clustering method. Our main concern here is to estimate the number of clusters K.

When a clustering algorithm is applied to a set of observations, a partition of the data is returned whether or not the data show a true clustering structure, that is, whether or not K = This fact causes no problems if clustering is done to obtain a practical grouping of the given set of objects, as for organizational or visualization purposes (for example, hierarchical clustering for displaying large gene-expression data matrices. However, if interest lies primarily in the recognition of an unknown classification of the data, an artificial clustering is not satisfactory, and clusters resulting from the algorithm must be investigated for their relevance and reproducibility. This task can be carried out by descriptive and graphical exploratory methods, or by relying on probabilistic models and suitable statistical significance tests.

We argue here that validating the results of a clustering procedure can be done effectively by focusing on prediction accuracy. Once new classes are identified and class labels are assigned to the observations, the next step is often to build a classifier for predicting the class of future observations. The reproducibility or predictability of cluster assignments becomes very important in this context, and therefore provides a motivation for using ideas from supervised learning in an unsupervised setting. Res ampling methods such as bagging and boosting have been applied successfully in the field of supervised learning to improve prediction accuracy.

We propose here a novel re sampling method, which combines ideas from discriminate and cluster analysis for estimating the number of clusters in a dataset.

Although the proposed re sampling methods are applicable to general clustering problems and procedures, particular attention is given to the clustering basis of gene expression data using the partitioning around methods are leads to failure.

IV. METHODOLOGY

In ASM, spam detection is influenced by review and author features. Normalized continuous author features in [0, 1] are modeled as following a Beta distribution ($ya,rf\sim \psi k\in\{\hat{s},n\ \}f$) (Table 1). This enables ASM to capture more fine grained dependencies of author's behaviors with spamming. However, review features being more objective, we found that they are better captured when modeled as binary variables being emitted from a Bernoulli distribution ($xa,rf\sim Bern\ \ \theta k\in\{\hat{s},n\ \}f$ (Table 1). $\theta k\in\{\hat{s},n\ \}f$ for each review feature $f\in\{DUP,EXT,DEV,ETF,RA\}$ and $\psi k\in\{\hat{s},n\ \}f$ for each author feature $f\in\{CS,MNR,BST,RFR\}$ denote the per class/cluster (spam vs. non-spam) probability of emitting feature f.

Latent variables sa and πr denote the spamicity of an author, a and the (spam/non-spam) class of each review, r. The objective of ASM is to learn the latent behavior distributions for spam and non-spam clusters (K = 2) along with spamicites of authors from the observed features. We now detail its generative process.

1. For each class/cluster, $k\in\{\hat{s},n\ \}$:

Draw $\theta kf\in\{DDD,\ldots,RR\}\sim Beta(\ \gamma f)$

2. For each (author), $a\in\{1\ldots A\}$:

i. Draw spamicity, $sa\sim Beta(\alpha a)$;

ii. For each review, $ra\in\{1\ldots Ra\}$:

a. Draw its class, $\pi ra\sim Bern(sa)$

b. Emit review features $f\in\{DUP,\ldots,RA\}$:

$xraf \sim Bern\ \ \theta\pi raf\ \ ;$

c. Emit author features $f\in\{CS,..,RFR\}$:

$yraf \sim \psi\pi raf$;

We note that the observed author features are placed in the review plate (Figure 1). This is because each author behavior can be thought of as percolating through reviews of that author and emitted across each review to some extent. Doing this renders two key advantages: i) It permits us to exploit a larger co-occurrence domain. ii) It paves the way for a simpler sampling distribution providing for faster inference.

V. Observation

1. As *k* increases, we find a monotonic degradation in review classification performance (except HS) which is expected as the spamicities of top and bottom authors get closer which makes the corresponding review classification harder.
2. For *k* = 5%, ASM models performs best on F1 and Accuracy metrics. Next in order are FSum, RankBoost, and SVMRank. It is interesting that the simple baseline FSum performs quite well for *k* = 5%. The reason is attributed to the fact that the top positions are mostly populated by heavy spammers while the bottom positions are populated by genuine reviewers and hence a naïve un-weighted FSum could capture this phenomenon.
3. For *k* = 10, 15%, FSum does not perform so well (SVMRank and RankBoost outperform Fsum). This is because for *k* = 10, 15%, the ranked positions involve more difficult cases of authors/reviewers and a mere sum is not able to balance the feature weights as not all features are equally discriminating.
4. For *k* = 10%, SVMRank and RankBoost outperform ASM-UP and perform close to ASM-IP. ASM-HE still outperforms SVMRank and RankBoost by 4% in F1 and 2-3 % in accuracy.
5. For *k* = 15%, ASM variants outperform other methods and increase F1 by a margin of 2-10% and accuracy by 3-7%.
6. Performance of HS remains much poorer and similar for each *k* showing that it is not able to rank spammers well, indicating that helpfulness is not a good metric for spam detection. In fact, helpfulness votes are subject to abuse.

VI. Conclusion

This paper proposed a novel and principled method to exploit observed reviewing behaviors to detect opinion spammers (fake reviewers) in an unsupervised Bayesian inference framework. To our knowledge, this is the first such attempt. Existing methods are mostly based on heuristics and/or ad-hoc labels for opinion spam detection. The proposed model has its basis in the theoretic foundation of probabilistic model based clustering. The Bayesian framework facilitates characterization of many behavioral phenomena of opinion spammers using the estimated latent population distributions. It also enables detection and posterior density analysis in a single framework. This cannot be done by any of the existing methods. The paper also proposed a novel way to evaluate the results of unsupervised opinion spam models using supervised classification without the need of any manually labeled data.

References

[1] L. Azzopardi, M. Girolami, and K. V. Risjbergen, "Investigating the relationship between language model perplexity and ir precision-recall measures," in Proc. 26th Int. Conf. Res. Develop. Inform. Retrieval, 2003, pp. 369–370.

[2] G. Heinrich, Parameter estimation for text analysis, "Univ. Leipzig, Leipzig, Germany, Tech. Rep., http://faculty.cs.byu.edu/~ringger CS601R/papers/HeinrichGibbsLDA.pdf, 2008.

[3] A. Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in Proc. 18th Eur. Conf. Mach. Learn., 2007, pp. 616–623.

[4] A. Klementiev, D. Roth, and K. Small, "Unsupervised rank aggregation with distance-based models," in Proc. 25th Int. Conf. Mach. Learn., 2008, pp. 472–479.

[5] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proc. 19thACMInt. Conf. Inform. Knowl. Manage., 2010, pp. 939–948.

[6] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 632–640.

[7] McAuliffe, D.B. and J. 2007. Supervised Topic Models. NIPS (2007).

[8] Mukherjee, A., Liu, B. and Glance, N. 2012. Spotting Fake Reviewer Groups in Consumer Reviews. WWW (2012).

[9] Mukherjee, A., Liu, B., Wang, J., Glance, N. and Jindal, N. 2011. Detecting Group Review Spam. WWW (2011).

[10] Feng, S., Banerjee R., Choi, Y. 2011. Syntactic Stylometry for Deception Detection. ACL (2011).