ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 1 , January 2020

Predictions Analysis of Netspam Reviews in Online Social Media Based on Loopy Belief Propagation.

K. VishnuPriya

Assistant Prof. East Point College of Engineering & Technology,Bangalore.

ABSTRACT: The NetSpam, which utilizes spam features for modeling review datasets as heterogeneous information networks to map spam detection procedure into a classification problem in such networks. Using the importance of spam features help us to obtain better results in terms of different metrics experimented on real-world review datasets from Yelp and Amazon websites. Netspam detects the spam and helps us to find the whether the reviews are real/ fake. In this work, which exploits the burstiness nature of reviews to identify review spammers? Bursts of reviews can be either due to sudden popularity of products or spam attacks. Reviewers and reviews appearing in a burst are often related in that sense that spammer tends to work without other spammers and genuine reviewers tent to appear together with other genuine reviewers. This paves the way for us to build a network of reviewers appearing in different bursts. We then model reviewers and their co-occurrence in bursts as a Markov Random Field (MRF), and employ the Loopy Belief Propagation (LBP) method to infer whether a reviewer is a spammer or not in the graph. We also propose several features and employ feature induced message passing in the LBP framework for network inference. We further propose a novel evaluation method to evaluate the spammer's automatically using supervised classification of their reviews. Additionally, we employ domain experts to perform a human evaluation of the identified spammers and non-spammers. Both the classification result and human evaluation result show that the proposed method outperforms strong baselines, which demonstrate the effectiveness of the method.

KEYWORDS: Markov Random Field, employ the Loopy Belief Propagation, Spam Finder and feature selection.

1. INTRODUCTION

Online social media portal plays an influential role information propagation which considered as an important source for producers in their advertising campaigns as well as for customers in selecting products and services. People rely a lot on the written reviews in their decision-making processes, and positive/negative reviews encouraging/discouraging them in their selection of products and services. These reviews thus have become an important factor in success of a business while positive reviews can bring benefits for a company, negative reviews can potentially impact credibility and cause economic losses. In the past years, people rely a lot on the written reviews in their decision-making processes, and positive/negative reviews encouraging/discouraging them in their selection of products and services. In addition, written reviews also help service providers to enhance the quality of their products and services. These reviews thus have become an important factor in success of a business while positive reviews can bring benefits for a company, negative reviews can potentially impact credibility and cause economic losses. The fact that anyone with any identity can leave comments as review provides a tempting opportunity for spammers to write fake reviews designed to mislead users' opinion.

II. MATERIAL AND METHODS

We propose a novel framework, named NetSpam, which utilizes spam features for modeling review datasets as heterogeneous information networks to map spam detection procedure into a classification problem in such networks. Using the importance of spam features help us to obtain better results in terms of different metrics experimented on real-world review datasets from Yelp and Amazon websites. This generative model of deception which, in conjunction with a deception classifier, we use to explore the prevalence of deception in six popular online review communities: Expedia, Hotels.com, Priceline, Trip Advisor and Yelp.. In this paper, we explore multiple heterogeneous pair wise features in virtue of some collusion signals found in reviewers' rating behaviors and linguistic patterns. In addition, an unsupervised and intuitive colluder detecting

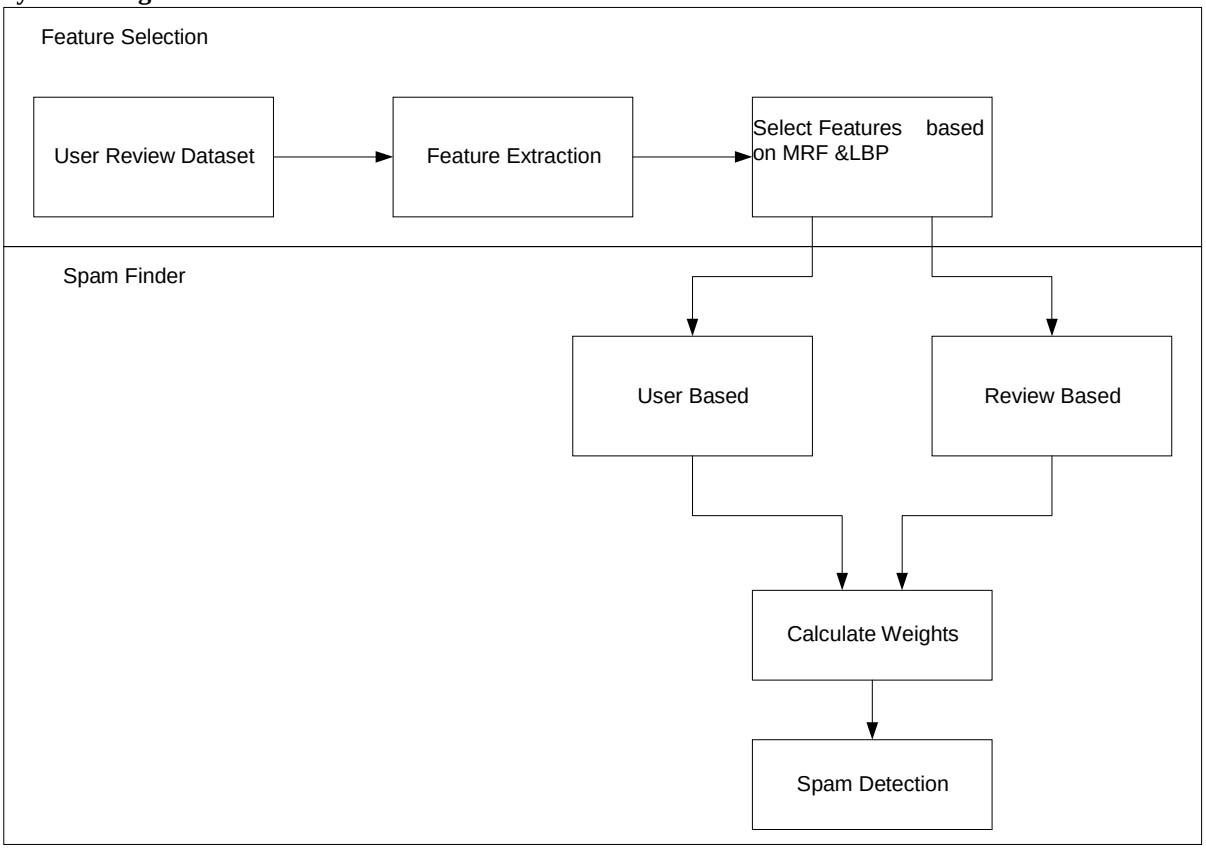
ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 1 , January 2020

framework has been proposed which can benefit from these pair wise features. Extensive experiments on real datasets how the effectiveness of our four method and satisfactory superiority over several competitor. It is important to identify and filter out the review spam. Previous work only focuses on some heuristic rules, such as helpfulness voting, orienting deviation, which limits the performance of this task. In this paper, we exploit machine learning methods to identify review spam. Toward the end, we manually build a spam collection from our crawled reviews. We first analyze the effect of various features in spam identification. We also observe that the review spammer consistently writes spam. This provides us another view to identify review spam: we can identify if the author of the review is spammer. Based on this observation, we provide a two view semi-supervised method, co-training, to exploit the large amount of unlabeled data. The experiment results show that our proposed method is effective..Our designed machine learning methods achieve significant improvements in comparison to the heuristic baselines.

System Design:



System architecture is the conceptual design that defines the structure and behavior of a system. An architecture description is a formal description of a system, organized in a way that supports reasoning about the structural properties of the system. It defines the system components or building blocks and provides a plan from which products can be procured, and systems developed, that will work together to implement the overall system.

The System architecture is shown below.

The system consists of two module components

1. Feature Selection
2. Spam Finder

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 1 , January 2020

Feature Selection: The input reviews are collected from various sources. From among all the attributes collected from user reviews, the most important features on reviews are extracted.

Spam Finder: The selected features are inputted to find the spam's using user based spam and review based spam by calculating weight age on the given features.

NetSpam Algorithm:

1) Algorithm steps for similarity matching based on weightage

wordCount = 0

TotalWeight = 0

While NOT EOF(reviews)

Read A word

wordCount = wordCount + 1

While NOT EOF(frequencies)
Read CurrentWord, CurrentWeight
If word = CurrentWord

TotalWeight = TotalWeight + CurrentWeight
END While

END While

Result = TotalWeight / wordCount
If Result > 1

Print Spam
Else

Print Legitimate
END

Algorithm steps for finding spam- Netspam()

Input : review- dataset, spam- feature- list,
Pre- labeled- reviews

Output : features- importance(W),
Spamicity- probability(Pr)

% *u*,*v*: review, y_u: spamicity probability of review *u* % $f(x_u)$: initial probability of review *u* being spam

% p_l: metapath based on feature *l*, *L*: features number

% *n*: number of reviews connected to a review

% $m^p{}_u{}^l$: the level of spam certainty

% $m^p{}_{u,v}{}^l$: the metapath value %Prior Knowledge if semi-supervised mode

I **if** $u \in pre-labeled-reviews$
$\{y_u = label(u)$

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 1 , January 2020

else

{

yu= 0

else % unsupervised mode

$$\{y_u = \frac{1}{L}\sum_{l=1}^{L} f(x_{lu})$$

%Network Schema Definition *schema* = defining schema based on spam-feature-list

% Metapath Definition and Creation for

$$p_l \in schema$$

$$\begin{cases} \textbf{for } u, v \in review - dataset \\ \quad \begin{cases} m_u^{p_l} = \frac{\lfloor s \times f(x_{lu}) \rfloor}{s} \\ m^p \quad \frac{\lfloor s \times f(x_{lv}) \rfloor}{s} \end{cases} \end{cases}$$

$$\begin{cases} v^l = \\ \textbf{if } m^{p_l} = m^{p_l} \end{cases}$$

do do

$$\begin{cases} \{mp_{u,v}^{p_l} = m_u^{p_l} \\ \quad \textbf{else} \\ \{mp_{u,v}^{p_l} = 0 \end{cases}$$

% Classification - Weight Calculation

for $p_l \in schemes$

do $\{W_{p_l} = \frac{\sum_{r=1}^{n}\sum_{s=1}^{n} mp_{r,s} \times y_r \times y_s}{\sum_{r=1}^{n}\sum_{s=1}^{n} mp_{r,s}^{p_l}}$

% Classification - Labeling for *u,v*∈*review* − *dataset*

Pru,v = 1 − Π*Lpl*=11 − *mppu,vl* × *Wpl*

do

Pru = *avg*(*Pru*,1,*Pru*,2,...,*Pru*,*n*) return (W, Pr)

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 1 , January 2020

Table 1.1: Features for users and reviews in four defined categories

Spam Feature	User-based	Review-based
Behavioral based Features	**Burstiness**: Spammers, usually write their spam reviews in short period of time for two reasons: first, because they want to impact readers and other users, and second because they are temporal users, they have to write as much as reviews they can in short time. $$x_{BST}(i) = \begin{cases} 0 & (L_i - F_i) \notin (0, \tau) \\ 1 - \frac{L_i - F_i}{\tau} & (L_i - F_i) \in (0, \tau) \end{cases} \quad (1)$$ Where $Li - Fi$ describes days between last and first review for τ = 28. Users with calculated value greater than 0.5 take value 1 and others take 0. **Negative Ratio:** Spammers tend to write reviews which defame businesses which are competitor with the ones they have contract with, this can be done with destructive reviews, or with rating those businesses with low scores. Hence, ratio of their scores tends to be low. Users with average rate equal to 2 or 1 take 1 and otherstake 0.	**Early Time Frame**: Spammers try to write their reviews asap, in order to keep their review in the top reviews which other users visit them sooner. $$x_{ETF}(i) = \begin{cases} 0 & (T_i - F_i) \notin (0, \delta) \\ 1 - \frac{T_i - F_i}{\delta} & (T_i - F_i) \in (0, \delta) \end{cases}$$ Where $L_i - F_i$ denotes days specified written review and first written review for a specific business. We have also δ= 7. Users with calculated value greater than 0.5 takes value 1 and others take0. **Rate Deviation using threshold**: Spammers, also tend to promote businesses they have contract with, so they rate these businesses with high scores. In result, there is high diversity in their given scores to different businesses which is the reason they have high variance and deviation. $$x_{DEV}(i) = \begin{cases} 0 & otherwise \\ 1 - \frac{r_{ij} - avg_{e \in E_{*j}} r(e)}{4} & \geq \beta_1 \end{cases}$$ Where $\beta 1$ is some threshold determined by recursive minimal entropy partitioning. Reviews are close to each other based on their calculated value, take same values (in [0, 1)).
Linguistic based Features	**Average Content Similarity, Maximum Content Similarity:** Spammers, often write their reviews with same template and they prefer not to waste their time to write an original review. In result, they have similar reviews. Users have close calculated values take same values (in [0, 1)).	**Number of first Person Pronouns, Ratio of Exclamation Sentences containing '!'** : First, studies show that spammers use second personal pronouns much more than first personal pronouns. In addition, spammers put '!' in their sentences as much as they can to increase impression on users and highlight their reviews among other ones

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 1 , January 2020

III. SIMULATION&RESULTS

Table 1.2: Metapaths used in the NetSpam framework.

Row	Notation	Type	MetaPath	Semantic
1	R-DEV-R	RB	Review-Threshold Rate Deviation-Review	Reviews with same Rate Deviation from average Item rate (based on recursive minimal entropy partitioning)
2	R-U-NR-U-R	UB	Review-User-Negative Ratio-User-Review	Reviews written by different Users with same Negative Ratio
3	R-ETF-R	RB	Review-Early Time Frame-Review	Reviews with same released date related to Item
4	R-U-BST-U-R	UB	Review-User-Burstiness-User-Review	Reviews written by different users in same Burst
5	R-RES-R	RL	Review-Ratio of Exclamation Sentences containing '!'-Review	Reviews with same number of Exclamation Sentences containing '!'
6	R-PP1-R	RL	Review-first Person Pronouns-Review	Reviews with same number of first Person Pronouns
7	R-U- ACS-U-R	UL	Review-User-Average Content Similarity-User-Review	Reviews written by different Users with same Average Content Similarity using cosine similarity score
8	R-U- MCS-U-R	UL	Review-User-Maximum Content Similarity-User-Review	Reviews written by different Users with same Maximum Content Similarity using cosine similarity score

$$W_{p_{ACS}} = \frac{1 \times 1 \times 0.9 + 1 \times 0 \times 0.2}{1 \times 1 + 1 \times 0 \times 1} \simeq 0.82 \qquad \Pr_{1,4} = 1 - \left(1 - mp_{1,4}^{p_{ETF}} \times W_{p_{ETF}}\right) = 1 - (1 - 0.5 \times 0.33) \simeq 0.166$$

$$\Rightarrow \Pr_4 = \operatorname{avg}\left(\Pr_{1,4}, \Pr_{3,4}\right) \simeq 0.165$$

$$W_{p_{ETF}} = \frac{1 \times 1 \times 0.5 + 1 \times 0 \times 0.5 + 1 \times 0 \times 0.5}{0.5 + 0.5 + 0.5} \simeq 0.33 \qquad W_{p_{BST}} = \frac{1 \times 0 \times 0.3}{0.3} = 0$$

$$\Pr_{3,4} = 1 - \prod_{i=1}^{m}\left(1 - mp_{3,4}^{p_i} \times W_{p_i}\right) = 1 - \left(1 - mp_{3,4}^{p_{ACS}} \times W_{p_{ACS}}\right) \times \left(1 - mp_{3,4}^{p_{BST}} \times W_{p_{BST}}\right) = 1 - (1 - 0.2 \times 0.82)(1 - 0.3 \times 0) = 0.164$$

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 1 , January 2020

RESULT

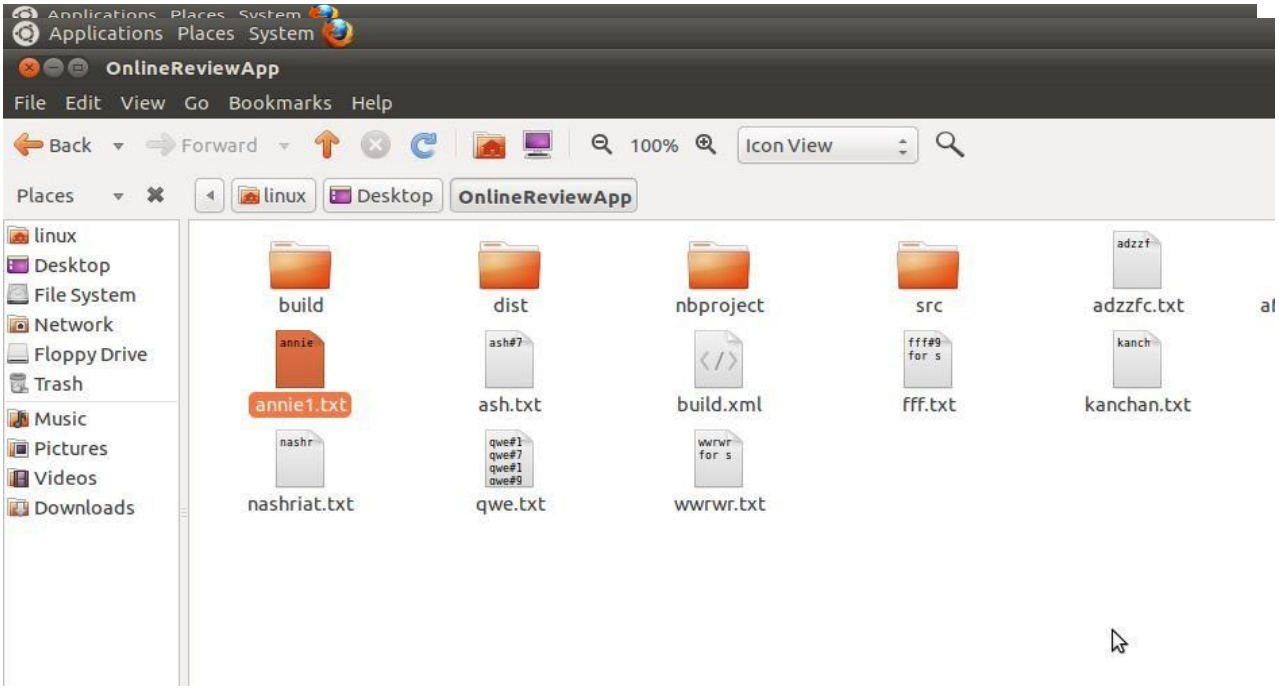


Fig 1.1: User entering a review

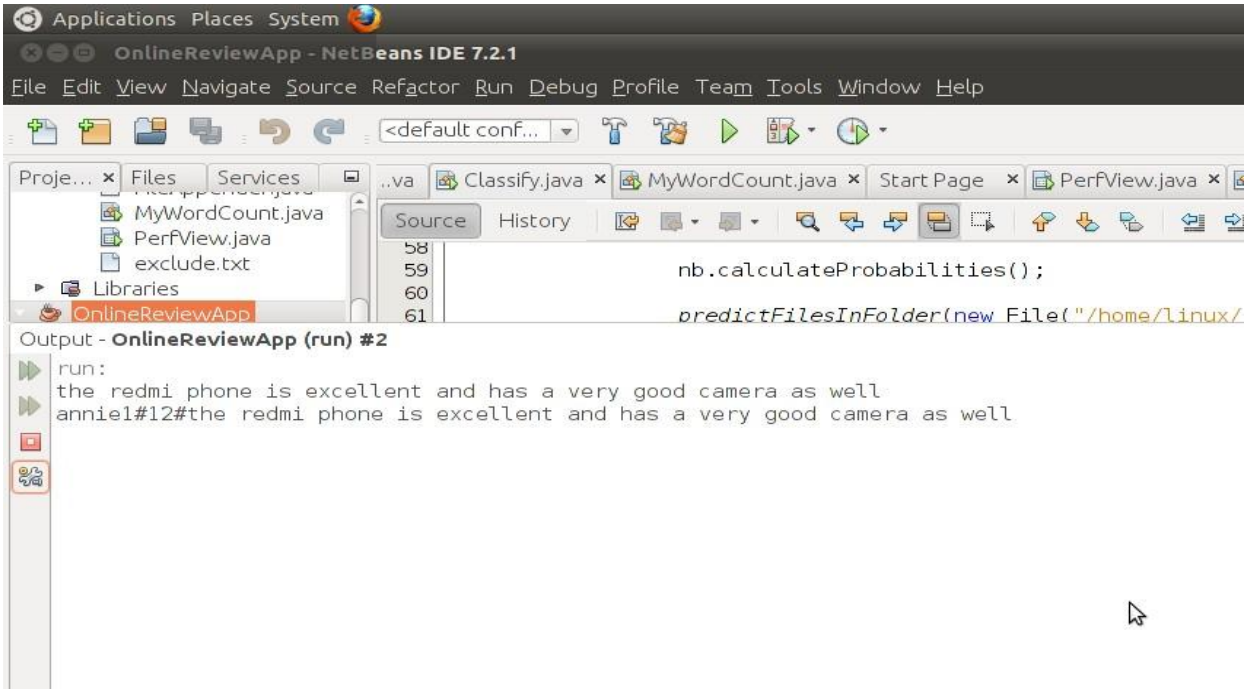


Fig 1.2: User review added

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 1 , January 2020

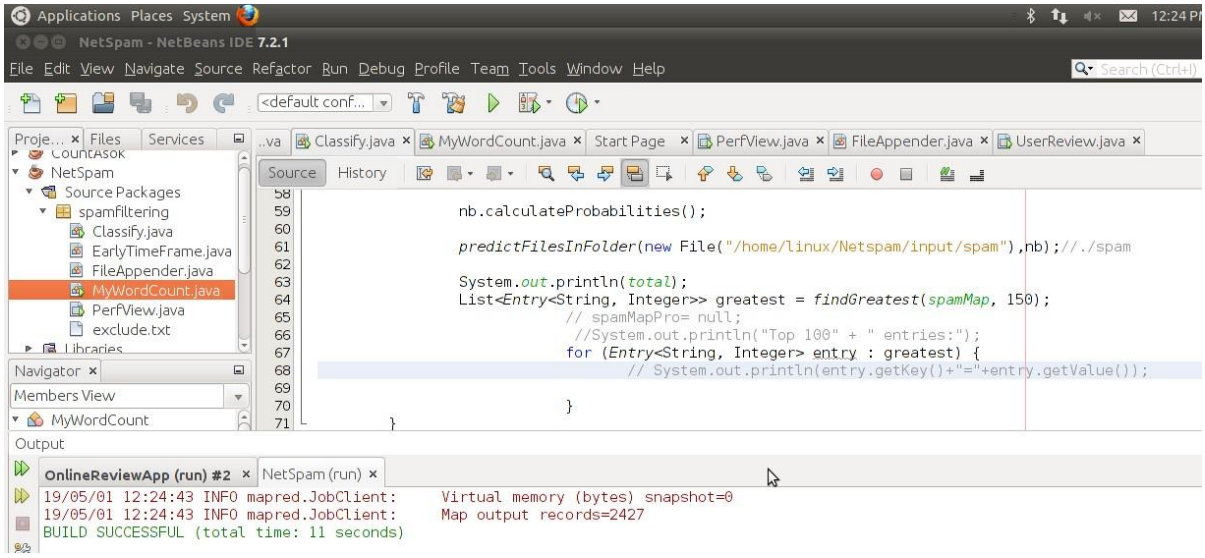


Fig 1.3: MyWordCount.java executed

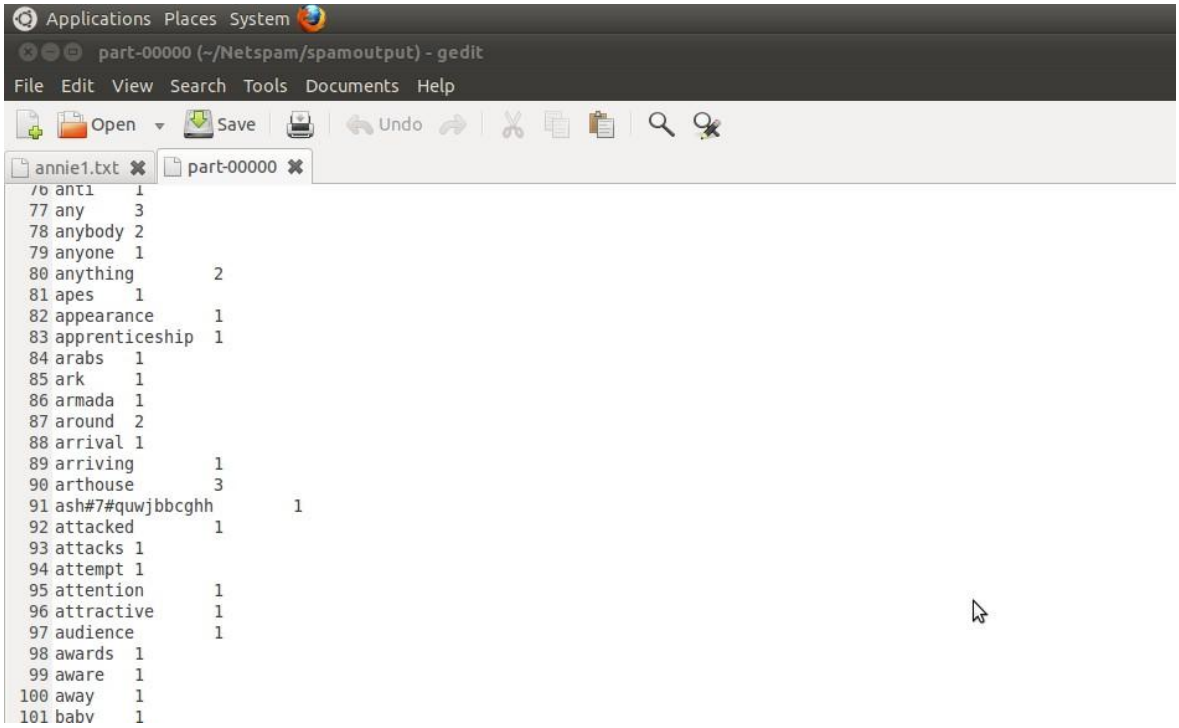
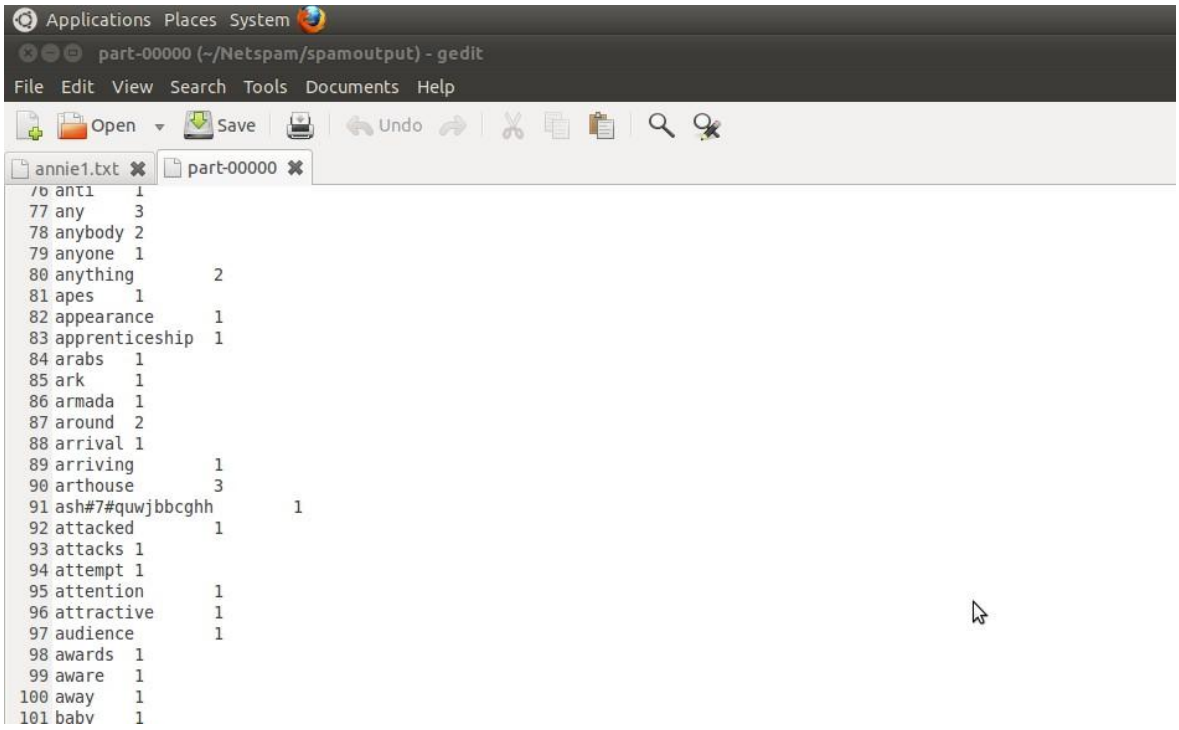


Fig 1.4 Spam Output folder (feature selection)

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 1 , January 2020



IV. DISCUSSIONS

The obtained results show based on Behavioural and Linguistic features method .The results are able to predict and improve the quality owner's products

V. CONCLUSION

This study introduces a novel spam detection framework namely NetSpam based on a metapath concept as well as a new graph-based method to label reviews relying on a rank-based labeling approach. The performance of the proposed framework is evaluated by using two real-world labeled datasets of yelp and Amazon websites. Our observation show that calculates weights by using this metapath concept can be very effective in identifying spam reviews and leads to a better performance. In addition, we found that even without a train set, NetSpam can calculate the importance of each feature and it yields better performance in the features' addition process, and performs better than previous works, with only a small number of features. Moreover, after defining four main categories for features our observations show that the reviews behavioral category performs better than other categories, in terms of AP, AUC as well as in the calculated weights. The results also confirm that using different supervisions, similar to the semi-supervised method, have no noticeable effective in determining most of the weighted features, just as indifferent dataset.

For future work, metapath concept can be applied to other problems in this field. For example, similar framework can be used to find spammer communities. For finding community, reviews can be connected through group spammer features (such as the proposed feature in and reviews with highest similarity based on metapath concept are known as communities. In addition, utilizing the products features is an interesting future work on this study as we used features more related to spotting spammers and spam reviews. Moreover, while networks has received considerable attention from various disciplines for over a decade. Information diffusion and content sharing in multilayer networks is still a young research. Addressing the problem of spam detection in such networks can be considered as a new research line in this field.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 1 , January 2020

REFERENCES

[1] J. Donfro, A whopping 20 % of yelp reviews are fake. http://www.businessinsider.com/20-percent-of-yelp-reviews-fake-2013-9. Accessed: 2015-07-30.
[2] M. Ott, C. Cardie, and J. T. Hancock. Estimating the prevalence of deception in online review communities. In ACM WWW, 2012.
[3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock. Finding deceptive opinion spam by any stretch of the imagination.In ACL, 2011.
[4] Ch. Xu and J. Zhang. Combating product review spam campaigns via multiple heterogeneous pair wise features. In SIAM International Conference on Data Mining, 2014.
[5] N. Jindal and B. Liu. Opinion spam and analysis. In WSDM, 2008.
[6] F. Li, M. Huang, Y. Yang, and X. Zhu. Learning to identify review spam. Proceedings of the 22nd International Joint Conference on Artificial Intelligence; IJCAI, 2011.
[7] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. Exploiting burstiness in reviews for review spammer detection. In ICWSM, 2013.
[8] A. j. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos. Trueview: Harnessing the power of multiple review sites. In ACM WWW, 2015.
[9] B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Towards detecting anomalous user behavior in online social networks. In USENIX, 2014.
NETSPAM: A Network based Spam Detection Framework for Reviews in Online Social Media Dept. of CSE, EPCET 2018-19 Page 44
[10] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao. Spotting fake reviews via collective PU learning. In ICDM, 2014.
[11] L. Akoglu, R. Chandy, and C. Faloutsos. Opinion fraud detection in online reviews bynetwork effects. In ICWSM, 2013.
[12] R. Shebuti and L. Akoglu. Collective opinion spam detection: bridging review networks and metadata. In ACM KDD, 2015.
[13] S. Feng, R. Banerjee and Y. Choi. Syntactic stylometry for deception detection. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers; ACL, 2012.
[14] N. Jindal, B. Liu, and E.-P. Lim. Finding unusual review patterns using unexpected rules. In ACM CIKM, 2012.
[15] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting product review spammers using rating behaviors. In ACM CIKM, 2010.
[16] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh. Spotting opinion spammers using behavioral footprints. In ACM KDD, 2013.
[17] S. Xie, G. Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In ACM KDD, 2012.
[18] G. Wang, S. Xie, B. Liu, and P. S. Yu. Review graph based online store review spammer detection. IEEE ICDM, 2011.
[19] Y. Sun and J. Han. Mining Heterogeneous Information Networks; Principles and Methodologies, In ICCCE, 2012.