International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 11 | Nov -2017 www.irjet.net p-ISSN: 2395-0072

Survey in Online social media Skelton by network based spam

Satish Tukaram Pokharkar[1], Ajit Jaysingrao Shete[2], Vishal Dyandeo Ghogare[3]

[1]*Satish .T.Pokharkar, SCSCOE Rahuri Factory, Maharashtra, India*
[2]*Ajit Jaysingrao Shete, SCSCOE Rahuri Factory, Maharashtra, India*
[3]*Vishal Dyandeo Ghogare,* Ashok *Institute of Engineering and Technology, Ashoknagar, Maharashtra, India*

Abstract - *These days, a major piece of individuals depend on accessible substance in web-based social networking in their choices (e.g. audits and criticism on a subject or item). The likelihood that anyone can clear out a survey give a brilliant chance to spammers to compose spam surveys about items and administrations for various interests. Recognizing these spammers and the spam content is a hotly debated issue of research and in spite of the fact that an extensive number of studies have been done as of late toward this end, yet so far the techniques set forth still scarcely recognize spam surveys, and none of them demonstrate the significance of each removed element sort. In this examination, we propose a novel system, named Net Spam, which uses spam highlights for displaying audit datasets as heterogeneous data systems to delineate identification strategy into a characterization issue in such systems. Utilizing the significance of spam highlights help us to acquire better outcomes as far as distinctive measurements investigated genuine audit datasets from Yelp and Amazon sites. The outcomes demonstrate that Net Spam beats the current strategies and among four classes of highlights; including audit behavioral, client behavioral, review linguistic, client semantic, the primary kind of highlights performs better Than alternate classifications*

Key Words: Social Media, Social Network, Spammer, Spam Review, Fake Review, Heterogeneous Information Networks

1. INTRODUCTION

Online Social Media entries assume a persuasive part in Data spread which is considered as a vital hotspot for makers in their publicizing efforts as well with respect to clients in choosing items and administrations. In the previous years, individuals depend a considerable measure on the composed surveys in their basic leadership procedures, and positive/negative surveys empowering/debilitating them in their choice of items furthermore, administrations. What's more, composed surveys additionally help benefit suppliers to improve the nature of their items and administrations.

These surveys in this manner have turned into a vital factor in progress of a business while positive audits can bring benefits for a organization, negative surveys can possibly affect validity what's more, cause monetary misfortunes. The way that anybody with any character can leave remarks as audit, gives an enticing open door for spammers to compose counterfeit audits intended to delude clients' sentiment. These deceptive audits are at that point duplicated by the sharing capacity of web-based social networking and proliferation over the web. The surveys written to change clients' impression of how great an item or an administration are considered as spam and are regularly composed in return for cash As appeared in [1], 20% of the surveys in the Yelp site are all things considered spam surveys. Then again, a lot of writing has been distributed on the systems used to recognize spam and spammers and additionally extraordinary kind of investigation on this subject These methods can be characterized into various classifications; some utilizing semantic examples in content [2], [3], [4], which are for the most part in view of bigram, and unigram, others are in light of behavioural examples that depend on highlights separated from designs in clients' conduct which are for the most part metadata based. Regardless of this incredible arrangement of endeavours, numerous angles have been missed or stayed unsolved. One of them is a classifier that can ascertain include weights that demonstrate each element's level of significance in deciding spam surveys. The general idea of our proposed structure is to show a given survey dataset as a Heterogeneous Information Network (HIN)and to outline issue of spam discovery into a HIN order issue. Specifically, we show survey dataset as a HIN in which surveys are associated through various hub sorts (for example, highlights and clients). A weighting calculation is at that point utilized to compute each component's significance (or weight). These weights are used to figure the last names for surveys utilizing both unsupervised and administered approaches.
To assess the proposed arrangement, we utilized two specimen survey datasets from Yelp and Amazon sites. In light of our perceptions, characterizing two perspectives for highlights (survey client furthermore, behavioural-phonetic), the arranged highlights as review behavioural have more weights and yield better execution on spotting spam audits in both semi-managed and unsupervised methodologies. Likewise, we exhibit that utilizing diverse supervisions, for example, 1%, 2.5% and 5% or utilizing an unsupervised approach, make no perceptible minor departure from the execution of our approach. We watched that component weights can be included or evacuated for marking and subsequently time many-sided quality can be scaled for a particular level of exactness. As the consequence of this weighting step, we can utilize less highlights with more weights to get better precision with Less time many-sided quality. Also, ordering highlights in four real classes (survey behavioural, client behavioural, review linguistic, client phonetic), encourages us to see how much every classification of highlights is added to spam recognition.

(I) we propose Net Spam system that is a novel network based approach which models survey organizes as heterogeneous data systems. The grouping step utilizes distinctive Meta path sorts which are imaginative in the spam recognition space.

(ii) another weighting strategy for spam highlights is proposed to decide the relative significance of each component what's more, indicates how viable each of highlights are in recognizing spasms from typical surveys. Past works [12], [20] too planned to address the significance of highlights for the most part in term of got precision, yet not as a work in work in their structure (i.e., their approach is reliant to ground truth for deciding each component significance). As we clarify in our unsupervised approach, Net Spam can discover highlights significance even without ground truth, and just by depending on Meta path definition and in light of qualities ascertained for each survey.

(iii) Net Spam enhances the precision contrasted with the state of-the craftsmanship as far as time intricacy, which exceptionally depends to the quantity of highlights used to recognize a spam survey; subsequently, utilizing highlights with more weights will brought about recognizing Counterfeit surveys less demanding with less time intricacy.

2. PRELIMINARIES

As specified before, we demonstrate the issue as a heterogeneous system where hubs are either genuine segments in a dataset, (for example, audits, clients and items) or spam highlights. To better comprehend the proposed structure we first exhibit a diagram of a portion of the ideas and definitions in heterogeneous data systems [23], [22], [24]

2.1.1 Definitions1 (Heterogeneous Information Network)

Assume we have r(> 1) sorts of hubs and s(> 1) sorts of connection interfaces between the hubs, at that point a heterogeneous data arrange is characterized as a diagram G = (V; E) where every hub v ε V and each connection e εE has a place with one specific hub sort and connection sort separately. On the off chance that two connections have a place with a similar sort, the sorts of beginning hub and consummation hub of those connections are the same

2.1.2 Definitions 2 (Network Schema)

Given a heterogeneous data organize G = (V; E), a system outline T = (A; R) is a meta path with the protest sort mapping : V ! A also, interface mapping - : E ! R, which is a chart characterized over question sort A, with joins as relations from R. The pattern Depicts the met structure of a given system (i.e., what number of hub sorts there are and where the conceivable connections exist).

2.1.3Definition 3 (Metapath)

As said above, there are no edges between two hubs of a similar sort, yet there are ways. Given a heterogeneous data arrange G = (V; E), a metapath P is characterized by a succession of relations in the system outline T = (A; R), indicated in the frame A1(R1)A2(R2).....(R(l 1))Al , which characterizes a composite connection P = R1oR 2...R (l 1) between two hubs, where o is the synthesis administrator on relations. For accommodation, a metapath can be spoken to by an arrangement of hub sorts when there is no equivocalness, i.e., P = A1A2:::A l . The metapath broadens the idea of connection sorts to way sorts and portrays the diverse relations among hub sorts through circuitous connections, i.e. ways, and furthermore infers different semantics

2.1.4Definition 4(Classification problem in

heterogeneous information networks)Given a heterogeneous data arrange G = (V; E), assume V' is a subset of V that contains hubs of the objective sort (i.e., the kind of hubs to be grouped). K means the quantity of the class, and for each class, say C1.....C k , we have some pre-marked hubs in V' related with a solitary client. The characterization errand is to foresee the marks for all the unlabeled hubs in V'.

2.1.2 Feature Types

In this paper, we utilize a broadened meaning of the metapath idea as takes after. A metapath is characterized as a way between two hubs, which shows the association of two hubs through their mutual highlights. When we discuss metadata, we allude to its general definition, which is information about information. In our case, the information is the composed audit, and by metadata we mean information about the audits, including client who composed the audit, the business that the survey is composed for, rating esteem of the audit, date of composed survey lastly its name as spam or veritable audit. Specifically, in this work highlights for clients and audits fall into the classes as take after

Review-Behavioral (RB) based features:

This feature type is based on metadata and not the review text itself. The RBcategory contains two features; Early time frame (ETF) and Threshold rating deviation of review (DEV) [16]

Review-Linguistic (RL) based features:

This component sort depends on metadata and not simply the audit content. The RB classification contains two highlights; Early time span (ETF) and Edge rating deviation of audit (DEV) [16].

User-Behavioral (UB) based features:

These highlights are particular to every individual client and they are computed per client, so we can utilize these highlights to sum up the greater part of the surveys composed by that particular client. This classification has two primary highlights; the Burstiness of surveys composed by a solitary client [7], and the normal of a clients' negative proportion given to distinctive organizations [20].

User-Linguistic (UL) based features: These features are Extracted from the users' language and shows how users are describing their feeling or opinion about what they've Experienced as a customer of a certain business. We use this type of features to understand how a spammer communicates in terms of wording. There are two features engaged for our framework in this category; Average Content Similarity (ACS) and Maximum Content Similarity (MCS). These two features how much two reviews written by two different users are similar to each other, as spammers tend to write very similar views by using template pre-written text [11].

3 N ETSPAM; T HE PROPOSED SOLUTION

3.1 Prior Knowledge

Behavioral based Features (User-based);

Burstiness [20]: 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 Negative Ratio [20]: 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 others take 0.

Behavioral based Features (Review-based):

Early Time Frame [16]: Spammers try to write their reviews ASAP, in order to keep their review in the top reviews which other users visit them sooner Rate Deviation using threshold [16]: 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?

3.2 Network Schema Definition

The following stage is characterizing system blueprint in view of guaranteed rundown of spam highlights which decides the highlights occupied with spam discovery. This Schema are general meanings of metapaths and show all in all how unique system parts are associated. For instance, if the rundown of highlights incorporates NR, ACS, PP1 and ETF, the yield blueprint is as introduced in Fig1

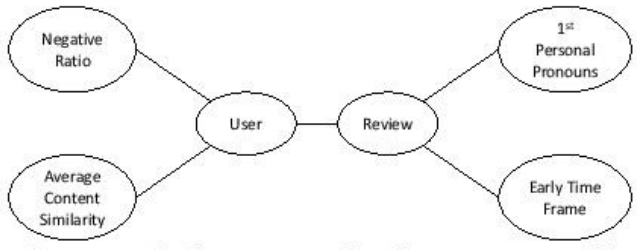


Fig. 1: An example for a network schema generated based on a given spam features list; NR, ACS, PP1 and ETF

3.3 Metapath Definition and Creation

a metapth is characterized by a grouping of relations in the system schema As appeared, the length of client based metapaths is 4 and the length of review based metapth is 2. For metapath creation, we characterize an expanded rendition of the metapath idea considering distinctive levels of spam conviction. Specifically, two audits are associated with each other on the off chance that they share same esteem. Hassanzadeh et al. [25] propose a fluffy based system and show for spam discovery, it is better to utilize fluffy rationale for deciding an audit's mark as an As appeared, the length of client based metapaths is 4 and the length of review based metapth is 2. For metapath creation, we characterize an expanded adaptation of the metapath idea considering diverse levels of spam. Specifically, two surveys are associated with each other on the off chance that they share same esteem. Hassanzadeh et al. [25] propose a fluffy based system and demonstrate for spam identification, it is better to utilize fluffy rationale for deciding an audit's name as spam or non-spam. Without a doubt, there are diverse levels of spam conviction. We utilize a stage capacity to decide these levels. In specific, given a survey u, the levels of spam sureness for metapath p l (i.e., highlight l) is ascertained as m p l u = bs□f (xlu)c s , where s signifies the quantity of levels. Subsequent to registering m p l u for all surveys and metapaths, two audits u and v with the same metapath esteems (i.e., m p l u = m p l v) for metapath p l are Associated with each other through that metapath and make one connection of survey arrange. The metapath esteem between them signified as m p l u;v = m p l u . Utilizing s with a higher esteem will expand the quantity of Each component's metapaths and thus less audits would be Associated with each other through these highlights. On the other hand, utilizing lower an incentive for s drives us to have bipolar esteems (which implies audits take esteem 0 or 1). Since we require enough spam and non-spam surveys for each progression, with less number of surveys associated with each other for each progression, the spam likelihood of audits take uniform dispersion, yet with lower estimation of s we have enough audits to ascertain last spam city for each audit. Along these lines, precision for bring down levels of s Diminishes in view of the bipolar issue, and it decades for higher estimations of s, since they take uniform dissemination. In the proposed system, we considered s = 20, i.em p l u 2 f{0; 0:05; 0:10; …… 0:85; 0:90; 0:95g}

Empowered by fast advances in sequencing innovation, met genomic contemplates mean to portray whole groups of microorganisms bypassing the requirement for refined individual bacterial individuals. One noteworthy objective of met genomic thinks about is to recognize particular useful adjustments of microbial groups to their environments. The useful profile and the plenitudes for an example can be evaluated by mapping met genomic successions to the worldwide metabolic system comprising of thousands of sub-atomic responses. Here we depict a capable logical technique (Metapth) that can recognize differentially rich pathways in met genomic datasets**,** depending on a mix of met genomic succession information and earlier metabolic pathway learning.

3.4 Classification

The arrangement part of Net Spam incorporates two stages; (I) weight count which decides the significance of each spam include in spotting spam surveys, (ii) Labeling which figures the last likelihood of each audit being spam. Next we portray them in detail.

Weight Calculation:

This progression registers the heaviness of each metapth. We accept that hubs' characterization is finished in view of their relations to different hubs in the audit arrange; connected hubs may have a high likelihood of taking the same names. The relations in a heterogeneous data organize include the immediate connection as well as the way that can be measured by utilizing the metapth idea. Along these lines, we require to use the metapth characterized in the past advance, which speak to heterogeneous relations among hubs. In addition, this step will have the capacity to figure the heaviness of every connection way (i.e., the significance of the metapth), which will be utilized as a part of the following stage (Labeling) to gauge the mark of each unlabeled survey. The weights of the metapth will answer an essential question; which metapth (i.e., spam highlight) is better at positioning spam surveys? Also, the weights help us to get it the development instrument of a spam survey. What's more, since some of these spam highlights may acquire impressive computational expenses (for instance, processing etymological based highlights through NLP techniques in a substantial audit dataset), picking the more profitable highlights in the spam identification methodology prompts better execution at whatever point the calculation cost is an issue.

Labeling:

It is worth to take note of that in making the HIN, as much as the number of connections between a survey and different audits increment, its likelihood to have a name like them increment as well, since it accept that a hub connection to different hubs appear their likeness. Specifically, more connections between a hub and other non-spam audits, greater likelihood for a survey to be non-spam and the other way around. At the end of the day, if a survey has heaps of connections with non-spam audits, it implies that it shares highlights with different audits with low spam city and thus its likelihood to be a non-spam survey increments the requirement for refined individual bacterial individuals. One noteworthy objective of met genomic thinks about is to recognize particular useful adjustments of microbial groups to their environments. The useful profile and the plenitudes for an example can be evaluated by mapping met genomic successions to the worldwide metabolic system comprising of thousands of sub-atomic responses. Here we depict a capable logical technique (Metapth) that can recognize differentially rich pathways in met genomic datasets**,** depending on a mix of met genomic

4. NETSPAM Algorithm:

$Input: review-dataset, spam-feature-list,$
$pre-labeled-reviews$
$Output: features-importance(W),$
$spamicity-probability(Pr)$
$\%\ u, v$: review, y_u: spamicity probability of review u
$\%\ f(x_{lu})$: 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
$\%\ m_u^{p_l}$: the level of spam certainty
$\%\ m_{u,v}^{p_l}$: the metapath value
%Prior Knowledge
if semi-supervised mode
$\begin{cases} \textbf{if } u \in pre-labeled-reviews \\ \{y_u = label(u) \\ \quad \textbf{else} \\ \{y_u = 0 \end{cases}$
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 $p_l \in schema$
$\textbf{do} \begin{cases} \textbf{for } u, v \in review-dataset \\ \quad \textbf{do} \begin{cases} m_u^{p_l} = \frac{\lfloor s \times f(x_{lu}) \rfloor}{s} \\ m_v^{p_l} = \frac{\lfloor s \times f(x_{lv}) \rfloor}{s} \\ \textbf{if } m_u^{p_l} = m_v^{p_l} \\ \{mp_{u,v}^{p_l} = m_u^{p_l} \\ \quad \textbf{else} \\ \{mp_{u,v}^{p_l} = 0 \end{cases} \end{cases}$
% Classification - Weight Calculation
for $p_l \in schemes$
$\textbf{do}\ \{W_{p_l} = \frac{\sum_{r=1}^{n}\sum_{s=1}^{n} mp_{r,s}^{p_l} \times y_r \times y_s}{\sum_{r=1}^{n}\sum_{s=1}^{n} mp_{r,s}^{p_l}}$
% Classification - Labeling
for $u, v \in review-dataset$
$\textbf{do} \begin{cases} Pr_{u,v} = 1 - \Pi_{p_l=1}^{L} 1 - mp_{u,v}^{p_l} \times W_{p_l} \\ \\ Pr_u = avg(Pr_{u,1}, Pr_{u,2}, ..., Pr_{u,n}) \end{cases}$
return (W, Pr)

5. E XPERIMENTAL EVALUATION

Datasets:

Dataset	Reviews (spam%)	Users	Business (Resto. & hotels)
Main	608,598 (13%)	260,277	5,044
Review-based	62,990 (13%)	48,121	3,278
Item-based	66,841 (34%)	52,453	4,588
User-based	183,963 (19%)	150,278	4,568
Amazon	8,160 (-)	7685	243

Incorporates an outline of the datasets and their attributes. We utilized a dataset from Yelp, presented in [12], which incorporates very nearly 608,598 surveys composed by clients of eateries and lodgings in NYC. The dataset incorporates the commentators' impressions and remarks about the quality, and different perspectives identified with an eateries (or inns). The dataset additionally contains named surveys as ground truth (purported close ground-truth [12]), which demonstrates whether a survey is spam or, on the other hand not. Cry dataset was named utilizing sifting calculation connected with by the Yelp recommender, and albeit none of recommenders are immaculate, however as indicated by [36] it produces trustable outcomes. It discloses enlisting somebody to compose extraordinary counterfeit surveys on various web-based social networking locales, it is the cry calculation that can spot spam surveys and rank one particular spammer at the highest point of spammers. Different characteristics in the dataset are rate of commentators, the date of the composed audit, and date of real visit, and additionally the client's and the eatery's id (name).We made three different datasets from this primary dataset as take after:- Review-based dataset, incorporates 10% of the surveys from the Main dataset, haphazardly chose utilizing uniform circulation. - Item-based dataset, makes out of 10% of the haphazardly chose surveys of everything, likewise in view of uniform dissemination (similarly as with Review-based dataset). - User-based dataset, incorporates haphazardly chose surveys utilizing uniform conveyance in which one survey is chosen from each 10 surveys of single client and if number of audits was under 10, uniform appropriation has been changed in request to no less than one survey from each client get chose

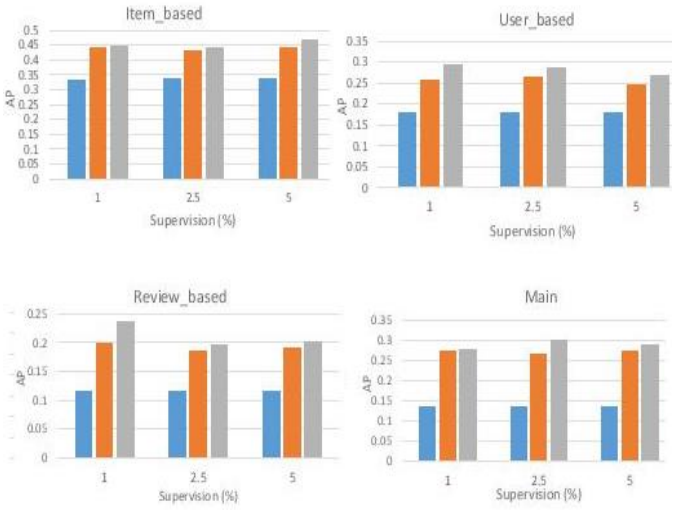
Evaluation Metrics:

We have utilized Average Precision (AP) and Area Under the Bend (AUC) as two measurements in our assessment. AUC measures precision of our positioning in view of False Positive Ratio (FPRas y-hub) against True Positive Ratio (TPR as x-pivot) and incorporate esteems in view of these two measured esteems. The estimation of these metric increments as the proposed strategy performs well in positioning, and tight clamp versa. Let A be the rundown of arranged spam audits with the goal that A(i) means a survey arranged on the I th record in A. On the off chance that the quantity of spam (non-spam) audits some time recently audit in the j th file is equivalent to nj and the aggregate number of spam (non-spam) audits is equivalent to f , then TPR (FPR) for the j th is registered as nj f . To figure the AUC, we set T P R esteems as the x-hub and F P R esteems on the y-hub and at that point incorporate the zone under the bend for the bend that employments their qualities

Main Results:

In this segment, we assess Net Spam from alternate point of view and contrast it and two different methodologies, Random approach and SPeaglePlus [12]. To contrast and the first one, we have built up a system in which audits are associated with each other arbitrarily. Second approach utilize a well-known diagram based calculation called as "LBP" to ascertain last marks. Our perceptions indicate Net Spam, outflanks these current strategies. At that point effect of investigation on our perception is performed lastly we will analyze our system in unsupervised mode. In conclusion, we research time many-sided quality of the proposed structure and the cover system on its execution

Accuracy:



FigAP for Random, SPeaglePlus and NetSpam approaches in different datasets and supervisions (1%, 2.5% and 5%)

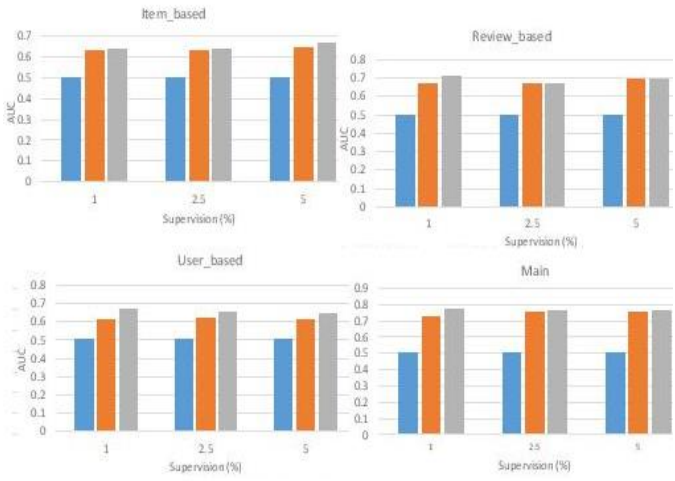


Fig AUC for Random, SPeaglePlus and NetSpam approaches in different datasets and supervisions (1%, 2.5% and 5%).

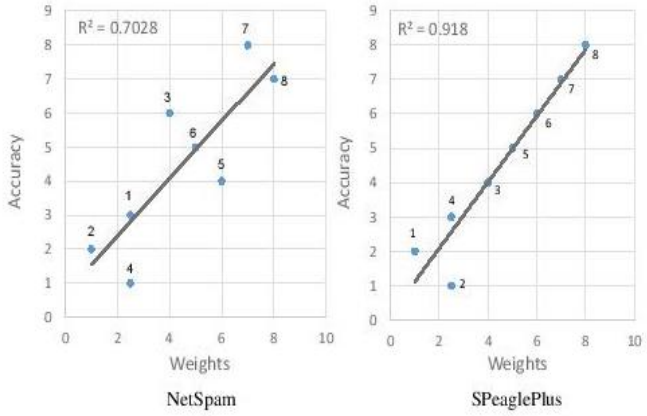


Fig Regression graph of features vs. accuracy

4. CONCLUSIONS

This examination presents a novel spam recognition system Specifically Net Spam in light of a metadata idea too as a new chart based strategy to name audits depending on a rank-based naming methodology. The execution of the proposed system is assessed by utilizing two certifiable named datasets of Yelp and Amazon sites. Our perceptions appear that figured weights by utilizing this meta path idea can be Exceptionally powerful in distinguishing spam surveys and prompts a superior execution. What's more, we found that even without a prepare set, Net Spam can figure the significance of each component also, it yields better execution in the highlights' expansion process, and performs superior to anything past works, with just a modest number of highlights. In addition, in the wake of characterizing four primary classes for highlights our perceptions demonstrate that the reviews behavioral classification performs superior to different classifications, in terms of AP, AUC and in addition in the computed weights. The comes about additionally affirm that utilizing distinctive supervisions, comparative to the semi-administered technique, have no perceptible impact on deciding a large portion of the weighted highlights, similarly as in various datasets. For future work, multipath idea can be connected to other issues in this field. For instance, comparable structure can be used to discover spammer groups. For discovering group, surveys can be associated through gathering spammer highlights (for example, the proposed highlight in [29]) and audits with most astounding comparability in light of metapth idea are known as groups. Furthermore, using the item includes is an Intriguing future work on this investigation as we utilized highlights more identified with spotting spammers and spam audits. Also, while single systems has gotten significant consideration from different orders for over 10 years, data dissemination what's more, content partaking in multilayer systems is as yet a youthful research Addressing the issue of spam recognition in such systems can be considered as another examination line in this field

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 ArtificialIntelligence; IJCAI, 2011.

[7] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Gosh. 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.

[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 by network effects. In ICWSM, 2013.

[12] R. Shebuti and L. Akoglu. Collective opinion spam detection: bridging review network sand 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. Gosh. Spotting opinion spammers using behavioral footprints. InACM 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 storage view spammer detection. IEEE ICDM, 2011.

[19] Y. Sun and J. Han. Mining Heterogeneous Information Networks; Principles and Methodologies, In ICCCE, 2012.

[20] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance. What Yelp Fake Review Filter Might Be Doing?, In ICWSM, 2013.

[21] S. Feng, L. Xing, A. Gogar, and Y. Choi. Distributional footprints of deceptive product reviews. In ICWSM, 2012.

[22] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. In VLDB,2011.

[23] Y. Sun and J. Han. Rankclus: integrating clustering with ranking for heterogeneous information network analysis. In Proceedings of the 12thInternational Conference on Extending Database Technology: Advances in Database Technology, 2009.

[24] C. Luo, R. Guan, Z. Wang, and C. Lin. HetPathMine: Novel Transductive Classification Algorithm on Heterogeneous Information Networks. InECIR, 2014.

[25] R. Hassanzadeh. Anomaly Detection in Online Social Networks: Using Data mining Techniques and Fuzzy Logic. Queensland University of Technology, Nov. 2014.

[26] M. Luca and G. Zervas. Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud., SSRN Electronic Journal, 2016.

[27] E. D. Wayne and A. Djunaidy. Fake Review Detection From a Product Review Using Modified Method of Iterative Computation Framework. In Proceeding MATEC Web of Conferences. 2016.

[28] M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa. Reducing Feature set Explosion to Facilitate Real-World Review Spam Detection. In Proceedingof 29th International Florida Artificial Intelligence Research Society Conference. 2016.

[29] A. Mukherjee, B. Liu, and N. Glance. Spotting Fake Reviewer Grouping Consumer Reviews. In ACM WWW, 2012