Analysis on Increasing Intelligence of Viral Marketing in Social Network Using Information Diffusion

S. N. Salunkhe[1] A. V. Chavan[2] R. P. Choudhar[3] U. S. Lathor[4]
[1,2,3,4]B.E Student
[1,2,3,4]Department of Computer Engineering
[1,2,3,4]S.B. Patil C.O.E. Indapur, 413106, Savitribai Phule, Pune University, Maharashtra, India

Abstract— The rising of mobile social networks opens opportunities for infectious agent selling. However, before absolutely utilizing mobile social networks as a platform for infectious agent selling, several challenges have to be compelled to be addressed. during this paper, we have a tendency to address the matter of distinguishing a little variety of people through whom the knowledge will be subtle to the network as presently as potential, spoken because the diffusion reduction downside. Diffusion reduction beneath the probabilistic diffusion model will be developed as AN uneven k- centre downside that is NP-hard, and also the best best-known approximation algorithmic rule for the uneven k-centre downside has approximation quantitative relation of log n and time complexness O(n5). Clearly, the performance and also the time complexness of the approximation algorithmic rule don't seem to be satiable in large-scale mobile social networks. To subsume this downside, we have a tendency to propose a community primarily based algorithmic rule and a distributed set-cover algorithmic rule. The performance of the planned algorithms is evaluated by in depth experiments on each artificial networks and a true trace. The results show that the community primarily based algorithmic rule has the simplest performance in both synthetic networks and there altrace compared to existing algorithms, and also the distributed set-cover algorithmic rule out performs the approximation algorithmic rule within the real trace in terms of diffusion time.

Key words: Data Diffusion, Mobile Social Networks, Community Structure

I. INTRODUCTION

Social network plays a very important role for spreading data, plan and influence among its members. Nowadays, social networks are evolving to on-line social networks like Facebook, Twitter, and Google+ that link humans, computers and also the net, and data spreading in social networks has been modified from the means of “word-of-mouth” to “word-of-text”, “word-of-voice”, “word-of-photo” and “word-of-video”. Additionally, with the proliferation of good mobile devices, like smartphone and pill, folks will simply go surfing with their mobile devices, meantime a lot of and a lot of native mobile social networks are created like Foursquare, Instagram, and Path. Moreover, Bluetooth and wireless local area network Direct extend communications between mobile devices from the restrictions of cellular infrastructure; user quality and social property bring various ad-hoc communication opportunities.

As the essence of infectious agent selling applications is data diffusion from little variety of people to the whole network by “word-of-mouth”, during this paper, we have a tendency to address the matter of distinguishing a little variety of people through whom the knowledge will be subtle to the whole network as presently as potential, spoken because the diffusion reduction downside. Diffusion reduction is of course vital to infectious agent selling applications. For instance, the “word-of-mouth” promotion caught to be disseminated to the network as presently as potential, and so it might be of interest to several firms further as people that need to extend complete awareness, or publicize advertisements or innovative concepts through “word-of-mouth”. For instance, a corporation would love to quickly raise the notice of a brand new product in a very network. The corporate at first offers free samples of the merchandise to a little variety of people within the network (the product is pricey or the corporate has restricted budge such they will solely select a little variety of people). The corporate hopes that the first elite users can unfold the knowledge of the new product to their friends, and their friends can propagate the knowledge to their friends’ friends then on.

II. LITERATURE SURVEY

A. “Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates Selection”- Zongqing Lu, Yonggang Wen,Weizhan Zhang, Qinghua Zheng, and Guohong Cao .[1]

In this article, at first we highlight the problem due to the complexity of social networks, few models exist to interpret social network marketing, social network marketing using Heat Diffusion Processes.

B. “SelfInterest Driven incentives for ad dissemination in autonomous mobile social networks.”-T. Ning, Z. Yang, H. Wu, and Z. Han.[2]

In this paper they discuss the to eliminate the needs of accurate knowledge about whom and how many credits ad provider should pay. SelfInterestDriven (SID) incentive scheme to stimulate cooperation among selfish nodes.

C. “Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags”,D. M. Romero, B. Meeder, and J. Kleinberg[3]

In this paper they discuss the difficult to evaluate since it requires a setting where many different kinds of information spread in a shared environment. They provide the solution for that is develop the simulation based and generative models to analyze.

D. “Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags”,D. M. Romero, B. Meeder, and J. Kleinberg[3]

In this paper they discuss the difficult to evaluate since it requires a setting where many different kinds of information spread in a shared environment. They provide the solution for that is develop the simulation based and generative models to analyse.

E. "To maximize the performance and Set-cover outperforms Approximation and Nave in terms of diffusion time",Z. Lu, Y. Wen, and G. Cao .[5]

They intend to improve community detection in weighted networks and exploit community for data forwarding in DTN and worm containment in OSN. Novel community detection algorithm, and then introduce two metrics called intracentrality and intercentrality, to characterize nodes in communities. Adaptive framework, and novel approach for scalable community detection.

III. SYSTEM DESCRIPTION

The system contains the various modules:
In this system, we are going to find the intent of short-text document on social media.

A. Collection of Tweets

For this firstly collects tweets from Social Media API for extracting options.

B. Feature Extractor

Feature extractor is finished victimisation completely different guided pattern like declarative data guided pattern, social data guided pattern, distinction mining pattern.

C. Subset Generator

when this we are going to generate subsets of options and options are validate victimisation classification rule.

D. Classification

The category classification models square measure in the main wont to assign the category label among the obtainable class values to a brand new tuple. This can be done by victimisation classifier Model.

E. Knowledgebase

there'll be predefined intents and that we are use content to avoid wasting that intents and validate options. Thus information are hold on in knowledgebase.

F. Prediction Phase

Then in prediction part we are going to match the options and intents victimisation classifier model. The classifier model will be engineered by applying varied learning algorithms like multi-classification such i.e. Naive Bayesian classification, K-Nearest Neighbour (KNN). Finally feature and intent are match at classifier model. And potential intent of short-text are find.

IV. ALGORITHM

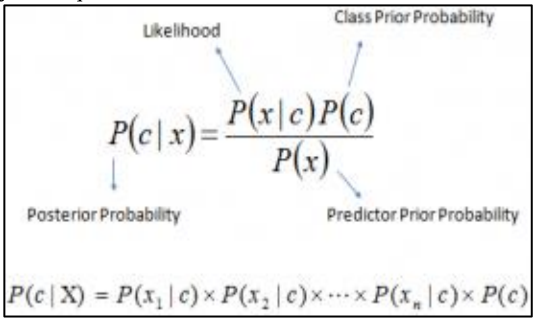
A. Naïve Bayes

Naïve Bayes classifiers assume that there are not any dependencies amongst attributes. This assumption is named category conditional independence. It's created to change the computations concerned and, thus is named "naive". This classifier is additionally known as moron Thomas Bayes, easy Thomas Bayes, or freelance Thomas Bayes.

It's a classification technique supported Bayes' Theorem with associate assumption of independence among predictors. In easy terms, a Naive Thomas Bayes classifier assumes that the presence of a specific feature during a class is unrelated to the presence of the other feature. For instance, a fruit could also be thought-about to be associate apple if it's red, round, and regarding three inches in diameter. Not with standing these options rely on one another or upon the existence of the opposite options, all of those properties severally contribute to the chance that this fruit is associate apple which is why it's referred to as 'Naive'.

Naive Thomas Bayes model is straightforward to create and notably helpful for terribly giant information sets. Beside simplicity, Naive Thomas Bayes is thought to exceed even extremely subtle classification strategies.

Bayes theorem provides the simplest way of shrewd posterior chance P(c|x) from P(c), P(x) and P(x|c). Verify the equation below:



Where,

- P(c|x) is that the posterior chance of sophistication (c, target) given predictor (x, attributes).
- P(c) is that the previous chance of sophistication.
- P(x|c) is that the probability that is that the chance of predictor given category.
- P(x) is that the previous chance of predictor.

1) Applications of Naïve Bayes

- Real time Prediction: Naive Thomas Bayes is associate eager learning classifier and It's positive quick. Thus, it may well be used for creating predictions in real time.
- Multi category Prediction: This algorithmic rule is additionally standard for multi category prediction feature. Here we will predict the chance of multiple categories of target variable.
- Text classification/ Spam Filtering/ Sentiment Analysis: Naive Thomas Bayes classifiers largely utilized in text classification (due to raised end in multi class issues and independence rule) have higher success rate as compared to different algorithms. As a result, it's wide utilized in Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to spot positive and negative client sentiments)
- Recommendation System: Naive Thomas Bayes Classifier and cooperative Filtering along builds a Recommendation System that uses machine learning and data processing techniques to filter unseen info and predict whether or not a user would really like a given resource or not

B. Distributed Set-Cover Algorithm

The approximation formula and therefore the community primarily based algorithm square measure centralized and need international info of the network; i.e., pairwise expected diffusion time is required for the approximation formula and community structure is needed for the community primarily based formula. However, such info may not be offered or value an excessive amount of in some situations, like mobile social networks created from expedient node contacts. Furthermore, networks would possibly dynamically evolve over time and so the contact frequency between nodes (the edge weight) varies over time, which is able to have an effect on the accuracy for hard the pairwise expected diffusion time and detection the communities. Thus, during this section, we propose a distributed set-cover formula to deal with these issues, wherever every node collects up-to-date information and therefore the collected info is exploited to solve the diffusion decrease drawback. For an explicit fundamental measure   and a node u, there is a set of nodes to that u will diffuse info inside, spoken because the diffusion set of u. Suppose   is equal to the minimum diffusion time of the set of diffusion nodes, precisely,

$$\gamma = \min_{\substack{S \subseteq V \\ |S| \leq k}} \max_{v \in V} |(S, v)|,$$

The set of diffusion nodes S will be simply known by selecting the nodes, wherever the union of the diffusion sets for the chosen nodes is that the set of network nodes V. Though it's not possible to own the decreased diffusion time beforehand, this conjures up the look of the distributed set-cover formula.

The distributed set-cover formula includes 2 phases: discovering the diffusion set and characteristic the k- node set. For a given, that could be a system parameter, the first part leverages inquisitory messages to search out the diffusion set for every node during a distributed way; the second part iteratively selects the node to maximise the union of the diffusion sets for the chosen nodes.

Protocol 1: Discovering the diffusion set

Input: Δt *and* γ

Event: *Every* Δt.
Object: *All nodes:*

I. Generate a probing message including TTL and a set of traversed nodes, where TTL is set to γ and the set of traversed nodes is initialized to include the message generator.
II. Add the message into local message queue.

Event: *When two nodes contact with each other.*
Object: *Each of them*
Outgoing:

I. Randomly select a probing message, whose set of traversed nodes does not include the receiver, and send it out with the probability of information diffusion from sender to receiver.

Incoming:

I. When received a probing message, deduct the expected diffusion time from receiver to sender from TTL of the message.
II. If TTL ≥ 0, add the set of traversed nodes into the *up-to-date* diffusion set.
III. If TTL > 0, include itself into the set of traversed nodes and store the message into its message queue, otherwise, discard the message.

C. *Top-k Algorithm*

Given a social network composed of N individuals, Algorithm 1 shows the steps in finding the top-k influential individuals. The basic idea of this approximation algorithm is: first calculate the influence set of each individual, and then find the k most influential individuals.

Input: Graph of a social network; Parameter θ
Output: Top-k influential individuals
foreach *Individual i* **do**
  $f(0) = 0$; $f_i(0) = h_0$;
  Execute the heat diffusion process $f(t) = e^{\alpha t H} f(0)$;
  foreach *Individual j* **do**
    if $f_j(t) \geq \theta$ **then**
      Add *Individual j* into set $I_i(t)$
    end
  end
end
Sort $\{I_1(t), I_2(t), \ldots, I_N(t)\}$ by the set size;
Output top-k individuals;

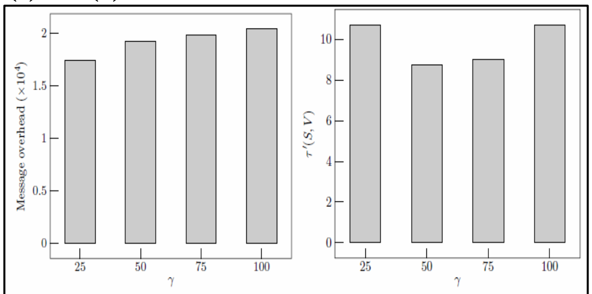
V. ANALYSIS

Naïve Bayes will exceed a lot of subtle classification strategies. Besides that it's conjointly exhibited high accuracy and speed once applied to giant info. Moreover, it's in no time for each learning and predicting. Its learning time is linear within the range of examples and its prediction time is freelance of the amount of examples [10].Naïve Thomas Bayes classifier is additionally quick, consistent, and simple to keep up and correct within the classification of attribute information. And from computation purpose of read, Naïve Thomas Bayes is a lot of economical each within the learning and within the classification task than call Tree.

Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) have completely different choices together with the selection of kernel operate for every. They're each sensitive to parameter improvement (i.e. completely different parameter choice will considerably modification their output). So, if you have got a result showing that NBC is playing higher than SVM. This is often solely true for the chosen parameters.

In general, if the idea of independence in NBC is glad by the variables of your dataset and also the degree of sophistication overlapping is little (i.e. potential linear call boundary), NBC would be expected to realize sensible. For a few datasets, with improvement mistreatment wrapper feature choice, for instance, NBC might defeat different classifiers. Not with standing it achieves a comparable performance, NBC are a lot of fascinating attributable to its high speed.

Performance of **Set-cover** with varying γ and Δt in terms of message overhead and expected diffusion time, where k = 50, Δt =2 days for (a) and (b), and k = 50, γ = 50 for (c) and (d).



(a) message overhead vs γ (b) expected diffusion time vs γ

Fig. 1:

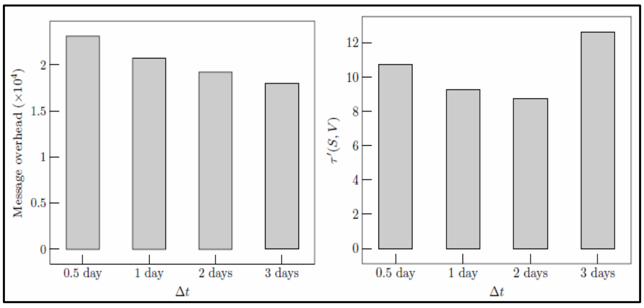


Fig. 2:

VI. CONCLUSION

A problem of distinctive intent in user text posts or social information. We feature out a deep analysis of the structure and content of posts showing intent and gift a feature extraction technique that captures them effectively. We have a tendency to then train a classifier victimisation these options to classify every post into Intent. We have a tendency to believe that our work will give vital insights to applications that specialize in exploiting free-text intentions from social media.

REFERENCES

[1] Hao Ma, Haixuan Yang, Michael R. Lyu and Irwin King, "Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates Selection," 2008.
[2] Ting Ning, Zhipeng Yang and Hongyi Wu, "Self-Interest-Driven incentives for ad dissemination in autonomous mobile social networks," 2013.
[3] W. Peng, F. Li, X. Zou, and J. Wu, "A privacy-preserving socialawar incentive system for word-of-mouth advertisement dissemination on smart mobile devices," in Proc. of IEEE SECON, 2012.
[4] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in Proc. of ACM WWW, 2012.
[5] D. M. Romero, B. Meeder, and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter," in Proc. of ACM WWW, 2011.
[6] D. Lopez-Pintado, "Diffusion in complex social networks," ´ Games and Economic Behavior, vol. 62, no. 2, pp. 573–590, 2008.
[7] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in Proc. of ACM SIGKDD, 2010.
[8] D. M. Romero, B. Meeder, and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter," in Proc. of ACM WWW, 2011.
[9] P. Domingos and M. Richardson, "Mining the network value of customers," in Proc. of ACM SIGKDD, 2001.
[10] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," in Proc. of ACM WWW, 2004.

