www.ijiarec.com

ISSN:2348-2079

Volume-8 Issue-3

International Journal of Intellectual Advancements and Research in Engineering Computations

Risk assessment in social networks based on user anomalous behaviours

M.Karthikeyan[1], D.Sangeetha[2]

[1]Assistant Professor, Department of Computer Science and Engineering, Bharathiyar Institute of Engineering for Women, Deviyakurichi, Attur- 636112

[2]PG Scholar Master of Engineering, Department of Computer Science and Engineering, Bharathiyar Institute of Engineering for Women, Deviyakurichi, Attur- 636112

ABSTRACT

Although the dramatic increase in OSN usage, there are still a lot of security and privacy concerns. In such a scenario, it would be very beneficial to have a mechanism able to assign a risk score to each OSN user. In this paper, we propose a risk assessment based on the idea that the more a user behavior diverges from what it can be considered as a 'normal behavior', the more it should be considered risky. In doing this, we have taken in into account that OSN population is really heterogeneous in observed behaviors. As such, it is not possible to define a unique standard behavioral model that fits all OSN users' behaviors. However, we expect that similar people tend to follow the similar rules with the results of similar behavioral models. For this reason, we propose a risk assessment organized into two phases: similar users are first grouped together, then, for each identified group, we build one or more models for normal behavior. The carried out experiments on a real Facebook dataset show that the proposed model outperforms a simplified behavioral-based risk assessment where behavioral models are built over the whole OSN population, without a group identification phase.

Keywords: Security and Privacy Concerns, Risk Assessment Organized, Normal Behavior, OSN Population

INTRODUCTION

Online Social Networks (OSNS) are today one of the most popular interactive medium to communicate, share, and disseminate a considerable amount of human life information. Daily and continuous communications imply the exchange of several types of content, including free text, image, audio, and video data. According to Facebook statistics1 average user creates 90 pieces of content each month, whereas more than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) Are shared each month. The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data. They are instrumental to provide an active support in complex and sophisticated tasks involved in OSN management [1-5] For instance access control or information filtering. Information filtering has been greatly explored for what concerns textual documents and, more recently, web content However, the aim of the majority of these proposals is mainly to provide users a classification mechanism to avoid they are overwhelmed by useless data. In osns, information filtering can also be used for a different, more sensitive, purpose. This is due to the fact That in osns there is the possibility of posting or commenting other posts on particular public/private areas, called in general walls. Information filtering can therefore be used to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted messages [6-10].

Author for correspondence:
Department of Computer Science and Engineering, Bharathiyar Institute of Engineering for Women, Deviyakurichi, Attur- 636112

RELATED WORK

The main contribution of this paper is the design of a system providing customizable content-based message filtering for OSNs, based on ML techniques. As we have pointed out in the introduction, to the best of our knowledge, we are the first proposing such kind of application for OSNs. However, our work has relationships both with the state of the art in content-based filtering, as well as with the field of policy-based personalization for OSNs and, more in general, web contents. Therefore, in what follows, we survey the literature in both these fields.

LITERATURE SURVEY

Cuneyt GurcanAkcora, Barbara Carminati, and Elena Ferrari

Several efforts have been made for more privacy aware Online Social Networks (OSNs) to protect personal data against various privacy threats. However, despite the relevance of these proposals, we believe there is still the lack of a conceptual model on top of which privacy tools have to be designed. Central to this model should be the concept of risk. Therefore, in this paper, we propose a risk measure for OSNs. The aim is to associate a risk level with social network users in order to provide other users with a measure of how much it might be risky, in terms of disclosure of private information, to have interactions with them. We compute risk levels based on similarity and benefit measures, by also taking into account the user risk attitudes. In particular, we adopt an active learning approach for risk estimation,

Todd K Moon. The expectation-maximization algorithm

A common task in signal processing is the estimation of the parameters of a probability distribution function. Perhaps the most frequently encountered estimation problem is the estimation of the mean of a signal in noise. In many parameter estimation problems the situation is more complicated because direct access to the data necessary to estimate the parameters is impossible, or some of the data are missing. Such difficulties arise when an outcome is a result of an accumulation of simpler outcomes, or when outcomes are clumped together, for example, in a binning or histogram operation. There may also be data dropouts or clustering in such a way that the number of underlying data points is unknown (censoring and/or truncation).

PROPOSED SYSTEM

The aim of the present work is therefore to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls. We exploit Machine Learning (ML) text categorization techniques to automatically assign with each short text message a set of categories based on its content.

The major efforts in building a robust short text classifier are concentrated in the extraction and selection of a set of characterizing and discriminate features. The solutions investigated in this project are an extension of those adopted in a previous work by us from whom we inherit the learning model and the elicitation procedure for generating pre-classified data.

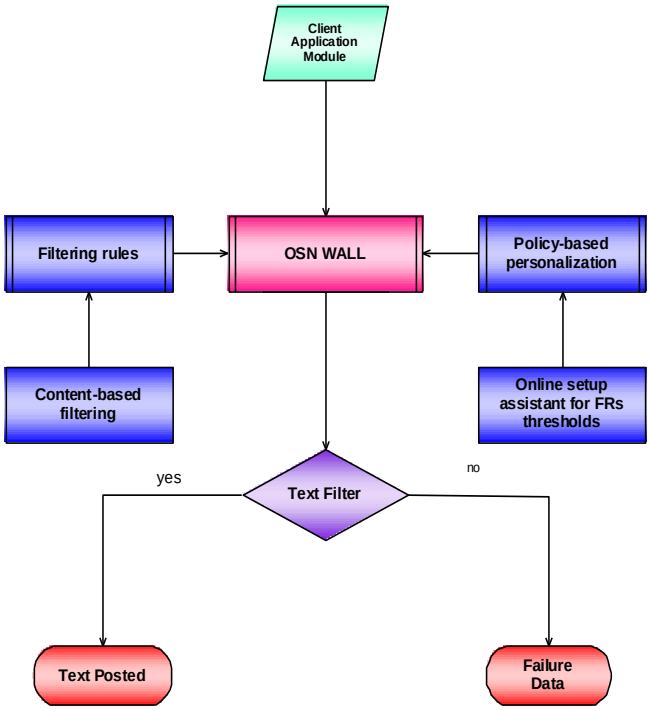
The original set of features, derived from endogenous properties of short texts, is enlarged here including exogenous knowledge related to the context from which the messages originate. As far as the learning model is concerned, we confirm in the current project the use of neural learning which is today recognized as one of the most efficient solutions in text classification. We base the overall short text classification strategy on Radial Basis Function Networks (RBFN) for their proven capabilities in acting as soft classifiers, in managing noisy data and intrinsically vague classes. Moreover, the speed 2 in performing the learning phase creates the premise for an adequate use in OSN domains, as well as facilitates the experimental evaluation tasks.

Advantages

- Experimentally evaluate an automated system
- Automatically assign with each short text message
- Discriminate features.
- Adopted in a previous work
- Speed 2 in performing

Copyrights © International Journal of Intellectual Advancements and Research in Engineering Computations, www.ijiarec.com

System Model



MODULES

- Content based filtering
- Filtering rules (FRs)
- Short text classification
- Blacklists management

Content based filtering

The application of content-based filtering on messages posted on OSN user walls poses additional challenges given the short length of these messages other than the wide range of topics that can be discussed. Short text classification has received up to now few attentions in the scientific community. Recent work highlights difficulties in defining robust features, essentially due to the fact that the description of the short text is concise, with many misspellings, nonstandard terms, and noise.

Filtering Rules (FRS)

We consider three main issues that, in our opinion, should affect a message filtering decision. First of all, in osns like in everyday life, the same message may have different meanings and relevance based on who writes it. As a consequence, FRs should allow users to state constraints on message creators. Creators on which a FR applies can be selected on the basis of several different criteria; one of the most relevant is by imposing conditions on their profile's attributes.

Short text classification

This technique used for text classification work well on data sets with large documents such as newswires corpora but suffer when the documents in the corpus are short. In this context, critical aspects are the definition of a set of characterizing and discriminant features allowing the representation of underlying concepts and the collection of a complete and consistent set of supervised examples.

Copyrights © International Journal of Intellectual Advancements and Research in Engineering Computations, www.ijiarec.com

Blacklists Management

Our system is a BL mechanism to avoid messages from undesired creators, independent from their contents. BLs are directly managed by the system, which should be able to determine who are the users to be inserted in the BL and decide when users retention in the BL is finished. To enhance flexibility, such information are given to the system through a set of rules, hereafter called BL rules. Such rules are not defined by the SNM, therefore they are not meant as general high level directives to be applied to the whole community.

METHODOLOGY

Content-Based Filtering Information filtering systems are designed to classify a stream of dynamically generated information dispatched asynchronously by an information producer and present to the user those information that are likely to satisfy his/her requirements. In content-based filtering, each user is assumed to operate independently. As a result, a content-based filtering system selects information items based on the correlation between the content of the items and the user preferences as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences. While electronic mail was the original domain of early work on information filtering, subsequent papers have addressed diversified domains including newswire articles, Internet "news" articles, and broader network resources. Documents processed in content-based filtering are mostly textual in nature and this makes content-based filtering close to text classification. The activity of filtering can be modeled, in fact, as a case of single label, binary classification, partitioning incoming documents into relevant and no relevant categories.

More complex filtering systems include multilabel text categorization automatically labeling messages into partial thematic categories. Content-based filtering is mainly based on the use of the ML paradigm according to which a classifier is automatically induced by learning from a set of pre classified examples. A remarkable variety of related work hasrecently appeared which differ for the adopted feature extraction methods, model learning, and collection ofsamples. The feature extraction procedure maps text into a compact representation of its content and is uniformly applied to training and generalization phases. Several experiments prove that Bag-of-Words(BoW) approaches yield good performance and prevail in general over more sophisticated text representation that may have superior semantics but lower statistical quality.

As far as the learning model is concerned, there are a number of major approaches in content-based filtering and text classification in general showing mutual advantages and disadvantages in function of application dependent issues. In a detailed comparison analysis has been conducted confirming superiority of Boosting-based classifiers, Neural Networks, and Support Vector Machines over other popular methods, such as Rocchio and Naı ¨ve Bayesian. However, it is worth to note that most of the work related to text filtering by ML has been applied for long-form text and the assessed performance of the text classification methods strictly depends on the nature of textual documents.

CONCLUSION

In this project, we have presented a system to filter undesired messages from OSN walls. The system exploits a ML soft classifier to enforce customizable content-dependent FRs. The development of a GUI and a set of related tools to make easier BL and FR specification is also a direction we plan to investigate, since usability is a key requirement for such kind of applications. In particular, we aim at investigating a tool able to automatically recommend trust values for those contacts user does not personally known. We do believe that such a tool should suggest trust value based on users actions, behaviors, and reputation in OSN, which might imply to enhance OSN with audit mechanisms. However, the design of these audit-based tools is complicated by several issues, like the implications an audit system might have on user's privacy and/or the limitations on what it is possible to audit in current OSNs.

Copyrights © International Journal of Intellectual Advancements and Research in Engineering Computations, www.ijiarec.com

REFERENCES

[1]. CuneytGurcanAkcora, Barbara Carminati, and Elena Ferrari. Privacy in social networks how risky is your social graph? In Data Engineering (ICDE), IEEE 28th International Conference on, IEEE, 2012, 9–19.

[2]. Christa SC Asterhan and Tammy Eisenmann. Online and faceto-face discussions in the classroom: A study on the experiences of'active'and'silent'students. In Proceedings of the 9th international conference on Computer supported collaborative learning-. International Society of the Learning Sciences, 1, 2009, 132–136.

[3]. Leyla Bilge, Thorsten Strufe, DavideBalzarotti, and EnginKirda. All your contacts are belonging to us automated identity theft attacks on social networks. In Proceedings of the 18th international conference on World Wide Web, ACM, 2009, 551–560.

[4]. YazanBoshmaf, Konstantin Beznosov, and MateiRipeanu. Graphbasedsybil detection in social and information systems. In Advances in Social Networks Analysis and Mining (ASONAM), IEEE/ACM International Conference on, IEEE, 2013, 466–473.

[5]. YazanBoshmaf, DionysiosLogothetis, GeorgosSiganos, Jorge Ler´ıa, Jose Lorenzo, MateiRipeanu, and Konstantin Beznosov. ´Integro: Leveraging victim prediction for robust fake account detection in osns. In Proc. of NDSS, 2015.

[6]. YazanBoshmaf, IldarMuslukhov, Konstantin Beznosov, and MateiRipeanu. The socialbot network- when bots socialize for fame and money. In Proceedings of the 27th Annual Computer Security Applications Conference, ACM, 2011, 93–102.

[7]. Paul S Bradley, Usama Fayyad, and Cory Reina. Scaling em (expectation-maximization) clustering to large databases. Technical report, Technical Report MSR-TR-98-35, Microsoft Research Redmond, 1998.

[8]. Qiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro. Aiding the detection of fake accounts in large scale social online services. In NSDI, 2012, 197–210.

[9]. George Danezis and Prateek Mittal. Sybilinfer- detecting sybil nodes using social networks. In NDSS, 2009.

[10]. Vacha Dave, SaikatGuha, and Yin Zhang. Measuring and fingerprinting click-spam in ad networks. In Proceedings of the ACM SIGCOMM conference on Applications, technologies, architectures, and protocols for computer communication, ACM, 2012, 175–186.

Copyrights © International Journal of Intellectual Advancements and Research in Engineering Computations, www.ijiarec.com