

Deep learning approaches for video-based anomalous activity detection

Karishma Pawar¹ · Vahida Attar¹

Received: 16 August 2017 / Revised: 16 April 2018 / Accepted: 26 April 2018 /
Published online: 3 May 2018
© Springer Science+Business Media, I.L.C., part of Springer Nature 2018

Abstract The pervasive use of cameras at indoor and outdoor premises on account of recording the activities has resulted into deluge of long video data. Such surveillance videos are characterized by single or multiple entities (persons, objects) performing sequential/concurrent activities. It is often interesting to detect suspicious behavior of such entities in an automated manner without any intervention of human personnel, and to this end, anomalous activity detection from surveillance videos is an important research domain in Computer Vision. Detecting the anomalous activities from videos is very challenging due to equivocal nature of anomalies, context at which events took place, lack of ample size of anomalous ground truth training data and also other factors associated with variation in environment conditions, illumination conditions and working status of capturing cameras. Though automated visual surveillance is one of the highly sought-after research domains, use of deep learning techniques for anomalous activity detection is still in nascent stage. Deep learning models like convolution neural networks, auto-encoders, Long Short Term Memory network models have achieved remarkable performance on different domains like image classification, object detection, speech processing, and expediting towards achieving excellence in anomaly detection tasks. This paper aims at studying and analyzing deep learning techniques for video-based anomalous activity detection. As outcome of the study, the graphical taxonomy has been put forth based on kinds of anomalies,

This article belongs to the Topical Collection: *Special Issue on Deep vs. Shallow: Learning for Emerging Web-scale Data Computing and Applications*
Guest Editors: Jingkuan Song, Shuqiang Jiang, Elisa Ricci, and Zi Huang.

✉ Karishma Pawar
kvppawar@gmail.com

Vahida Attar
vahida.comp@coep.ac.in

¹ Department of Computer Engineering & IT, College of Engineering Pune (COEP), Pune, India

Comparative study of machine learning algorithms for anomaly detection in Cloud infrastructure

¹Mr. Satish Kumbhar, ²Nikita Paranjape, ³Rasika Bhawe, ⁴Akshay Lahoti

¹Assistant Professor, Department of Computer Engineering, College of Engineering, Pune, Maharashtra 411005

²B.Tech Student, Department of Computer Engineering, College of Engineering, Pune, Maharashtra 411005

³B.Tech Student, Department of Computer Engineering, College of Engineering, Pune, Maharashtra 411005

⁴B.Tech Student, Department of Computer Engineering, College of Engineering, Pune, Maharashtra 411005

Email: ¹ssk.comp@coep.ac.in, ⁴lahotias14.comp@coep.ac.in

Contact: ¹+919860574798, ²+918390023697, ³+919607210054, ⁴+919987985460

Abstract—Cloud is one of the emerging technologies in the field of computer science and is extremely popular because of its use of elastic resources to provide optimized, cost-effective and on-demand services. As technology started to grow in scale and complexity, the need for automated anomaly detection and monitoring system has become important. Inappropriate exploitation of Cloud resources can often lead to faults like crashing of VMs, decreased efficiency of cloud system etc. thereby leading to violations of the Service Level Agreement (SLA). These faults are often preceded by anomalies in the behavior of the VMs. Hence, the anomalies can be used as indicators of faults which potentially violate the SLAs. We have created a system that will monitor the VMs, detect anomalies and warn the system administrator before any problem escalates. We present in this paper a comparative study of various machine learning algorithms used for detecting anomalies in cloud

Keywords- Anomaly detection; Machine Learning; Cloud infrastructure

I. INTRODUCTION

Anomaly detection is the process of finding the patterns in a dataset whose behavior is abnormal or unexpected. Such unexpected behavior is also termed as an anomaly or an outlier [9]. The anomalies cannot always be categorized as an attack but it can be a surprising, previously unknown behavior which can escalate into a bigger problem. This can potentially violate the SLA [1].

Machine Learning automatically trains a model from historic data without being explicitly programmed and improves its accuracy with experience. This model predicts values of data in near future. In anomaly detection, the machine learning algorithm is supplied initially with the data. This data is a mixture of normal and anomalous data that has been labelled. These algorithms create models that are then used to predict whether the cloud system is in anomalous state or not [8].

This paper is organized as follows. Section II explains the setting up of the infrastructure. Section III describes the methods used to simulate anomalies on the VMs. Section IV explains how the data was collected and preprocessed. Section V describes the machine learning algorithms used. Section VI will put forth the results we got on the testing data and lastly Section VII describes the future scope of this research.

II. SETTING UP INFRASTRUCTURE

A. Cloud

OpenStack is a free and open-source software platform for cloud computing, mostly deployed as infrastructure-as-a-service, whereby virtual servers and other resources are made available to customers [2]. It can be installed as mentioned in

[3]. For testing, we installed OpenStack on a machine which has 4 processors, 32 GB of RAM.

B. Metrics monitoring tool

We have installed Zabbix, enterprise open monitoring software for networks and applications, to monitor our virtual machines and collect data. This entails installing a Zabbix server on one virtual machine, and a Zabbix agent on all other virtual machines that we wish to monitor. The agent monitors the virtual machine it is installed on and periodically sends data to the server (data is sent per minute). The server machine needs to be configured to accept data from all the agents and agents need to be configured to send data to the correct server [4].

III. SIMULATING ANOMALIES

In order to find out how the VM behaves when an anomaly occurs, we simulated some anomalies. There are several tools to simulate such anomalies. These tools can be used as commands on the terminal and do not require a user interface. We used the stress tool [5]. By varying different command line parameters passed to these tools, we were able to customize the severity of each of the following anomalies [6].

A. CPU Utilization

It creates number of process specified by the user and consumes the CPU resource for a specified amount of time.

Syntax: stress -c <no_of_processes> -t <time_to_run>

Example: stress -c 4 -t 180s

B. Available Memory

Available memory is the amount of RAM that is free.

Home > Other

nQuery A Natural Language Statement to SQL Query Generator

 7  14  Download (0)

Increase Efficiency Using Automation and /
—Download Our Free eBook to Get Startec

nQuery - A Natural Language Statement to SQL Query Generator

Nandan Sukthankar, Sanket Maharnawar, Pranay Deshmukh
Yashodhara Haribhakta, Vibhavari Kamble

Department of Computer Engineering and Information Technology
College of Engineering Pune

Wellesley Road, Shivajinagar, Pune, Maharashtra, India

<http://www.coep.org.in>

{nandans16, sanketmaharnawar, pranay.s.deshmukh}@gmail.com
{ybl.comp, vvk.comp}@coep.ac.in

Abstract

In this research, an intelligent system is designed between the user and the database system which accepts natural language input and then converts it into an SQL query. The research focuses on incorporating complex queries along with simple queries irrespective of the database. The system accommodates aggregate functions, multiple conditions in WHERE clause, advanced clauses like ORDER BY, GROUP BY and HAVING. The system handles single sentence natural language inputs, which are with respect to selected database. The research currently concentrates on MySQL database system. The natural language statement goes through various stages of Natural Language Processing like morphological, lexical, syntactic and semantic analysis resulting in SQL query formation.

1 Introduction

Today, virtually every relational database management system (RDBMS) uses Structured Query Language (SQL) for querying and maintaining the database. Users accessing relational databases need to learn SQL and build queries in the right syntax for retrieving the data. It becomes a big hurdle for all those who are not technically knowl-

is regarded as a promising and important endeavor in the field of computer research.

nQuery will translate natural language queries into SQL before retrieving data from database. It will deal with single sentence inputs given by the user using a particular database. The system mainly focuses on data retrieval but also provides the facility to convert DML natural language statements to SQL. However, the system will output queries which can be used for querying the MySQL database system only. The aim of the system is to reduce the complexity of database querying. The approach our system uses, extracts certain keywords from the natural language statement and goes through various steps of Natural Language Processing. This system focuses on table mapping, attribute mapping and clause tagging to generate the resultant query.

2 Related Work

Over the years, certain systems which focus only on a particular database have been built to serve a particular purpose. (Woods, 1973) developed a system called LUNAR, that answered questions about rock samples brought back from the moon. LIFER/LADDER designed by (Hendrix, 1978) was designed as a natural language interface to

“Matsar” Sentiment Analysis

Yashodhara V. Haribhakta*, Sunil Barhate

Department of Computer Engineering and IT, College of Engineering, Pune, Maharashtra, India

ABSTRACT

Matsar is Marathi language word and the meaning is *envy, jealousy, negativity* and *distrust*. The following sentence “He got this project because of luck rather than his work” conveys *Matsar* sentiment. In this competitive world, it increases negativity and jealousy among people due to the success of other people or status of other people etc. Generally it seems between friends, co-workers, businesses etc. It will be helpful in schools, colleges, organizations to get to know the relation between two individuals. In this work, we have proposed a pattern based approach to detect whether the text contains *Matsar* Sentiment or not. The proposed approach gives accuracy of 99% using KNN algorithm, 98% using DT algorithm, 97% using SVM algorithm and 94% using NB algorithm.

Keywords: *Matsar*, Sentiment Analysis, Patterns, Pos-Tagging

I. INTRODUCTION

Sentiment Analysis is a process of extracting information from the text. It is also known as opinion mining. The aim of Sentiment Analysis is to detect the sentiment, opinion, attitude from the text. Lot of research has been done in analyzing sentiments in the text, such as, subjectivity detection, polarity detection, aspect based sentiment summarization, sarcasm detection, spam detection, emotion detection, mood detection, attitude detection, wish detection, sentiment summarization, etc.

Subjectivity Detection implies whether the text contains any sentiment or not, that is, whether text sentiment is subjective or objective. It is unlike of detecting polarity of text. Some remarkable research works on subjectivity detection are [1] [2]. In Sentiment Prediction, sentiments are detected from the text, i.e. positive, negative or neutral. Sentiment detection is done at document level, sentence level and phrase level. Some remarkable research works of Sentiment Prediction are [3][4]. Aspect Based Sentiment Summarization is bit ahead of sentiment analysis. The summarizer creates opinion summary about feature or aspect of products. For example, a Phone has different features like screen, sound, design, etc., then the objective is to give an outline as rating scores on each of these features. Some remarkable research works on this domain are [5] [6] [7]. Text Summarization for Opinion helps to create summary format for text summaries. Some remarkable research

works on this domain are [8] [9]. Some reviews may be more helpful compared to others. Instead of displaying user reviews in sequential order, sorting reviews by its helpfulness would improve user productivity. Predicting Helpfulness aims at automatically predicting the helpfulness of user reviews instead of just relying on users rating. Some remarkable research works on this domain are [10] [11].

In this competitive world, it is observed that, usually, individuals do not appreciate success of others. They carry jealous feelings for their subordinate success. The text “Ram got this project because of luck rather than his hard work” is a *Matsar* text, the speaker is not giving importance to Ram’s hard work and trying to project that Ram got the project because of luck and not due to his hard work.

The aim of the proposed work is to detect *Matsar* sentiment in English language text. The approach used in this work is extraction of pattern based features defined as “Ordered sequence of content words(CW) and General Expression Words(GEW)”. The CW are *Matsar* sentiment contextual words whereas GEW can be replaced by general expressions. This work focuses on pattern creation, determining maximum length of pattern and their effects, finding optimal value of a parameter which is used to optimize the score of partial matching, and finding optimal value of maximum length of pattern to detect *Matsar* sentiment in text. Since no dataset is available for *Matsar* sentiment, manually the data has been generated for detecting *Matsar* and *non-matsar* text. So a dataset comprising of

Parts of Speech Tagger for Pali Language

Yashodhara Haribhakta*¹, Laxmi Nadageri²

*^{1,2}Department of Computer Engineering and IT, College of Engineering, Pune, Maharashtra, India

ABSTRACT

Parts of Speech tagging is the process of labelling the words in the text with their appropriate labels. The labels assigned are noun, verb, adjective, adverb, pronoun... etc. For performing natural language processing, Parts of Speech tagging is an essential requirement. It is very simple statistical model for many Natural Language Processing applications. In this paper, we propose a parts of speech tagger for Pali language. Pali though considered as extinct, has very rich literature comprising works on Logic, History, Medicine, Pharmacology etc. It is an Indo-Aryan language. The general approach used for development of Pali tagger is a Rule based approach. It also presents the tagset used for Pali language. The paper shows the performance of proposed Rule based tagger for a dataset up to 300 sentences / 1000 words. The learning algorithms Support Vector Machine and Decision Tree have been used for measuring the performance on Pali tagged corpus.

Keywords : Parts of Speech tagging, tagger, Rule based tagger, Pali language.

I. INTRODUCTION

Pali is an Indo-Aryan language which died out as a literary language in fourteenth century but survived somewhere as sacred language which is used mainly for religious purpose. The original teachings of Gautama Buddha were written into Pali language. The initial Pali text were found written on Palm leaves. Pali Text Society founded in 1881 took initiatives to scan such palm leaves and store Pali text into electronic form. Digitalization of such historical language is important to increase usage of its literature. The work presented here is an initiative step to create an annotated language resources to use in an applications of Natural Language Processing (NLP) like language learning, language teaching, text classification, question answering etc. The first attempt of creating an annotated language resource is Parts of speech (POS) tagger. POS tagging is a process of assigning a category to a word from a defined set of categories belonging to a language.

The main objective of proposed work is to build a POS tagger for Pali language. The approach used is Rule based. The Rule based tagger uses Pali grammar rules.

Words of Pali text are labelled by looking at the affixes attached to each word. In case there is an ambiguity while assigning label, it is resolved looking at label of the context words. The Pali tagset comprises of 16 tags including unknown tag UN. Using this tagset the data from www.tipitaka.org has been used for manual labelling. So, a manually labelled dataset of 300 sentences has been generated.

The second objective is an evaluation of the performance of machine learning algorithms on labelled Pali dataset. Two popular learning methods are used: Support Vector Machine (SVM) and Decision Tree (DT). LibLinear and J48 are the implementations of SVM and DT respectively. They are used by optimizing their parameters to achieve the maximum accuracy. All the experiments done on learning algorithms and their results are discussed in this paper.

In section 2 shows related work done on POS taggers for Indian languages. In section 3 Pali language is introduced. Pali dataset is described in section 4. General introduction of POS taggers is given in section 5. Section 6 describes POS tagger for Pali language and results are discussed in section 7. Machine learning

Categorization of News Articles using Sentiment Analysis

Yashodhara Haribhakta^{*1}, Kiran Shriniwas Doddi²

^{*1,2} Department of Computer Engineering and Information Technology, College of Engineering Pune, Pune, Maharashtra, India

ABSTRACT

The advent use of new online social media such as articles, blogs, message boards, news channels, and in general web content has dramatically changed the way people look at various things around them. Today, it's a daily practice for many people to read news online. People's perspective tends to undergo a change as per the news content they read. The majority of the content that we read today is on the negative aspects of various things e.g. corruption, rapes, thefts etc. Reading such news is spreading negativity amongst the people. Positive news seems to have gone into a hiding. The positivity surrounding the good news has been drastically reduced by the number of bad news. This has made a great practical use of Sentiment Analysis and there has been more innovation in this area in recent era. It traditionally emphasizes on classification of text document into positive and negative categories. The objective of this paper is to provide a platform for serving good news and create a positive environment. This is achieved by finding the sentiments of the news articles and filtering out the negative articles which carry negative sentiments. This would enable us to focus only on the good news which will help spread positivity around society and would allow people to think positively. To achieve our objective, we have proposed an algorithm for classification of News articles. This includes data aggregator tool and processing engine at the server side as a Sentiment classifier and a platform for user where positive news being served to read.

Keywords: Document classification, Sentiment Analysis, Support vector machine (SVM)

I. INTRODUCTION

With the arrival of internet, there has been a radical change in the social life, routine and decisions of common people. Today, it's everyday activity and regular practice for each person to read news online and watch advertisements regarding a movie, a product or a book before actually placing money into it. As it has changed their lifestyle, it also has impact on the social life of an individual. The exposure to new online social media such as articles, blogs, message boards, news channels such as Web content is influencing their social life and the way people look at various things around them. People's perspective tends to undergo a change as per the content they read.

The social media has now occupied the major space on the Web. The new user-centric Web hosts a huge amount of data every day. Users are not only consuming the web, but they are also a part of web and co-creators of content on web. The user is now

contributing to social media ranging from articles, blog posts, news, tweets, reviews, photo/video upload, etc. This is creating a large amount of the data on the Web as unstructured text.

The majority of the content that we read today is on the negative aspects of various things e.g. corruption, rapes, thefts etc. Reading such news is spreading negativity amongst the people. Positive news has been dominated and getting less attention. The positivity surrounding the good news has been drastically reduced by the number of bad news.

The objective of this project is to provide a platform for serving good news and create a positive environment. The new challenging task here is to analyze large volume of unstructured text to be more specific news articles and devise suitable algorithms to understand the opinion of others and find positive and negative aspect of it. This would enable us to focus only on the good news which will help spread positivity around and would allow people to think positively.

Journal of Computer Based Parallel Programming (e-ISSN: 2582-2179)

HOME ABOUT LOGIN REGISTER CATEGORIES
SEARCH CURRENT ARCHIVES ANNOUNCEMENTS CONTACT

OPEN JOURNAL
SYSTEMS

US

[Journal Help](#)

Home > Vol 3, No 1 (2018) > Mahadule

Open Access Subscription Access

USER

Username

Password

Remember me

SUBSCRIPTION

Login to verify
subscription

NOTIFICATIONS

- [View](#)
- [Subscribe](#)

JOURNAL
CONTENT

Search

Search Scope

All

Browse

- [By Issue](#)
- [By Author](#)
- [By Title](#)
- [Other Journals](#)
- [Categories](#)

FONT SIZE

INFORMATION

- [For Readers](#)
- [For Authors](#)
- [For Librarians](#)

Host Based Intrusion Detection System Using Decision Tree and Naïve Bayes Algorithms

Mr Abhijeet Mahadule, Mr Atul Rathod, Mr Shubham Singh, Mr Rahul B. Adhao

Abstract

A network intrusion is any unauthorized access to a computer network. For detecting a network intrusion, the defenders should have a clear understanding of how attacks work. In a network environment, intrusions possess a major security issue which can be an unauthorized activity on a computer network which is generally difficult to detect. Through this project we aim to monitor computer network for malicious activity using Intrusion Detection System (IDS). In this work, we aim to use Supervised Machine Learning Algorithms such as Naive Bayes Classifier and Decision Tree Classifiers, and compare their output and efficiency. The machine learning algorithms are used on a labelled dataset, which classifies the connections as good or bad. As a result, the accuracy of the classification result has to be maximized by maintaining low false-negative and low false-positive rates.

Full Text:

[PDF](#)

References

Harish C., Palash Chaturvedi, Amit K. Saxena, "A Systematic Literature Survey in IDS", International Journal on Recent and Innovation Trends in Computing and Communication, Vol 5, Issue:6, 2017

Sravan K. Jonnalagadda, Ravi P. Reddy, "A Literature Survey and Comprehensive Study of Intrusion Detection", International Journal of Computer Applications, Volume 81, 2013

Mrutyunjaya Panda, Manas R. Patra, "Network Intrusion detection using Naive Bayes", International Journal of Computer Science and Network Security, Vol 7, 2007

Gavin wolf, Dr. Taghi, M. Khoshgoftaar, "Using Machine Learning for Network Intrusion Detection: The Need for Representative Data", Florida Atlantic University Boca Raton, Florida USA, May 2016

Suseela T., Q.A. Zhu, "Hierarchical Kohonen net for anomaly detection in network security J. Huff", IEEE Transactions on System, Man and Cybernetics, Part(B), Cybernetics (Volume: 35, Issue:2, April 2005)

Ajith Abraham, Sandhya Peddabachigari, C. Grosan, Johnson Thomas, "Modelling intrusion detection system using hybrid intelligent systems", Journal of Network and Computer Applications, 2007

Iwan Syarif, Adam Prugel-Bennett, Gary Wills, "Data Mining Approaches for Network Intrusion Detection: from Dimensionality Reduction to Misuse and Anomaly Detection", Journal of Information Technology, Vol 3, 2012



Corpus ID: 212493523

Share This Paper

Survey on Municipal Waste Collection Management in Smart City

5 Citations

Background Citations 2

1Student, Dept. of Computer engineering, College of Engineering Pune, Maharashtra, India 2Professor, Dept. of Computer engineering, College of Engineering Pune, Maharashtra, India

View All

Abstract At present waste management is a major concern in the metropolitan cities of the developing and developed countries. As the population is growing, the garbage is also increasing day by day. Garbage management is becoming a global problem. Due to the lack of care and attention by the authorities the garbage bins are mostly seem to be overflowing. It has to be taken into care by corresponding authorities and should think what method can be followed to overcome this. This huge unmanaged accumulation of garbage is polluting the environment, spoiling the beauty of the area and also leading to the health hazard. To overcome this situation an efficient smart municipal waste management system has to be developed. In this era of Internet, Internet of Things (IOT) can be used effectively to manage this waste as many effective methods can be found out easily. This is the survey paper which involves the various ideas to solve this problem using some algorithms that can be easily implemented. Collapse

No Paper Link Available

Save

Alert

Abstract

5 Citations

16 References

Related Papers

5 Citations

Search Date Range Citation Type Year PDF Publication Type Author More Filters More Filters Filters Sort by Releva

SMART BOT WITH AUTOMATIC GARBAGE COLLECTING SYSTEM

Mudite Koushal Yadav, Rohith Malvada, Saou Rahaf Goudi Engineering 2020

1Department of Mechanical Engineering, St Martins Engineering College, Telangana, 2Department of Mechanical Engineering, St Martins Engineering College, Telan 3Department of Mechanical... Expand

PDF Save Alert

Sustainable Waste Management Model

Shrey Datta, Shrey Datta, I. Agarwal, Saou Rahaf Goudi, Indu Divyale, Dhakshinam Business 2018 Eleventh International Conference on Contemporary Computing (IC3) 2018

TLDR The idea is to divide the management system into four modules-division of the city into zones to reduce the problem size, assignment of scrape yards to eac zone, optimization of waste collection route and setting an efficient budget for investing in waste processing machines, thereby designing an overall waste manage model. Expand

5 Save Alert

AN ANALYSIS OF GARBAGE MECHANISM FOR SMART CITIES

P. Manikandan S. Sathyanarayanan A. Aravind M. Durga L. Keerthana Engineering 2019

To make smart cities, safer and efficient, wireless communications can play an important role. While implementing this proper waste disposal management, it will a the spreading of diseases. In... Expand

PDF Save Alert

Waste Management by a Robot- A Smart and Autonomous Technique