International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Social Media Comment Extractor using NLP

Akhil S[1]*, Prof. Neelima P*[2]*, Dr.T.Mahalekshmi*[3]

[1]Final year Student, Sree Narayana Instiute of Technology, Kollam, Kerala,India
[2]Assistant Professor, Sree Narayana Instiute of Technology, Kollam, Kerala,India
[3]Principal, Sree Narayana Instiute of Technology, Kollam, Kerala,India

ABSTRACT

This undertaking centers around the issue of short message synopsis on the remark stream of a particular message from informal organization administrations (SNS). Because of the great notoriety of SNS, the amount of remarks might increment at a high rate just later a social message is distributed. Spurred by the way that clients might want to get a short comprehension of a remark stream without perusing the entire remark list, we endeavor to bunch remarks with comparable substance together and create a compact assessment synopsis for this message. Since particular clients will demand the synopsis without warning, existing grouping techniques can't be straightforwardly applied and can't meet the constant need of this application. In this venture, we model a clever steady bunching issue for input stream rundown on SNS. In addition, we propose a calculation that can gradually refresh bunching results with most recent approaching remarks continuously. Besides, we plan an initially perception interface to help clients effectively and quickly get an outline synopsis. From broad test results and a genuine case exhibition, we check that the has the benefits of high effectiveness, high versatility, and better taking care of exceptions, which legitimizes the practicability of this strategy on the objective issue.

Keywords—***Short text; text distance; formal distance; semantic distance; network comment clustering; golden section method***

Introduction

With the prevalence of portable internet providers, the Internet clients arrived at 668 million [1] and the extent of clients in the utilization of versatile Internet clients represented over 88.9% [1]in China by June 2015, the Internet clients will quite often utilize more succinct and more customized short text express close to home

sees, so the short text online remarks brought new provokes and freedoms to Chinese data processing.The short text online audits that source in Microblog and WeChat and so forth typically start to an occasion or point,fast spread and have a wide effect, mirroring the general population's disposition to public occasions, the articulation content with solid subjectivity.It not just communicates the commentator's perspective, yet additionally influences the other member's perspective. In this paper, in light of the streamlining of short text network survey grouping investigation, in light of semantic and structure two parts of online audit of the client to improve the examination, it can get a handle on

individuals' perspectives and positions on different hotly debated issues, for the nation, undertakings and society are of incredible importance.

In the past research of the online comment clustering, the special character of short text processing is not considered and the successful method of dealing with the traditional long text is used to deal with the short text. Such as based on the method of Bias and Naive Bayesian [2] and the use of all text sets or context relations will be short text as a whole to reduce the impact of text length [3-5]. In the paper [2] model based on n gram is used in the information filtering of Twitter, and the length of the input Twitter is used as a variable, and the effect of the term is balanced by using the dynamic average method. The limitation of Bag-of-words in short text classification is analyzed by data, and the feature extraction method based on user information context is proposed in this paper [3]. A corpus based on Set - the set of knowledge short text semantic approximation method of measuring the degree, give full consideration to the characteristics of short text in the large scale corpus, to achieve good performance in recognition reporting questions presented in literature [4]. The paper [5] uses LSA+ICA as a whole to extract the semantics of short text, thus avoiding the problem of short text length. These methods do not take into account the special nature of the short text processing, the processing effect and the real requirements of the short text has a huge gap. As the short text feature and the user processing of the short text feature in the research of network reviews, the method of text form and semantic similarity is more important. Based on text similarity method, through compare the text contains words and word order to calculate the passage between the semantic similarities of the paper [6]. Based on the method of semantic similarity of text in the paper [7-9], ontology, is used to calculate the semantic similarity of short text. Method based on semantic similarity of text and text semantic similarity based on the method of the paper [10]. The paper [9] proposed a method of text similarity calculation based on LDA topic model and evaluated the

clustering effect of the text similarity matrix. The paper [10] presents an improved semantic distance measure, which is a comprehensive representation of formal distance and unit semantic distance. In short, the research work considered the combination between the two, or simply adding two methods, without considering its intrinsic link and weight factor. On the basis of the above research, the paper introduces the method of weight factor and golden section optimization to optimize the clustering based on the short text.

we investigate the issue of steady short text synopsis on remark streams from interpersonal organization administrations. We model this issue as an steady bunching task and propose the IncreSTS (representing Incremental Short Text Summarization) algorithm to find the top-k groups including unique gatherings of conclusions towards one social message. For each
remark group, significant and normal terms will be
extricated to develop a key-term cloud. This key-term
cloud gives an initially show that clients can without much of a stretch and quickly comprehend the central matters of comparative remarks in a bunch. Besides, delegate marks in each gathering will likewise be identified. Our objective is to produce an enlightening, brief, and great interface that can assist clients with getting an outline understanding without perusing all remarks.Note that this paper doesn't zero in on the strategies of normal language handling (truncated as NLP).With utilizing some fundamental NLP methods, each comment is changed to a bunch of n-gram terms. On the other hand, we define new closeness measures to adequately decide the distance among remarks and
groups. As per the new definitions, we propose a completely gradual calculation that is nearly boundary
free and can deal with the exception issue. Moreover,the most significant benefit of our calculation is its high efficiency, showing that it can create bunching results with most recent approaching remarks progressively.These capacities unquestionably address the issue of remark stream synopsis on SNS.To check the viability of IncreSTS calculation, we gather genuine remark streams from Facebook and con pipe broad examinations with near strategies to show the strength and prevalence of our methodology.Generally, the commitments of this paper can be summarized as follows.
We model a clever gradual grouping issue in light of the prerequisites of remark stream aggregate summarization on SNS.
We propose IncreSTS calculation that can incrementally count update bunching results with most recent approaching remarks progressively.
We plan an initially show, which is compact, enlightening, and great, to help clients effectively and quickly get an outline understanding of a remark stream.

II. BACKGROND

Technologies used in this Project :

- **Python**

Python is a general-purpose, object-oriented programming language with high-level programming capabilities. It has become famous because of its apparent and easily understandable syntax, portability, and easy to learn.Python is a programming language that includes features of C and Java. It provides the style of writing an elegant code like C, and for object-oriented programming, it offers classes and objects like Java..Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.There are three main areas where PHP scripts are used.Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

- **Django**

Django is an MVT web framework that is used to build web applications. The huge Django web-framework comes with so many “batteries included” that developers often get amazed as to how everything manages to work together. The principle behind adding so many batteries is to have common web functionalities in the framework itself instead of adding latter as a separate library.
One of the main reasons behind the popularity of Django framework is the huge Django community. The community is so huge that a separate website was devoted to it where developers from all corners developed third-party packages including authentication, authorization, full-fledged Django powered CMS systems, e-commerce add-ons and so on. There is a high probability that what you are trying to develop is already developed by somebody and you just need to pull that into your project.

- **MySQL**

MySQL is an open source, SQL Relational Database Management System (RDBMS) that is free for many uses (more detail on that later). Early in its history, MySQL occasionally faced opposition due to its lack of support for some core SQL constructs such as sub selects and foreign keys. Ultimately, however, MySQL found a broad, enthusiastic user base for its liberal licensing terms, perky performance, and ease of use. Its acceptance was aided in part by the wide variety of other technologies such as PHP, Java, Perl, Python, and the like that have encouraged its use through stable, well-documented modules and extensions. MySQL has not failed to reward the loyalty of these users with the addition of both sub selects and foreign keys as of the 4.1 series. Like many competing products, both free and commercial, MySQL isn‘t a database until you give it some structure and form.

III. EXISTING SYSTEM

A. Description

The existing system mainly enforce the user to read the entire comments one by one to get the perspective of the comment.This system consumes more time and space of the user and becomes a hectic task for the user and user may mislead by the comments and get discouraged to continue reading the comments .The system doesn't provide a platform to filter the comments, thus making the process difficult for the user .As can be observed, not only the quantity of comments is large, but also the generation rate is remarkably high. Users unnecessarily and almost impossibly go over the whole comment list of each message. However, we may still desire to know what are they talking about and what are the opinions of these discussion participants.

B. Disadvantages of Existing System

Existing Systems are mainly forcing the users to spend more time in social media
Spending more time on computer screen
Time Consuming.
Not user friendly.

IV. PROPOSED SYSTEM

A. Description

Here the system is concentrating on the problem of short text summarization on the comment stream of a specific message from social network services (SNS). Due to the high popularity of SNS, the quantity of comments may increase at a high rate right after a social message is published. Motivated by the fact that users may desire to get a brief understanding of a comment stream without reading the whole comment list, we attempt to group comments with similar content together and generate a concise opinion summary for this message. Since distinct users will request the summary at any moment, existing clustering methods cannot be directly applied and cannot meet the real-time need of this application. In this system, we model a novel incremental clustering problem for comment stream summarization on SNS. Moreover, we propose IncreSTS algorithm that can incrementally update clustering results with latest incoming comments in real time. Furthermore, we design an at-a-glance visualization interface to help users easily and rapidly get an overview summary.This will reduce the users to spend more time in the screen.

B. Working of Proposed System

In this system, the user can perform multiple role. He/she can put a comment on his own posts as well as on other's post too. The user can even add new people as friends and can lead a conversation with them This system ensures time efficiency, as the user need not to spend time on reading the entire comments to get the abstract or synopsis of the viewer's point .This also ensures that the user won't get disoriented with multiple perspectives that can mislead the user Filtration of views enable the user to neatly understand the commenter's perspective

C.Advantages of Proposed System

- User friendly.
- More flexible than existing system.
- Less time consuming.
- Comforts every type of user.
- Adaptable even for all kinds of peoples.

V. MAIN MODULES OF THE SYSTEM

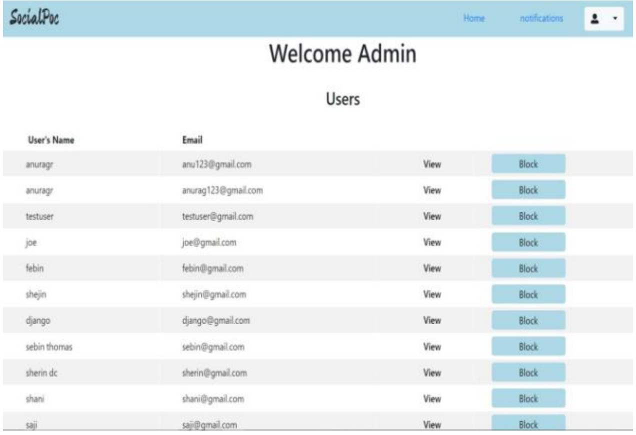
Admin

The administrator has the overall control of the system. If your user name and password is matching with administrator‘s user id and password, then you will enter in to the site as administrator.admin will Verify and block the users and admin should have the provision to view the details of users..

Users

In this module, the User is the one who uses the site. User can have Registration and Login page, user can add the post, view the post, can create a friend circle, have the provision to view the details of those who put the comments and user along can give the comments too.

VI. RESULTS AND DISCUSSIONS



Welcome Admin

Users

User's Name	Email		
anuragr	anu123@gmail.com	View	Block
anuragr	anurag123@gmail.com	View	Block
testuser	testuser@gmail.com	View	Block
joe	joe@gmail.com	View	Block
febin	febin@gmail.com	View	Block
shejin	shejin@gmail.com	View	Block
django	django@gmail.com	View	Block
sebin thomas	sebin@gmail.com	View	Block
sherin dc	sherin@gmail.com	View	Block
shani	shani@gmail.com	View	Block
saji	saji@gmail.com	View	Block

Fig 1 Admin Homepage

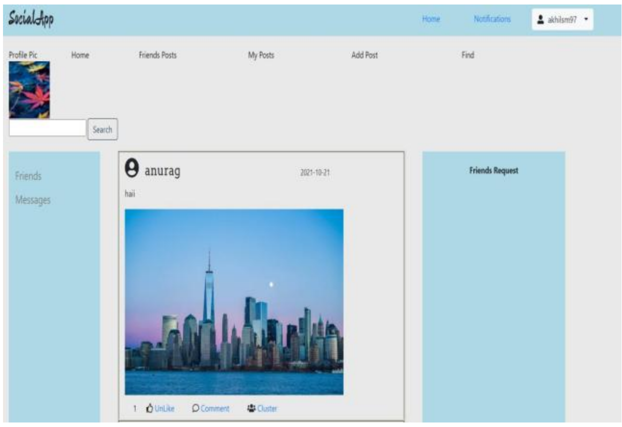


Fig 2 UserHomepage

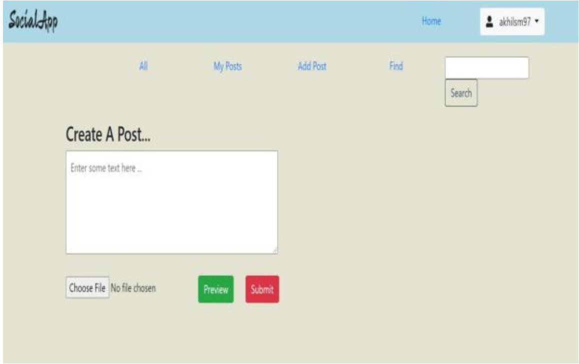
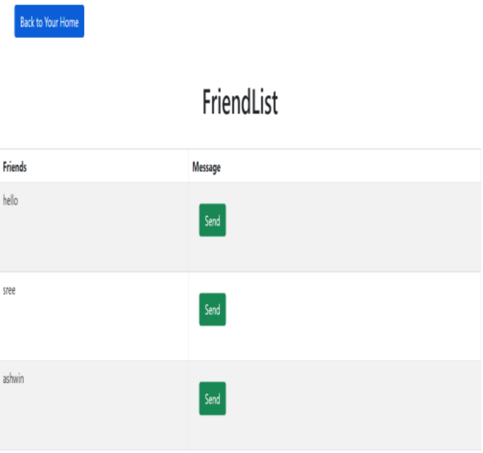


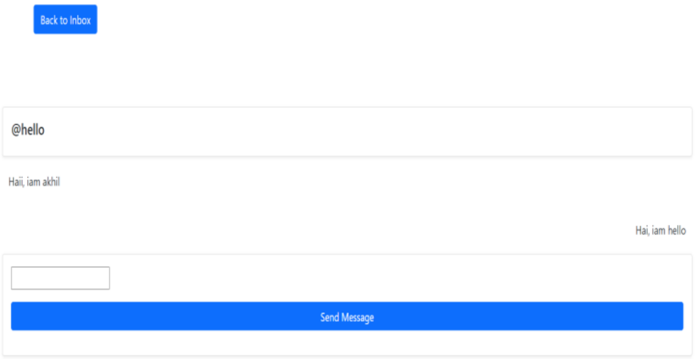
Fig 3 Add Post



FriendList

Friends	Message
hello	Send
sree	Send
ashwin	Send

Fig 4 Friend List

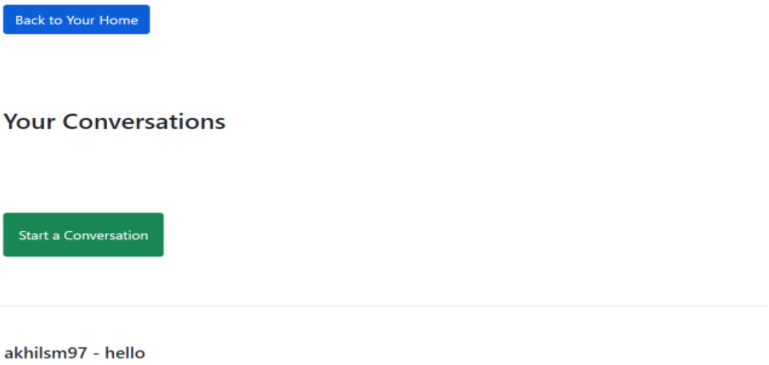


@hello

Haii, iam akhil

Hai, iam hello

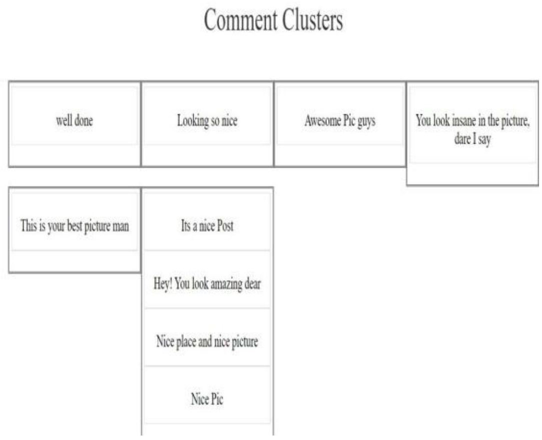
Fig 5 Public chat



Your Conversations

akhilsm97 - hello

Fig 6 Friends chat



Comment Clusters

well done

Looking so nice

Awesome Pic guys

You look insane in the picture, dare I say

This is your best picture man

Its a nice Post

Hey! You look amazing dear

Nice place and nice picture

Nice Pic

Fig 7 Clusters of comments

VII. CONCLUSION

SOCIAL MEDIA COMMENT EXTRACTOR USING NLP provide that users to get a brief understanding of a comment stream without reading the whole comment list, we group comments with similar content together and generate a concise opinion summary for this message.So that the user can efficiently manages the huge comments stream.For that system ensures time efficiency, as the user need not to spend time on reading the entire comments to get the abstract or synopsis of the viewer's point .This also ensures that the user won't get disoriented with multiple perspectives that can mislead the user Filtration of views enable the user to neatly understand the commenter's perspective.

REFERENCE

[1] http://www.cnnic.cn/gywm/xwzx/rdxw/2015/201507/t20150 723 _52626.htm

[2] Kyosuke N, Takahide H, Ko F. Improving tweet stream classification by detecting changes in word probability. In: Proc. of the ACM SIGIR 2012. 2012. 971í 980. [doi:
10.1145/2348283.2348412].

[3] Sriram B, Fuhry D, Demir E, Ferhatosmanoglu H, Demirbas M. Short text classification in twitter to improve information filtering.In: Proc. of the ACM SIGIR 2010. 2010. 841í 842. [doi: 10.1145/1835449.1835643]

[4] Mihalcea R, Courtney C, Strapparava C. Corpus-Based and knowledge based measures of text semantic similarity. In: Proc. of the AAAI 2006. 775- 780.

[5] Pu Q, Yang GW. Short-Text classification based on ICA and LSA. In: Proc. of the ISNN 2006. LNCS 3972, Heidelberg: Springer Verlag, 2006. 265-270. [doi: 10.1007/1176002339]

[6] Chatterjee N. A statistical approach for similarity measurement between sentences for EBMT. In: Proc. of Symp. on Translation Support Systems.

[7] Liu YP, Li S, Zhao TJ. System combination based on wsd using WordNet. Acta Automatica Sinica, 2010,36(11):1575í 1580 (in Chinese with English abstract). [doi: 10.3724/SP.J.1004.2010.01575]

[8] Yang Z, Fan KF, Lei JJ, Guo J. Text manifold based on semantic analysis. Acta Electronica Sinica, 2009,37(3):557í 561 (in Chinese with English abstract). [doi: 10.3321/j.issn:0372-2112.2009.03.024]

[9] Wang ZZ, He M, Du YP. Text Similarity Computing Based on Topic Model LDA, Computer Science, Dec 2013 Vo1.40 No.12 229-232

[10] Yang Z, Wang LT, Lai YX. Online comment clustering based on an improved semantic distance. Ruan Jian Xue Bao/Journal of Software, 2014,25(12):2777í 2789 (in Chinese). http://www.jos.org.cn/1000- 9825/4729.htm