Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 04, April -2018

SOCIAL Q&A:AN ONLINE SOCIAL NETWORK BASED QUESTION AND ANSWER SYSTEM

Prateeksha Chaurasia

Bharati vidyapeeth college of engineering,Pune

Abstract — *Question and Answer (Q&A) system presume a basic component in our day by day life for information and details sharing. Clients submit inquiries and select inquiries to reply in the framework. Due to the fact of the rapidly building client| populace and the amount of inquiries, it is improbable for a client to unearth an inquiry by opportunity that he can reply. philanthropy does not urge all clients to give solutions, also great solutions with a quick response hold up time. The essential purpose of this paper is to improve the execution of Q&A frameworks by currently sending inquiries to clients who are able and prepared to response the inquiries. To this end, we have composed and actualized SocialQ&A, an on the web informal local community based mostly Q&A technique. Social Q&A use the informal local community properties of standard intrigue and frequent believe in companion relationship to acknowledge an asker through kinship who are properly on the way to response the inquiry, and improve the clien security. We likewise improve SocialQ&A with security and productivity upgrades by guaranteeing client safety and distinguishes, and recovering answers naturally for repetitive inquiries. We depict the engineering and calculations, and led far reaching extensive scale reenactment to assess Social Q&A in correlation with different methods Our outcomes propose that informal organizations can be utilized to improve the appropriate response high quality and asker's holding up time. We likewise actualized a genuine model of Social Q&A, and dissect the Q&A |perform of real clients and inquiries from a minor scale certifiable Social Q&A technique.*

Keywords- *Question and answer systems, Social networks, Information search, Knn, NLP, SVM, TFIDF, Sentence level clustering*

1. INTRODUCTION

Internet is a essential wellspring of data, in which the measure of data is immense and continually establishing. Clients depend on internet search resource to find out certain data in this data base. Several clients could request or answer questions recognized} with delicate factors this kind of health-related problem, political activism or even sexual introduction. In spite of the truth that the client may possibly require the response at the earliest chance, he/despite almost everything she requirements the security assurance to remain away from possible divulgence of individual data The prepare of SocialQ&A depends on two informal organization properties. To begin with, social companions have a tendency to have comparative interests (e.g., lab people majoring in{Pc frameworks). 2nd , social companions have a tendency to be trustworthy and charitable due to the fact of the property of "kinship cultivates collaboration" By making use ofPseudo code for the inquiry answer mapping calculation ,Knn calculation and Support vector machines.

Considering that Social Q&A is based on informal communities. The asker and answerer are social close to each other. In this method, making certain the security is important and challenge. To handle this problem, we propose Social Q&A, an on the internet interpersonal organization based mostly Q&A framework, that efficiently advances inquiries to individuals {clients with the most astounding probability (ability and eagerness) of noting them with mastery and enthusiasm for the inquiries' subjects. The outline of Social Q&A depends on two informal organization properties. In the initial area, social companions have a tendency to have comparative interests (e.g., lab individuals an majoring in Computer frameworks) [9]. Second, social companions have a tendency to be trustworthy and philanthropic since of the residence of\ "fellowship encourages collaboration" . In like manner, Social Q& A favors routing inquiries between companions and recognizes an inquiry's possible answerers by taking into consideration two measurements: the enthusiasm of the companion in the direction of in direction of} the inquiry and the social closeness of the companion to the asker/forwarder. Hence, the suitable response beneficiaries have high probability of providing exceptional answers in a quick span. Not quite the exact same as the current Q&A frameworks, due to the fact of the significance of clients safety, we future acquaint security and effectiveness improvement with make certain clients security although clients utilizing informal organization noting questions. The commitments of this perform are as per the following:

_ The strategy of Social Q & A . Social Q& A is created out of three segments: User Interest, Question Categorizer, and Question-User Mapper. Client Curiosity Analyzer connects every single client with a vector of intrigue classes. Question Categorizer partners a vector of intrigue classifications to every single inquiry. At that point, in view of client intrigue and social closeness, Question-User Mapper recognizes prospective answerers for every single inquiry.

-The strategy of security and productivity upgrade techniques. Social Q& A consolidates three techniques to enhance its security and proficiency execution. The sprout channel based mostly individual data trade technique secures client'

safety which includes fellowship and intrigue data. The onion steering based answer sending technique secures the personalities of the asker and the answerer from becoming uncovered. The acceptable response recovery for repetitive inquiries consequently finds the responses for intermittent inquiries.

2. RESEARCH METHOD

Pseudo code for the question-answer mapping
Input : interest vector of a user ,his/her friends and question
Output: a list of potential answer providers
1: for each friend Uk in the friend set of Uj do
2:compute Ψ **I**,Uk based on eq.(1)
3: compute PSuk ,PAUk and PCUk based on eq.(2)
4:compute Ψ C,Uk based on eq.(3)
5: compute Ψ Uk based on eq.(4)
6: end for
7:Order the friends in descending order of Ψ Uk
8: notify the top N friends

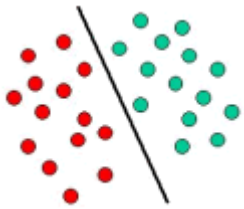
Knn algorithm :

1. Determine parameter k = number of nearest neighbor.
2. Calculate the distance between the query instance and all the training samples.
3. Sort the distance and determine nearest neighbor based on th k th minimum distance.
4. Gather the category y of the nearest neighbor.
5. Use simple majority of the category of nearest neighbor as the prediction value of query instance.

3 Support vector machines

In today's machine learning applications, support vector machines (SVM) are considered amust try—it offers one of themost robust and accurate methods among all well-known algorithms. It has a sound theoretical foundation, requires only a dozen examples for training, and is insensitive to the number of dimensions. In addition, efficient methods for training SVM are also being developed at a fast pace.

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. A schematic example is shown in the illustration below. In this example, the objects belong either to class GREEN or RED. The separating line defines a boundary on the right side of which all objects are GREEN and to the left of which all objects are RED. Any new object (white circle) falling to the right is labeled, i.e., classified, as GREEN (or classified as RED should it fall to the left of the separating line).



- Classification SVM Type 1 (also known as C-SVM classification)
- Classification SVM Type 2 (also known as nu-SVM classification)
- Regression SVM Type 1 (also known as epsilon-SVM regression)
- Regression SVM Type 2 (also known as nu-SVM regression)

1) CLASSIFICATION SVM TYPE 1

2) For this type of SVM, training involves the minimization of the error function:

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

subject to the constraints:

$$y_i \left(w^T \phi(x_i) + b \right) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, i = 1, \ldots, N$$

where C is the capacity constant, w is the vector of coefficients, b is a constant, and ξ_i represents parameters for handling nonseparable data (inputs). The index i labels the N training cases. Note that $y \in \pm 1$ represents the class labels and xi represents the independent variables. The kernel ϕ is used to transform data from the input (independent) to the feature

space. It should be noted that the larger the C, the more the error is penalized. Thus, C should be chosen with care to avoid over fitting.

3) CLASSIFICATION SVM TYPE 2

In contrast to Classification SVM Type 1, the Classification SVM Type 2 model minimizes the error function:

$$\frac{1}{2} w^T w - v\rho + \frac{1}{N}\sum_{i=1}^{N} \xi_i$$

subject to the constraints:

$$y_i\left(w^T \phi(x_i) + b\right) \geq \rho - \xi_i, \xi_i \geq 0, i = 1,\ldots, N \text{ and } \rho \geq 0$$

In a regression SVM, you have to estimate the functional dependence of the dependent variable y on a set of independent variables x. It assumes, like other regression problems, that the relationship between the independent and dependent variables is given by a deterministic function f plus the addition of some additive noise:

Regression SVM

y = f(x) + noise

The task is then to find a functional form for f that can correctly predict new cases that the SVM has not been presented with before. This can be achieved by training the SVM model on a sample set, i.e., training set, a process that involves, like classification (see above), the sequential optimization of an error function. Depending on the definition of this error function, two types of SVM models can be recognized:

Algorithm :TF_IDF

The **tf-idf** value increases proportionally to the number of times a word appears in the document, but is often offset by the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general. Nowadays, **tf-idf** is one of the most popular term-**weighting** schemes.

Document 1: The game of life is a game of everlasting learning
Document 2: The unexamined life is not worth living
Document 3: Never stop learning

Step 1: Term Frequency (TF)

Term Frequency also known as TF measures the number of times a term (word) occurs in a document. Given below are the terms and their frequency on each of the document.

Step 2: Inverse Document Frequency (IDF)

The main purpose of doing a search is to find out **relevantdocuments** matching the query. In the first step all terms are considered equally important. In fact certain terms that occur too frequently have little power in determining the relevance. We need a way to **weigh down** the effects of too frequently occurring terms. Also the terms that occur less in the document can be more relevant. We need a way to **weigh up** the effects of less frequently occurring terms. Logarithms helps us to solve this problem.

Let us compute IDF for the term **game:**

$idf(t,D) = log|\ D\ |1+|\ \{d \in D : t \in d\}\ |$

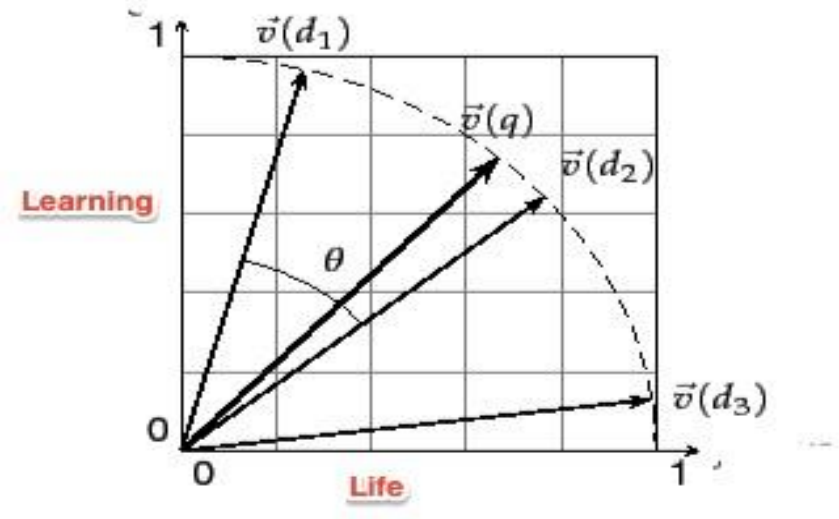
Step 3: TF * IDF

Remember we are trying to find out relevant documents for the query: **life learning** For each term in the query multiply its normalized term frequency with its IDF on each document. In Document1 for the term **life** the normalized term frequency is 0.1 and its IDF is 1.405507153. Multiplying them together we get **0.140550715** (0.1 * 1.405507153). Given below is TF * IDF calculations for **life and learning** in all the documents.

Step 4: Vector Space Model – Cosine Similarity

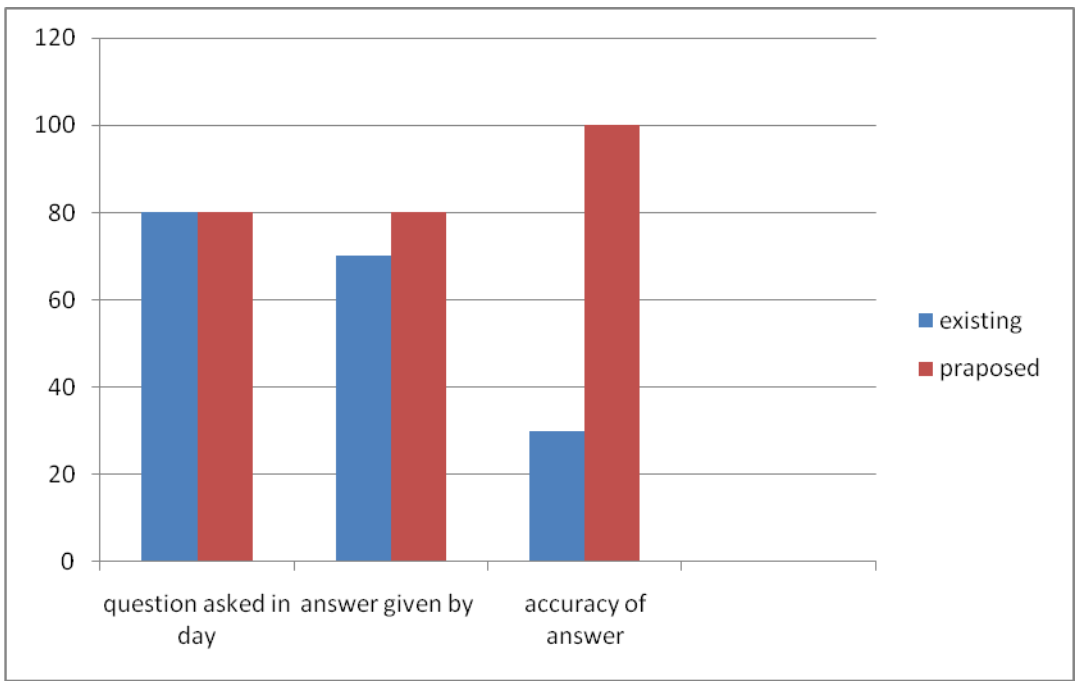
From each document we derive a vector. If you need some refresher on vector refer here. The set of documents in a collection then is viewed as a set of vectors in a vector space. Each term will have its own axis. Using the formula given below we can find out the similarity between any two documents.

$$cos(\theta) = \frac{v \cdot w}{\|v\| \|w\|} = \frac{\sum_{i=1}^{n} v_i w_i}{\sqrt{\sum_{i=1}^{n} v_i^2} \sqrt{\sum_{i=1}^{n} w_i^2}}$$



3. RESULTS AND ANALYSIS

Sr .no	Parameter	Existing	Proposed
1	Question asked in a day	80%	80%
2	Answer given by user	70%	80%
3	Accuracy of answer	30%	100%



4. CONCLUSION

In our project we are providing high quality reply with significantly less holding up time to quantity of users. For high quality answer and significantly less hold up time we have developed and prototyped an on the internet informal community based mostly Q&A framework, known as referred to Social Q &A.It utilizes the properties of an interpersonal organization to forward an inquiry to prospective answer suppliers, guaranteeing that a provided inquiry will get a best notch reply in a brief time system . It expels the excess weight from answer suppliers by straightforwardly conveying them the inquiries they may possibly be keen on, rather than requiring answer suppliers to appear by means of a massive gathering of inquiries as in Yahoo! Answers or flooding an inquiry to the majority of an asker's companions in an on the internet informal organization. The sprout channel based mostly improvement techniques scramble the intrigue and Friendship data traded in between clients to secure client safety , and record all n-grams of addressed inquiries to naturally recover answers for repetitive question .Considering that identical inquiries may possibly be displayed contrastingly and a related inquiry may possibly be addressed diversely in different circumstance.

ACKNOWLEDGEMENTS

REFERENCES

[1] M. R. Morris, J. Teevan, and K. Panovich. A Comparison ofInformation Seeking Using Search Engines and Social Networks.In In Proc. of ICWSM, 2010.
[2] M. R. Morris, J. Teevan, and K. Panovich. What do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&A Behavior. In Proc. of CHI, 2010.
[3] Z. Gyongyi, G. Koutrika, J. Pedersen, and H. Garcia-Molina.Questioning Yahoo! Answers.In Proc. of QAWeb, 2008.
[4] Yahoo!Answers Team. Yahoo! Answers BLOG. http://yahooanswers.tumblr.com/, [Accessed on 10/20/2014].
[5] B. Li and I. King. Routing Questions to Appropriate Answerers in Community Question Answering Services.In Proc. of CIKM, 2010.fusion," *SIGIR'06*, 2006.
[6] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In Proc. of WWW, 2008.

