Methods of Automated Question Answering

Aman
Department of
Applied Mathematics
Delhi Technological University
amankumar45695@gmail

Anikait Gupta
Department of
Applied Mathematics
Delhi Technological University
anikaitgupta3@gmail.com

Aditya Singh
Department of
Applied Mathematics
Delhi Technological University
adityasngh342@gmail.com

Abstract:
It is human nature to ask questions. Question answering is a very important part of learning. We ask questions to learn more about something and answers help us to gain a better perspective.

The internet is taking over everything in the world. Millions of queries are asked on the internet everyday, most of which are repeated. Hence, it is inefficient to answer the same question repeatedly. We can use the answers of previously asked questions to answer new questions.This can be done by using various Natural Language Processing (NLP) techniques like Bag of Words (BOW), word-to-vec, Glove, BERT, etc.

In this paper, we try to provide a quantitative analysis between these techniques. We also try to find out what is the best technique to make automated question answering mechanisms.

I. INTRODUCTION

Let us assume a user comes to a website of RTI (right to information) and asks a very common question "*Through this site, to which public authority may I submit a request?*". This website already has a whole bunch of FAQs (Frequently Asked Questions) which consist of information including something relevant for this question.When we visit the website FAQs we find a question *'To which Public Authority can I file a request through this portal?',* which is syntactically not the same question, but it is semantically close enough .If the user gets the answer for this question, it will be relevant. So, this is what the above stated techniques do. These techniques look at the user queries and find the closest FAQ to the user query.

Question answering system works in three stages.

1.First, question analysis i.e. what type of question is being asked. Questions can be a general question (with yes or no as answer) or wh- question (start with what ,why ,where etc.) or choice question or fotoid question (in which answer is in the text).
2.Second, document analysis which consists of extracting answers from given documents or from open sources such as wikipedia.
3.Third, answer analysis which consists of ranking the best answer from the set of answers that are obtained from the second step.
In this experiment we have taken a training dataset of FAQs asked on a website and we will ask a question. We will try to find out how close a bag of words will give the answer to our question.

II.APPROACH

The approach that will be followed here will be to convert these sentences into vectors. Firstly, convert given questions into vectors and secondly, convert the given FAQs into vectors and then find closeness between these two vectors.There are many ways to find closeness of the two vectors. For eg. cosine similarity,jaccard similarity etc. We will use cosine similarity here which looks at the angle between two vectors without regard to their magnitude.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2} \sqrt{\sum\limits_{i=1}^{n} B_i^2}},$$

For example:
Line 1:*Through this site, to which public authority may I submit a request?*
Line 2:*To which Public Authority can I file a request through this portal?'*

	submit	through	this	to	which	public	authority	may	i	submit	request	portal	can	site
line 1	0	1	1	1	1	1	1	1	1	1	1	0	0	1
line 2	1	1	1	1	1	1	1	0	1	1	0	1	1	0

Cosine similarity=7/(√11*√12)=.60

III. Text Preprocessing

Text preprocessing is the first step of data mining. As we want to give only necessary information to the machine and preprocessing technique helps to remove unnecessary data from given text. Preprocessing is done to remove the noise and to fill the missing values. Text preprocessing includes following steps:

1.Tokenization: A machine cannot understand a complete sentence at once.So,tokenisation is used to break down the sentence into words the way a machine can understand.

2.Lower casing: Words like 'machine' and 'Machine' mean different to computer if not converted into lowercase. So, lowercasing means converting words into lowercase.

3.Stopwords Removal: Stopwords are the words like (a, an, etc) which are not important features in differentiating two sentences.So, stopwords can be removed.

4.Stemming or Lemmatization: It is the process of bringing a word into its root form.It helps to decrease the vocabulary size two to three fold. For example: The word 'developed' and 'development' originated from the same word 'develop' hence can be brought back down to its base word while preprocessing.Lemmatization is the process of bringing words into its dictionary root form.

IV.MODELS

A.BAG OF WORD MODEL:

Bag of word is a natural language processing technique of feature extraction.This is the most simple approach of extracting features. This model basically deals with frequency of a word in the document and has nothing to do with grammatical details and order of occurance. It only examines whether a word is present in a document or not. Let us assume we are given following sentences:

Line 1: I go to a French restaurant.
Line 2: I go to an American restaurant.
Line 3: I go to a French restaurant in America.

After applying all the preprocessing steps, the words like a and an an removed and the word american is treated as america after applying stemming. Here, it is assumed that our dictionary contains all the words. The binary vector will look like:

Line 1: [1,1,1,1,1,1];
Line 2: [1,1,1,0,1,1];
Line 3: [1,1,1,1,1,1];

output:
We try to do an experiment on how accurate the bag of words approach is in predicting the answer of a given question. We will create a bag of words representing the question and that of faqs and find the closest match. The asked question is '*what does data scientist usually do*'

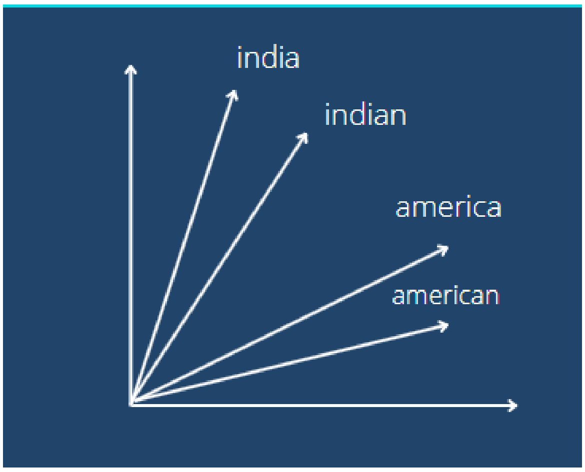
```
0 1.0 what does the job hunting experience look like
1 0.31622776601683794 any insights you can offer about the ds job
market
2 0.1643989873053573 whats the impact of covid on hiring for ds ro
les
3 0.19611613513818404 what skills and qualities do employers look
for in a data scientist
4 0.19611613513818404 do employers look for an advanced ml degree
5 1.0 how does a typical day of a data scientist look like
6 0.058722021951470346 is preparation of algorithms and data struc
tures needed for a data science interview
7 0.1643989873053573 what is the mathematical background required
to be a data scientist
8 0.1643989873053573 what are the various rounds in a data scienti
st interview
9 0.1414213562373095 what level of proficiency is needed for a dat
a scientist in coding


Question:  what does a data scientist usually do


Retrieved:  What does the job hunting experience look like ?
Job hunting experience involves networking to get in touch with th
e right people in various companies, applying to lots of jobs thro
ugh various channels, preparing for interviews – while interviews
are uncertain it is necessary to prepare well what you CAN prepar
e, smart scheduling of interviews to get the best job and salary y
ou can.
```

So, we can clearly see from our asked question and retrieved answer that it is not a good response.
This happens because the bag of word technique is only to look at how much overlap between the two sentences. It does not look like these sentences are similar in meaning. This issue can be addressed by word2vec technique.

B.WORD 2 VEC:
Word2Vec is an unsupervised learning technique in which words are represented in the form of vectors so that similar sounding words can be close in cosine similarity or they occur together in their vector representation. Word2Vec is widely used because Google has released a set of pretrained embedding on google news data set over 100 billion words which give three hundred dimension vectors on over millions words. So,training is not required. Second, an important property is that we can do arithmetic on a set of words to give meaning in some sense. Word2Vec architecture consists of a layer which performs dot product between weight matrix and input vector. The product is passed to the output layer which makes its dot product with vectors of output layer and then probability is calculated by softmax activation function.



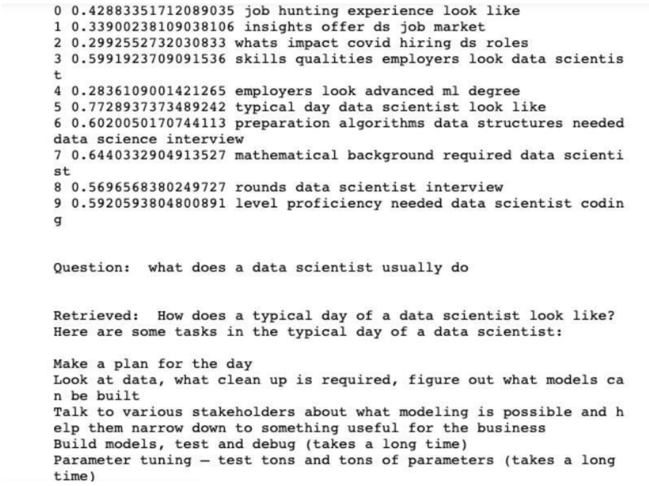
Suppose there are four vectors: India, Indian, America and American. India and America are closely related to each other as both are countries. A arithmetic can be derived from above given vectors:

America =India-Indian+American;

The thing that can be derived from the above equation is that doing arithmetic on a set of words gives their combined meaning more concisely if given a sentence to get its meaning just add up all the word vectors.

output:

We have used the google news model here.This is the result from word2vec model.



```
0 0.42883351712089035 job hunting experience look like
1 0.33900238109038106 insights offer ds job market
2 0.2992552732030833 whats impact covid hiring ds roles
3 0.5991923709091536 skills qualities employers look data scientis
t
4 0.2836109001421265 employers look advanced ml degree
5 0.7728937373489242 typical day data scientist look like
6 0.6020050170744113 preparation algorithms data structures needed
data science interview
7 0.6440332904913527 mathematical background required data scienti
st
8 0.5696568380249727 rounds data scientist interview
9 0.5920593804800891 level proficiency needed data scientist codin
g


Question:  what does a data scientist usually do


Retrieved:  How does a typical day of a data scientist look like?
Here are some tasks in the typical day of a data scientist:

Make a plan for the day
Look at data, what clean up is required, figure out what models ca
n be built
Talk to various stakeholders about what modeling is possible and h
elp them narrow down to something useful for the business
Build models, test and debug (takes a long time)
Parameter tuning – test tons and tons of parameters (takes a long
time)
```

From the above result it can be clearly seen that the word2vec model gives a more accurate result over a bag of words model. Here, the retrieved question is *'How does a typical day of data scientist look like'* which is more closely related to the asked question than that we get from the bag of word model in which we get a retrieved question which is far away from the asked question.

C.BERT:

BERT stands for Bidirectional Encoder Representations from Transformers. It captures a longer range of context. It is based on the concept of transformer architecture and masked learning model.BERT is a new model developed by researchers at Google for various natural language processing tasks. It can help computers to understand text using surrounding words. It was pre trained using a large wikipedia dataset and later fine tuned as per the required dataset. BERT is bidirectional as it understands from both left and right side during training.It is different from word2vec on the working method by randomly removing from the sentence and then trying to predict the sentence by some words.For instance given two sentences *'stand too my right'* and *'Your answer is right'*.In both the sentence the word *'right'* have different meaning according to context.BERT gives different embedding for both these two different context. The same word right has two different embeddings.This happens because BERT gives vector embedding for the whole sentence.

TYPES OF BERT MODEL:

There are two types of BERT available:

1. BERT base consists of 12 layers,12 attention heads and 110 million parameters.
2. BERT large which consists of 24 layers,16 attention heads and 340 parameters.

output:

We fetch the BertClient model from the server running in the background.There is no need to remove the stopwords because BERT is capable of using those subtle differences.

```
0 0.84057343 what does the job hunting experience look like
1 0.722398 any insights you can offer about the ds job market
2 0.70140886 whats the impact of covid on hiring for ds roles
3 0.8119571 what skills and qualities do employers look for in a d
ata scientist
4 0.7749537 do employers look for an advanced ml degree
5 0.86003697 how does a typical day of a data scientist look like
6 0.7659602 is preparation of algorithms and data structures neede
d for a data science interview
7 0.8592837 what is the mathematical background required to be a d
ata scientist
8 0.84925056 what are the various rounds in a data scientist inter
view
9 0.82339096 what level of proficiency is needed for a data scient
ist in coding


Question:  what does a data scientist usually do


Retrieved:  How does a typical day of a data scientist look like?
Here are some tasks in the typical day of a data scientist:

Make a plan for the day
Look at data, what clean up is required, figure out what models ca
n be built
Talk to various stakeholders about what modeling is possible and h
elp them narrow down to something useful for the business
Build models, test and debug (takes a long time)
```

As seen from the above output, we can see that we get the correct answer.The score is very high i.e. 0.8 6 which is greater than all above three models.

V.CONCLUSION

Automated question answering helps machines to generate answers without the help of humans . A lot of NLP based techniques have been proposed in the literature.. In this paper, we analyzed some of the most prominent Question Answering techniques including bag of words which deals with frequency of word and nothing to do with grammatical details ,word to vec which analyze words on basic of cosine similarity of their vector representation , GloVe which is trained using matrix factorisation techniques on word co-occurrence matrix. BERT which makes word embedding on the basis of their context in the sentence . We found that the BERT technique gives best results out of all the models because it deals with the whole context of the sentence in which the word is present

VI. FUTURE WORK

We may clearly deduce that this issue is far from being resolved. As previously stated, present approaches have a number of flaws, and more research is needed to overcome these limits and build an usable product that can reliably provide answers.

VII. REFERENCES

[1] Green BF, Wolf AK, Chomsky C, and Laughery K. Baseball: An automatic question answerer. In Proceedings of Western Computing Conference, Vol. 19, 1961, pp. 219–224.

[2] Clark P, Thompson J, and Porter B. A knowledge-based approach to question answering. In Proceedings of AAAI'99 Fall Symposium on Question-Answering Systems, 1999, pp. 43-51.

[3] Ravichandran D and Hovy E. Learning surface text patterns for a question answering system. In proceeding of 40th Annual Meeting on Association of Computational Linguistics, 2002, pp. 41-47.

[4] Zhang D and Lee WS. Web based pattern mining and matching approach to question answering. In Proceedings of the 11th Text REtrieval Conference, 2002.

[5] Sneiders E. Automated question answering using question templates that cover the conceptual model of the database. In Natural Language Processing and Information Systems, Springer Berlin Heidelberg, 2002, pp. 235-239.

[6] Greenwood M. and Gaizauskas R. Using a Named Entity Tagger to Generalise Surface Matching Text Patterns for Question Answering. In Proceedings of the Workshop on Natural Language Processing for Question Answering (EACL03), 2003, pp. 29-34.

[7] Moschitti A. Answer filtering via text categorization in question answering systems. In Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, 2003, pp. 241-248.

[8] Doaa Mohey El-Din, Hoda M.O. Mokhtar and Osama Ismael, "Online Paper Review Analysis", International Journal of Advanced Computer Science and Applications(IJACSA), 6(9), 2015.

[9] Lukasz, A., Tomasz, K., Piotr, S., and Włodzimierz, T.,"Simpler is Better? Lexicon-based Ensemble Sentiment Classification Beats Supervised Methods", International Workshop on Curbing Collusive Cyber-gossips in Social Networks (C3-2014), Proc. IEEE/ACM Int. Conf. Advances in Social Network Analysis and Mining, ASONAM, Beijing, China,August 17, 2014

[10] Ilia, C., & Natalia, L., "Extraction and Use of Opinion Words for Three-Way Review Classification Task", CDUD'11–Concept Discovery in unstructured data, 2011 [11] Georgios, P., & Mike, T., "More than Bag-of-Words: Sentence-based Document Representation for Sentiment Analysis", Proceedings of Recent Advances in Natural Language Processing, pages 546–552, Hissar, Bulgaria, September 2013.

[12] Rugved, D., Ketan, V., Suratsingh, R., & Tushar, J., "Comparative Study of Document Similarity Algorithms and Clustering Algorithms for Sentiment Analysis", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 5, September-October 2014 ISSN 2278-6856,2014

[13] Joseph, V. H., and Paul, J. K., "A COMPARISON OF THE RELATIONAL DATABASE MODEL AND THE ASSOCIATIVE DATABASE MODEL", Issues in Information Systems, Volume X, No. 1, 2009.

[14] Efthymios, K., Theresa, W., and Johanna, M.,"Twitter Sentiment Analysis: The Good the Bad and the OMG!",