ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 2, Issue 12, December 2013

QUESTION CLASSIFICATION FOR QUESTION ANSWERING SYSTEM USING BACK PROPAGATION FEED FORWARD ARTIFICIAL NEURAL NETWORK (BPFFBNN) APPROACH

Rishika Yadav
(Asst. Prof) SSCET Bhilai

Prof Megha Mishra.
(Sr. Asst. Prof.) SSCET Bhilai

Abstract
Question classification is one of the important tasks in question answering system. Question classification mainly includes two intermediate processes, feature extraction and question classification. In this paper for making knowledge base, we have extracted three features of the text question: lexical feature, semantic feature and syntactic feature. We have used Li & Roth two-layer taxonomy for categorization of questions. This taxonomy mainly divides the text question into 6 course grain categories and 50 fine grain categories. As per literature, many approaches to question classification have been proposed and reasonable results have been achieved. In this paper we have mainly used supervised machine learning technique of question classification. We have introduced a multilayer feed forward back propagation artificial neural network approach for question classification. This paper presents our research work on automatic question classification through this artificial neural network algorithm. We have discussed feature extraction process, algorithms, research work on question classification and results.

General Terms
Algorithms, Experimentation.

Index Terms back propagation, artificial neural network Question answering, text classification, machine learning, and neural network.

1. INTRODUCTION

Question retrieval is one of the main tasks in web based answering retrieval system and question classification is the main concern of the researchers all over the world and they are developing various methodologies to overcome this process complexity [12]. Question answering (QA) field is coming under the discipline of computer science within the field of retrieval and natural language processing (NLP). The main aim of which is to build the system which mechanically answers question inputted by humans in a natural language. A question answering system is a computer program which takes human input and gives an answer by querying structured database of information Language Processing (NLP), Information Retrieval (IR) and Information Extraction (IE) communities [11]. The main concept of this paper is to increase the efficiency of classifier with the use of different feature set. Following figure shows basic functional diagram of question answering system.

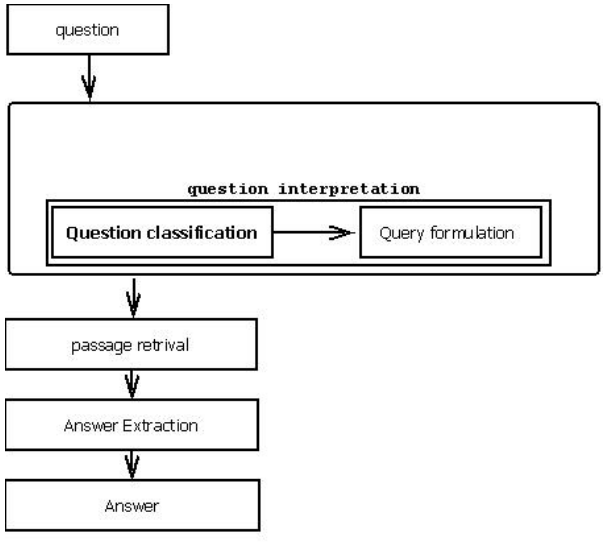

Figure1: functional diagram of Question Answering system

2. QUESTION CLASSIFICATION

Question classification is one of the main considerations in question answering system [2]. Question classification is the task of assigning a Boolean value to each pair hqj , cii 2 Q×C_p, where Q is the domain of questions and C_p= {c1, c2, . . . , c|C|} is a set of predefined categories. The complete work has been carried out in two phases: training phase and recognition phase. The strategies that have been applied for the classification is an artificial neural network (ANN). In the present work first we extract various features of the text with the use of various technologies.

Also the collection of the different category question has been done from the database of Text Retrieval Conference (TREC) [2]. In the present work how the question classification has been done has been rendered in a figure.

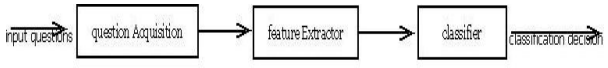

Fig 2 Question classification model

After that question acquisition has been done and different features of the questions are extracted .During the process of feature extraction various intermediate process is applied to improve the efficiency of the text we enter in input. In this thesis we mainly extract three features of the sentences and then we merge all three features to find the category of the questions. After feature extraction question classification is done. For questions classification we first use Li & Roth's two-layer taxonomy (X. Li & Roth, 2002) according to which questions should be categorized, and secondly, we must devise a strategy to use for the classifier.

Table 1: Li & Roth's two-layer taxonomy.

Coarse	**Fine**
ABBREVIATION	expansion ,abbreviation
DESCRIPTION	definition, manner, reason, description,
ENTITY	animal, colour, creative, currency, medical, disease, event, instrument, language, letter, other, substance, plant, product, religion, sport, term, symbol, technique, vehicle, word, food, body,
HUMAN	individual,description, group, title
LOCATION	city, mountain, other, state country,
NUMERIC	code,distance,date,money, order,other,percent,period,speed, temperature,size, weight, count

3 FEATURE EXTRACTION

In this work we mainly simulate the process of question classification with the use of supervised learning algorithms for classification. In this work,
We use a bag of word as lexical features. One of the main disputes in developing a supervised classifier for a particular domain is to identify and build a set of features. In this paper, we create transition patterns using three types of features:
3.1 Lexical
3.2 Semantic
3.3 Syntactic

3.1 Lexical Features

Lexical feature is the features or attributes of an instance which help identify the intended sense of target word are identified. In this work we use
Stemming and stop-word removal to decrease the size of word related question set.

3.1.1 Stop-word

This is a type of word which is filtering out prior to processing of natural language data. Fore. g. -'the whole', 'that was', 'that is'. These are non informative words which are not used in the classification process

3.1.2 Stemming removal – these are the document retrieval technique is commonly used in classification .It basically reduces the grammatical roots. For e.g. *Who did the Mahatma Gandhi killing?* Here after stemming removal killing is reduce to kill. We use a poster's stemming algorithm (1980).

3.2 semantic features

A semantic feature is a writing method which can be used to express the existence or non-existence of pre-set up semantic properties. This class includes a semantically improved version of the headword, and named entities.

3.2.1 Name Entity

Named entity is an important consideration for information extraction (IE) task [22] .In this project we use Stanford Named Entity Recognizer (NER) .And we use mainly 7 name entity classes MUC-7 (time, location , organization , person , money , percent , date). For e.g. what is India national flower? A named entity recognizer would (ideally) identify the following named entities (NE) -what is ($_{NE\_location}$ India) national flower.

3.2.2 Semantic headword (WordNet)
WordNet (Fellbaum, 1998) is a large English lexicon in which meaningfully related words are connected via cognitive synonyms (synsets) [14]. The WordNet is a useful tool for word semantics analysis and has been widely used in question classification. A natural way to use WordNet is via Hypernyms: B is a hyponym of A if every A is a (kind of) B. In the feature extraction process, we use two approaches to augment WordNet semantic Features, with the first augmenting the hypernym (super ordinates) and hyponyms (sub ordinates) of the headword [19]. And we make 50 clusters of fine grain category (X. Li & Roth, 2002).

3.3 syntactic features
Syntax feature is used to refer directly to the rules and principles regulate the sentence structure of input text sentence. With the use of syntactic feature we can design general rules that apply to all natural languages. This class of features include the question headword in and part-of-speech tags

3.3.1 Question headword
The headword mainly contains the information of the sentence [3]. For e.g. what is India national flower**?** In this question the flower is the main indication to correctly classify ENTITY: PLANT. For extraction of head word parse tree of the sentence is required .We use the Stanford parser for parsing process the following figure shows the parsing tree generated by Stanford parser.

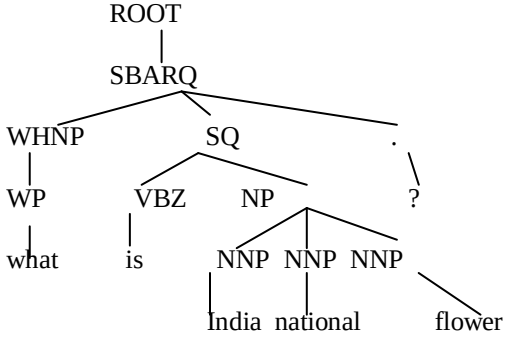

Figure 3: parse tree of the question-. what is India national flower ?

3.3.1.1 Question headword extraction algorithm

Process
EXTRACT_QUESTION_HEADWORD (*tree*, *rules*)
If TERMINAL? (*tree*)
Then
Return *to tree*
Else
Child ← APPLY-RULES (*tree, rules*)
Return EXTRACT_QUESTION_HEADWORD (*child*, *rules*)
End if
End process

The algorithm is mainly work by finding the head A_h of a non-terminal B with production rule $A \rightarrow Bi \cdots Bn$, using the head-rules which decide which of the $Bi \cdots Bn$ is the headword .This process is repeated continuously from A_h, until the terminal is reached

3.3.2 Parts of speech (POS)
Parts-of-speech (POS) tags are the grammatical classes of question token .POS are pre-terminal nodes for future use. For e.g. WP-VBZ-NNP-NNP-NNP is POS of the question "what is India national flower?"

4. Back propagation feed forward artificial neural network (BP-FFANN) classifier

The BP learning method became a popular methoTo train FFANN [18, 21]. The algorithm is a trajectory-driven technique that is corresponds to an error minimizing process. BP learning requires the neuron transfer function to be differentiable and sustain from the possibility of falling into local minima. The method is also known to be sustained to the initial weight settings, where many weight initialization techniques have been proposed to lessen such a possibility.

4.1 Differentiable activations function
We have to use a kind of activation function other than the step function used in perceptron [[18, 2]] because interconnected perceptrons produces the composite function is discontinuous, and therefore the error function is also discontinuous. One of the more popular activation functions for back propagation networks is the sigmoid,A real function sc : IR (0, 1) defined by the expression

$$sc(x) = \frac{1}{1 + e^{-cx}}.$$

The constant c is called reciprocal 1/c and c can be selected arbitrarily. The temperature parameter in stochastic neural networks. According to the value of c the shape of the sigmoid changes.

4.2 Regions in input space

The sigmoid's output range contains all numbers strictly between the given reason .Both extreme values can only be reached asymptotically. The computing units considered in this paper evaluate the sigmoid using the net amount of excitation as its argument. Given weights w1, . . . ,wn and a bias −α, a sigmoidal unit computes for the input A_1, . . . , An the output

$$\frac{1}{1 + \exp\left(\sum (i=1 \text{ to } n)\ wiAi - \alpha\right)}$$

4.3 Local minima of the error function

A price has to be paid for all the positive features of the sigmoid as activation function. The most important problem is that, local minima appear in the error function which would not be there if the step function had been used under some circumstances, .The function was computed for a single unit with weights, constant threshold, and four input-output
Patterns in the training set.

5 Back propagation feed forward artificial neural networks (BP-FFANN) classification algorithm

Backpropagation algorithm.
We Consider a network with a real input x and network function F [21]. The derivative F0 (x) is computed in two phases: part1-Feed-forward: the input x is fed into the network. The primitive functions at the nodes and their derivatives are evaluated at each node. The derivatives are stored. Part2-Backpropagation: the constant 1 is fed into the output unit and the network is run backwards. Incoming information to a node is added and the result is multiplied by the value stored in the left portion of the unit. The result is transmitted to the left of the unit. The result collected at the input unit is the derivative of the network function with respect to x.

5.1 (BP-FFANN) Training Algorithm

1 Initialize I=1; (W (I) randomly;
- While (stopping criterion is not satisfied or I< max-iteration
  - For each example (X, D)
    {1. Run the network with input X and compute the Output Y;
    2. Update the weight in backward order. starting from those of the output layer computed using the (generalized delta rule explained below;}
  - I=n+1;

End while

6 Result and discussion

6.1 Related work

Zhang & Lee [6] performed a number of experiments on question classification using the same taxonomy as Li & Roth, as well as the same training and testing data. In an initial experiment they compared different machine learning approaches with regards to the question classification problem: Nearest Neighbours (NN), Naıve Bayes (NB), Decision Trees (DT), SNoW, and SVM. NN, NB, and DT are by now fairly standard techniques and good descriptions of them can be found in for instance .The feature extracted and used as input to the machine learning algorithms in the initial experiment was bag of- words and bag-of-*n* grams (all continuous word sequences in the question). Questions were represented as binary vectors since the term frequency of each word or *n* gram in a question usually is 0 or 110. The results of the experiments are shown in table 6.1 [6]
The results of the SVM algorithm presented in table 6.1 are when the linear kernel is used. This kernel had as good performance as the RBF, polynomial, and sigmoid kernels

Table 2: Results from Zhang & Lee (2003) [7].

	Bag-of-words
Algorithm	Coarse grain (in%)
NN	75.6
NB	77.4
DT	84.2
SNoW	66.8
SVM	85.8

6.2 Experimental setup and results

The question BPANN classifier was carried out on the publicly available dataset of Text Retrieval Conference (TREC) [1]. In this work we take training set of 1000 questions, and a test set with 100 questions. The annotated question categories follow the question type taxonomy described in part 3.1 of this paper. This data set is also one of the most commonly used in the literature for supervised learning technique and we achieved 86% accuracy which out forms in all the previous result. Shown in the table below. 1. This

result should be achieved by following the steps--We take 1000 question for training data after training by feed forward back propagation artificial neural with 10 numbers of neurons and we use a TANSIG transfer function network we get these results shown in figure 6.3,6.4,6.5.

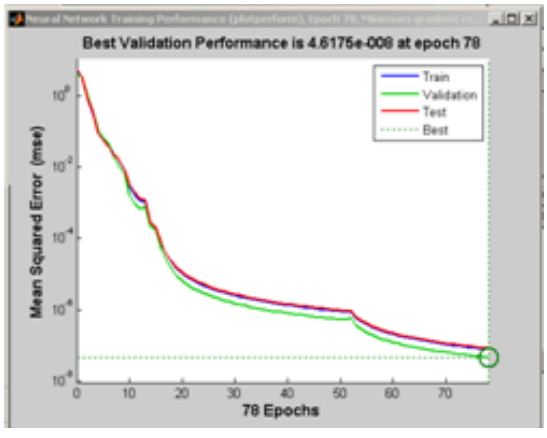

Figure 4: performance of training input dataset

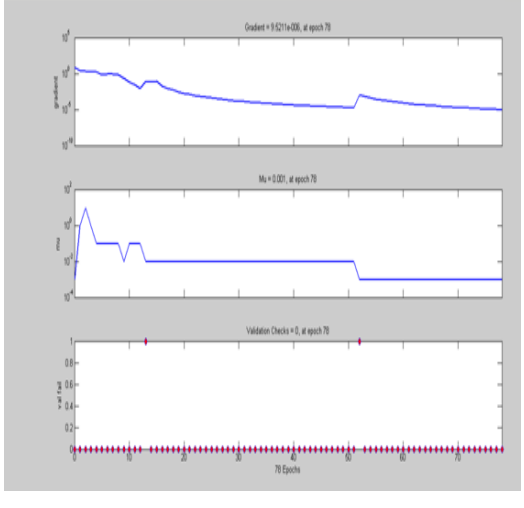

Figure5: Training states of input dataset

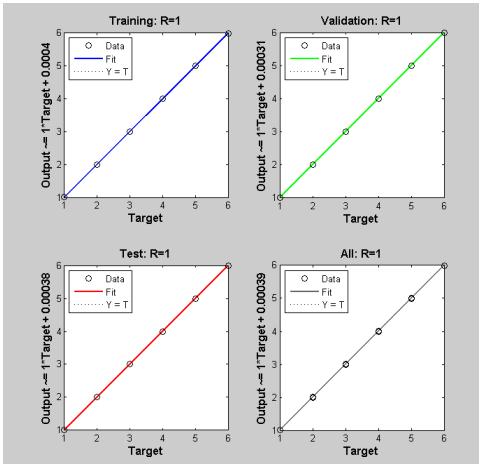

Fig 6 regression output dataset

After training phase we simulate the target data and test it with 100 question and out of testing 100 questions we get 86% right output. Which is shown in table 6.2

Table 3: Results from BPANN

	bag-of-words
Algorithm	Coarse Grain(in %)
BP-FFANN	86

REFERENCES

[1] Arun D Panicker, Athira U,Sreesha Venkitakrishnan, *"Question Classification using Machine Learning Approaches"* IJCA Journal, Volume 48 - Number 13, 2012.

[2] João Pedro Carlos Gomes da Silva "QA+ML@Wikipedia&Google", Ph. D Thesis, Departamento de Engenharia Informática, Instituto Superior Técnico, 2009.

[3] Pan, Y., Tang, Y., Lin, L., & Luo, Y "*Question classification with semantic tree kernel"*. In *Sigir '08: Proceedings of the 31st annual international acm sigir conference on research and development in information retrieval (pp. 837–838). New York, NY, USA: ACM.*, 2008.

[4] Petrov, S., & Klein, D. *Improved inference for unlexicalized parsing.* In Human language technologies 2007: The conference of the north american chapter of the association for computational linguistics;proceedings of the main conference (pp. 404–411). Rochester, New York: Association for Computational, April, 2007.

[5] Moschitti, A., & Basili, R. "*A tree kernel approach to question and answer classification in question answering systems"* In *Lrec* , 2006, *(p. 22-28).*

[6] Dell Zhang & Wee Sun Lee *"Question Classification using Support Vector Machines" by* Proceedings of the 26th annual international ACM, 2003

[7] Kadri Hacioglu & Wayne Ward "*Question Classification with Support Vector Machine and Error Correction Codes*" by,in NAACL-Short '03 Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003--short papers - Volume 2

[8] Vanitha Guda, Suresh Kumar Sanampudi "Approaches for Question Answering system*"*, IJEST, Vol. 3 No. 2 Feb 2011.

[9] Zhiping Zheng, "Answer bus question answering system" School of Information University of MichiganAnn Arbor, MI 48109.

[10] Changki Lee, Ji-Hyun Wang, Hyeon-Jin Kim, Myung-Gil Jang "Extracting Template for Knowledge-based Question-Answering Using Conditional Random Fields" (2004)

[11] Dina Demner-Fushman and Jimmy Lin, "*Answer Extraction, Semantic Clustering, and Extractive Summarization for Clinical Question Answering*" Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL, pages 841–848, Sydney, July 2006.

[12] Lotfi A. Zadeh,"*From Search Engines to Question Answering Systems – The Problems of World Knowledge"*, Relevance, Deduction and Precisiation" Computational Intelligence, Theory and Applications Volume 38, 2006, pp 163.

[13] Kadri Hacioglu,Wayne Ward, "*Question classification with support Vector Machine and Error correcting Codes*" (2003)

[14] Rainer Osswald,"*Constructing Lexical Semantic Hierarchies from Binary Semantic Features*"by (2007)

[15] Thomas G. Dietterich,"*Machine-Learning Research Four Current Directions*" by *(1997)*

[16] Aysun Guran,;_Nilgun GULER BAYAZIT, Mustafa Zahid GURBUZ1 "E_cient feature integration with Wikipedia-based semantic feature extraction for Turkish text summarization**",** Turkish Journal of Electrical Engineering & Computer Sciences, Turk J Elec Eng & Comp Sci,21: 1411 - 1425 doi:10.3906/elk-1201-15, (2013)

[17] Rish. T.J. Watson Research Center*, "An empirical study of naïve bayes classifier"* (2011)

[18] S. M. Kamruzzama, S. M. Kamruzzaman "*Pattern Classification using Simplified Neural Networks with Pruning Algorithm",ictm 2005.*

[19] Zhiheng Huang,Marcus Thint,Zengchang Qin *"Question Classification using HeadWords and their Hypernyms"* Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Honolulu, Association for Computational Linguistics, October 2008.

[20] Minh Le Nguyen,Thanh Tri Nguyen,Akira Shimazu "*Subtree Mining for Question Classification Problem",* IJCAI-07, 2007

[21] R. Rojas "*The Backpropagation Algorithm*", Neural Networks, Springer-Verlag, Berlin, 1996.

[22] David Nadeau, Satoshi Sekine*"A survey of named entity recognition and classification"*National Research Council Canada / New York University, 2007.

About authors

Rishika Yadav – B. E. (CSE) from MPCCET, currently works as asst prof. in SSCET also pursuing M. E. (CTA) from SSCET. Published 2 paper in national level paper presentation.

Megha Mishra – B. E. (CSE), M. E. (CTA), Ph. D. (pursuing) currently works as sr. asst. prof in SSCET, she published more than 10 paper in national and international journals.