AUTOMATIC QUESTION TAGGING WITH DEEP NEURAL NETWORKS

Silpa Rajeev S
Six semester MCA student
Sree Narayana Institute of Technology,
Kollam, Kerala,

Redhya M.
Assistant Professor,
Sree Narayana Institute of Technology,
Kollam, Kerala

T. Mahalakshmi
Principal,
Sree Narayana Institute of Technology,
Kollam, Kerala

Abstract: In recent years, computerized adaptive testing (CAT) has gained popularity as an important means to evaluate students' ability. Assigning tags to test questions is crucial in CAT. Manual tagging is widely used for constructing question banks; however, this approach is time-consuming and might lead to consistency issues. Automatic question tagging, an alternative, has not been studied extensively. In this paper, we propose a position-based attention model and keywords-based model to automatically tag questions with knowledge units. With regard to multiple-choice questions, the proposed models employ mechanisms to capture useful information from keywords to enhance tagging performance. Unlike traditional machine learning-based tagging methods, our models utilize deep neural networks to represent questions using contextual information. The experimental results show that our proposed models outperform some traditional classification and topic methods by a large margin on an English question bank dataset.

Key words : Deep neural networks, indexing methods, text analysis.

I INTRODUCTION

Computerized testing has become a popular method of assessing students with the aim of measuring their ability, adjusting their learning approach, and recommending materials to help them improve. Adaptively adjusting the approach and providing recommendations according to students individual level are crucial to improving the learning efficiency, this paradigm is called computerized adaptive testing (CAT)[1…4].This approach requires a well-structured question bank, which is a collection of question items stored in a database[5].Tagging is a simple and efficient method to organize resources[6]. As knowledge units are used as tags, we can utilize tagging technology to associate questions with them. Thus, questions will be well-organized for realizing various CAT functions.

Manual tagging, semi-automatic tagging, and automatic tagging are the three types of tagging methods .Manual tagging is the most commonly used method for organizing questions in the industry. However, manual tagging suffers from some limitations. First, manually tagging questions with knowledge units requires that the taggers be experts in that subject. Second, a question bank usually contains a large number of questions, which are updated constantly; this makes manual tagging expensive in terms of the time taken and associated cost. Semi-automatic tagging analyses content and returns tags that need to be further processed by users, making human help mandatory. Automatic tagging processes content without human intervention, resulting in more standard and consistent results at lower costs[7].

II COMPUTATINAL METHODS

The Question Tagging Application is based on two novel approaches – Position Based Attribute Modeling (PBAM) and Knowledge based Modeling.(KBAM). The methods relay on the basic algorithms used with Natural Language Processing. The tags obtained from NLP extraction are under gone tothese two models and evaluated separately. The observations and finding are followed here.

The conventional TFIDF based tagging is also compared with the novel methods[8].

Method 1

PBAM : The position of answer words in question are affecting the tag detection process in a most effective way. See the tag words has relevance in answers and questions that leads to title tags in a subject. The different levels of attention to words of an answer as opposed to that of a query. Hence, it is also called as the attention model. Specifically, the project impose different weights on different vectors of words. This idea was previously explored in Machine Reading Comprehension to find an answer to a query from a document

.In contrast to how they generate attention based on the interrelation between a query and document[9], we generate attention with the help of prior knowledge. That is the positions of answers. This prior knowledge can be obtained since blanks in a query reflect the positions of answers, a feature of multiple-choice questions.

In the experiment, first filter out punctuation and convert the words into lower case. Then, we split the question into tokens and transform them into one-hot representations. As a result, a word x is represented as an one-hot vector x. The entire question is represented as a matrix.

$[x_1; x_2; \_\ \_\ \_ ; x_{n-1}; x_n]$

$$p_i = \begin{cases} P_A, & x_i \text{ is a word of answers,} \\ P_N, & x_i \text{ is a normal word,} \end{cases}$$

where PA and PN indicate whether a word is part of an answer, and the initial attention weights for words. PA is greater than PN. We denote the position attention information of the question as a vector

$[p_1; p_2; \_\ \_\ \_ ; p_{n1}; p_n]$.

Given a question, the question matrix and position attention vector correspond. They are input to our neural network. Since a multiple-choice question is quite short, some methods of Natural Language Processing (NLP) preprocessing cannot be applied, such as stop words filtering and word stemming.

Method 2

KBAM : Short text provides little information because it contains a very limited number of words. From some related work, we were motivated to use keywords in the content to boost the tagging performance. In this scenario, the keywords of a question comprise answers based on the intuition that answers better reflect what knowledge units a question examines. Based on this, the work propose a keywords-based model to exploit information of answers in another way. Specifically, we pad more answer words and trained with a Knowledge database. The knowedge database consists of knowledge area and a lot of subknowledge words in it[10].

III DATA SETS

We have take a number of pdf documents that contain multiple choice questions. The questions are in a particular format. The question contain, question choices and correct answers. The questions are created in document format and converted to file. Such 100 files are tested and model and get the predicted tags.

IV EXISTING SYSTEM

Most of the question tagging is either manual or semi automatic in nature. In manual ,a human brain is to be used to find the tag relevance. In semi automatic, the relevance may be calculated with extra efforts.

Disadvantages of the existing system

- Chance of errors in accessing the question content
- The meaning aspect is less considered.
- The machine learning methodologies are not implied

V PROPOSED SYSTEM

The Proposed system uses two models Position-based attention model, keywords-based mode. The organize a question bank with the support of a knowledge map. The knowledge map can be established by knowledge units and their relationships from the corresponding syllabus using anontology technique.

The proposed automatic tagging methods provide questions with tags from a knowledge map.

Advantages of proposed system

- ○ Easy to evaluate subject distribution.
- ○ Multiple Tags are traced out with in the system.
- ○ Tag extraction to handle very short questions.

VI RESULTS AND DISCUSSIONS

Method 1

Table 1 gives the P

Parameters setup for CNN modelling.

VertexName (VertexType) nIn,nOut TotalParams ParamsShape Vertex Inputs

===

=====================================

input (InputVertex) -,- - - -

Output layer (OutputLayer) 65,2 132 W:{65,2}, b:{1,2} [input]

Total Parameters: 132

Trainable Parameters: 132

Frozen Parameters: 0

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.933	0.118	0.955	0.933	0.944	0.802	0.975	0.991	Y
	0.882	0.067	0.833	0.882	0.857	0.802	0.975	0.946	N
Weighted Avg.	0.919	0.104	0.921	0.919	0.920	0.802	0.975	0.978	

=== Confusion Matrix ===

a b <-- classified as

42 3 | a = Y

2 15 | b = N

Method 2

The knowledge levels are embedded to the CNN gateway.

The results are arrived as

Time taken to test model on test split: 0.02 seconds

=== Summary ===

Correctly Classified Instances	18	85.7143 %
Incorrectly Classified Instances	3	14.2857 %
Kappa statistic		0.6736
Mean absolute error		0.3085
Root mean squared error		0.3355
Relative absolute error		77.8094 %
Root relative squared error		77.889 %
Total Number of Instances		21

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.813	0.000	1.000	0.813	0.897	0.713	1.000	1.000	Y
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Weighted Avg.	0.857	0.045	0.911	0.857	0.866	0.713	1.000	1.000	

=== Confusion Matrix ===

a b <-- classified as

13 3 | a = Y

0 5 | b = N

The following figure show Architecture of Deep Neural Networks.

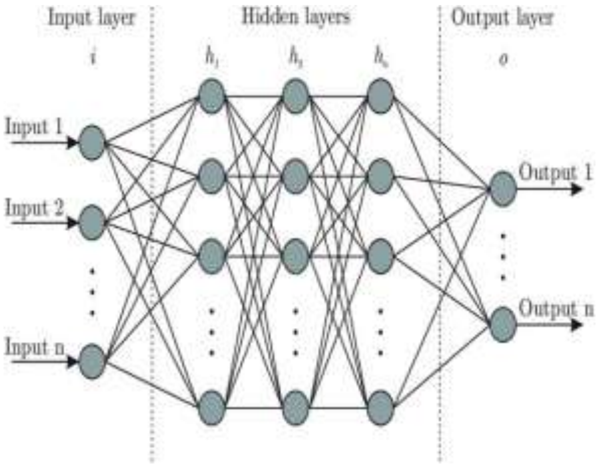


Fig: Deep Neural Networks

The accuracy, precision , recall and FRation are the measures obtained on the application of data from the two novel models. True-positives(TP),True-Negatives(TN),False-Positives(FP),False-Negatives(FN) are the four parameters for calculate Accuracy, Precision, Recall and F1 score.

Accuracy - High accuracy means this model is best. Accuracy rate is 0.803 which means model is approx. 80% accurate.

Accuracy = TP+TN/TP+FP+FN+TN

Precision – precision rate is 0.788 precision which is pretty good.

Precision = TP/TP+FP

Recall - Recall of 0.631 which is good for this model as it's above 0.5.

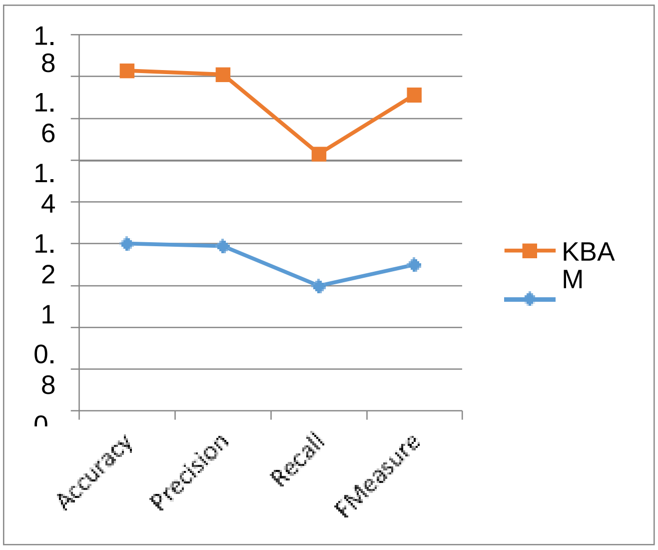
Recall = TP/TP+FN

F1 score - F1 Score is the weighted average of Precision and Recall. In this case, F1 score is 0.701 [11].

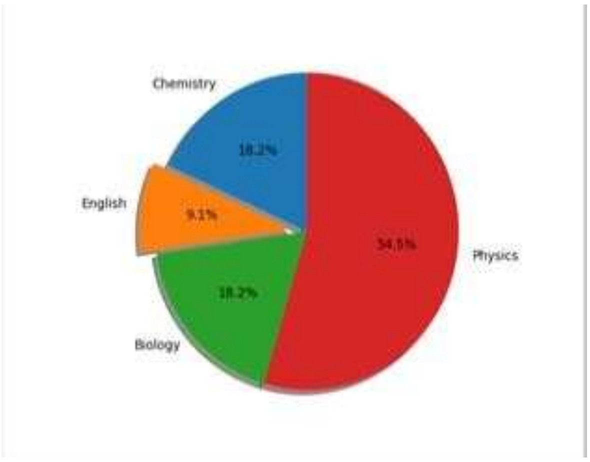
Comparision of Accuracy, Precision, recall and FMeasure in two models.

CNN Model	Accuracy	Precision	Recall	FMeasure
PABM	0.80	0.788	0.6	0.701
KBAM	0.83	0.823	0.63	0.81

Graphical Analysis of Performance Measures in two models are given below



The Percentage of Questions in each subject is evaluated as below.



VII CONCLUSION AND FUTURE SCOPE

The implementation of proposed DNN model using position based data and keyword based data is a novel idea in generating question tags. The systems generate meaningful tags better than the previous system. The study was aimed at automatic tagging from questions. The proposed model outperformed the previous models used for classification and tag extraction it is tested with questions bank from different subjects in English language also the performance of DNN is found to be better than the traditional neutral network system.
The implementation of Tag Generation using DNN was aimed to find the quality of question paper. The system is exclusively used for pdf based questions. In future, it can be changed to suit all type of documents like word and txt files.

VI. REFERENCES

[1] A.Al-Hmouz, J. Shen, R. Al-Hmouz, and J. Yan, "Modeling and simulation of an adaptive neuro-fuzzy inference system (anfis) for mobile learning," IEEE Trans. Learn. Technol., vol. 5, no. 3, pp. 226–237, 2012.

[2] P. Brusilovsky, "Knowledgetree: A distributed architecture for adaptive e-learning," in Proc. 13th Int. World Wide Web Conf. Alternate Track Papers & Posters. ACM, 2004, pp. 104–113.

[3] M. Mendicino, L. Razzaq, and N. T. Heffernan, "A comparison of traditional homework to computer-supported homework," J. Res.on Technol. in Edu., vol. 41, no. 3, pp. 331–359, 2009.

[4] J. D. Forbey and Y. S. Ben-Porath, "Computerized adaptive personality testing: a review and illustration with the mmpi-2 computerized adaptive version." Psychological Assessment, vol. 19, no. 1,p. 14, 2007.

[5] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs," Proc. IEEE,vol. 104, no. 1, pp. 11–33, 2016.

[6] H. L. Kim, A. Passant, J. G. Breslin, S. Scerri, and S. Decker,"Review and alignment of tag ontologies for semantically-linked data in collaborative tagging spaces," in Proc. IEEE Int. Conf.Semantic Comput. IEEE, 2008, pp. 315–322.

[7] M. Brut, F. Sedes, T. Jucan, R. Grigoras, and V. Charvillat, "An ontology-based modeling approach for developing a competencies-oriented collective intelligence," Learn. to Live in the Knowl. Soc., pp. 219–222, 2008.

[8] G. Salton, A. Wong, and C.-S. Yang, "A vector space model for automatic indexing," Commun. the ACM, vol. 18, no. 11, pp. 613–620, 1975.

[9] K. M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay,M. Suleyman, and P. Blunsom, "Teaching machines to read and comprehend," in Proc. Advances in Neural Inform. Process. Syst.,2015, pp. 1693–1701.

[10] Y. Cui, Z. Chen, S. Wei, S. Wang, T. Liu, and G. Hu, "Attentionover-attention neural networks for reading comprehension," ArXivPreprint ArXiv:1607.04423, 2016.

[11] J. L. Fleiss, "Measuring nominal scale agreement among manyraters." Psychological Bull., vol. 76, no. 5, p. 378, 1971.