



Overview of Question Categorisation Techniques

^{#1}Asha Bharat Sattane

¹asha.sattane@gmail.com

^{#1}ME Computer

Department of Computer Engineering

JSPM's Imperial College of Engineering & Research, Wagholi, Pune

ABSTRACT

Assessment is one of the essential engines which drive students for learning. The goal of assessment is to guide students about learning of necessary information. There should be appropriate categorization of questions to test student ability of using and applying the range of cognitive skills. To best of this study past research works have relied on matching questions against hand-crafted rules. However, rules require enormous effort to create and often suffer from being too specific. Recent research in question categorization is based on statistical approach to overcome these issues by employing machine learning techniques such as Support Vector Machine and Artificial Neural Network. To categorize question in the areas of assessment systems, information retrieval and educational environment this paper presents an updated literature survey. This helps to design new methods or approaches for improvement in question categorization. Question categorization involving other languages besides English has also been examined.

Keywords— Question categorization, Machine Learning, Cognitive level, Assessment, Support Vector Machine, Bloom's taxonomy.

ARTICLE INFO

Article History

Received :28th December 2015

Received in revised form :

31st December 2015

Accepted :5th January , 2016

Published online :

6th January 2016

I. INTRODUCTION

Education is a process of modelling learner's behaviour. Planning, teaching and assessment are the important stages of education which aims to make learner capable of decision making, good reasoning and critical thinking. Assessment is the crucial step of determining learner's conceptual development. It helps to build cognitive skills of learner. Further, critical thinking ability developed through using the higher-level thinking [1] skills of Bloom's Taxonomy. Zoller and Tsapalis [2] defined higher order cognitive skills (HOCS) items. HOCS helps learner to connect past learning with new learning, create new learning pathways, strengthen the existing pathways and consolidate new learning by storing in memory for future retrieval.

Early theories in learning process have identified many types of assessment. However, there is no cross check made with learning outcome. This leaves a question in teacher's mind such as "Are we teaching what we think we are teaching?" and "Are students learning what we think they are learning?"

Solution of above problem is question categorization based on Bloom's taxonomy. This paper consists of following sections. I. Glimpse of Bloom's Taxonomy II Complete Review III Conclusion

II. GLIMPSE OF BLOOM'S TAXONOMY

Benjamin Bloom [3] invented the taxonomy of educational objectives in 1956, called Bloom's Taxonomy, who is an educational psychologist at the

University of Chicago. The revised Bloom taxonomy is invented in 2001 [5];

A challenging task for learner is to create something new by their own efforts can be achieved by designing and assessing process, based on Bloom's Taxonomy. Bloom's was chosen because it is : i) widely recognizable and familiar to many academics, ii) generic and applicable across a wide range of subjects, iii) easy to apply to a range of question types owing to its simple structure [4]. Bloom's Taxonomy is a model of categorisation or a way to categorise thinking in six cognitive levels.

According to Bloom's Taxonomy, thinking skills is broken into six levels such as knowledge, comprehension, application, analysis, synthesis and evaluation. Anderson and Krathwohl in 2001 revised bloom's categorisation into Remembering, Understanding, Applying, Analysing, Evaluating, Creating [5].

A. Remembering

Remembering is when memory is used to produce definitions, facts, or lists, or recite or retrieve material from long-term memory.

B. Understanding

Determining and constructing the meaning of instructional messages, oral, written, and graphic communication, different materials. Activities included such as interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining.

C. Applying

Using procedures on situations where learned material is applied.

D. Analysing

Breaking concepts and components into smaller units, determining how the units relate to one another and to an overall structure.

E. Evaluating

Composing judgments based on criteria and standards. This generates specific reports and recommendations to demonstrate the process of evaluation.

F. Creating

Putting elements together to form a coherent or functional whole product. This is the difficult level as it reorganises elements to create new pattern.

Anderson taxonomy model is depicted in Fig. 1 below.

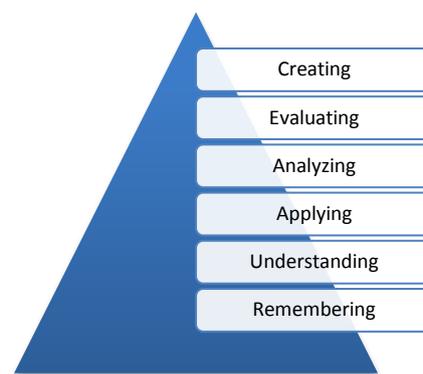


Figure 1. Bloom's revised taxonomy cognitive domain levels

In this revised Bloom's taxonomy cognitive levels have relationship among each other. Higher the level bigger will be complexity. Thus, assessment process starts from lower to higher level to encourage the learner for development of higher cognitive skills.

III. COMPLETE REVIEW

Question categorisation is mostly used in examination systems. Primary goal of categorising questions is to assign a label or unit to determine cognitive level of learner. Thus Bloom's taxonomy helps to categorize the question to measure cognitive skill of learner [6].

Some studies used both rule-based and learning based approaches together for categorising questions. In the study of Silva [8], which is one of the most successful works on question categorisation, first match the question with some pre-defined rules and then use the matched rules as features in the learning-based classifier. The same approach is used in the work by Huang [7]. This leaves a question behind setting the rules process for categorisation.

Zhang and Lee's question categorisation system is based on Support Vector Machine. They proposed use of special kernel function called tree kernel to enable the SVM to take advantage of the syntactic structures of questions. And described how the tree kernel can be computed efficiently by dynamic programming [9].

Krishnan et al. [10] used a short (typically one to three words) subsequence of question tokens or better known as informer span as features for question categorisation. To extract features it used a sequential graphical model with a novel feature representation derived from the parse tree of the question.

Author [11] concluded that, use of tokens in the work [10] could generate noisy information. To address the issues, it has been proposed head word feature and present two approaches to augment semantic features of such head words using WordNet [11].

For fact based question categorisation [12] has come up with analysis of statistical question categorisation. SVM classifier was performed across several data sets to test the performance of syntactic and semantic features.

In above reviews categorisation focuses on factual questions only, author [13] emphasises how SVM can be used in open ended questions.

All above work more concentrated on Question Answering (QA) and Information Retrieval (IR) system. Moreover question categorisation is also focused in educational environment in e-learning domain.

In the work of [14] Artificial Neural Network is used to categorise the question for e-learning systems. Multiple Types of question datasets is used by assigning the difficulty level to each question. Five steps for feature selection is used as query-text relevance, mean term frequency, length of Q&A, term frequency and distribution of Q & A in text. Lacking of semantic analysis and working on small data set are the concluded limitations by author.

Based on Bloom's taxonomy, [15] have used linear Support Vector Machines to categorise questions. For the effectiveness of categorisation author did not considered syntactic structure and semantic knowledge of questions.

Rule based classifier can also be used in categorising questions in accordance to Bloom taxonomy and the work by [16] obtained the result. To categorise the question in particular order author identified syntactic structure of question and used POS tagging and regular expressions to develop patterns and rule. Even results are obtained there is a need of large datasets and lots of rules.

In some research work no classifier methodology is also used. Work of [17] proposed keyword weight technique for categorisation based on Bloom's taxonomy. This technique can result more accuracy in case of overlapping of question in more than one cognitive level but applicable only on short open ended type of question.

The work by Nazlia Omar et al. [18] is similar to the work by [17] weighting techniques to categorise based on Bloom's taxonomy. However this method requires human intervention in determining the weight of particular question. and it is quite possible that different academicians or instructors may provide different weight for the same verb keyword. Besides

that, the rule based model needs to learn lots of patterns.

Accuracy observation details are shown in Table 1.

TABLE I
SUMMARY

#	Author	Limitation
1	Li and Silva [7] & [8]	Need to work on criteria of setting the rules
2	D.Zhang [9]	Only surface text Features are considered
3	V. Krishnan [10]	Noisy information generation after categorisation
4	Z. Huang [11]	Compact feature sets are used
5	D. Metzler [12]	Limited to factual questions only
6	J. Bullington [13]	Limited to open ended question only
7	T. Fei [14]	Small data sets are used
8	A. A. Yahya [15]	Not considered syntactic structure and semantic knowledge of questions
9	S. S. Haris [16]	Large data set required with lots of pre-defined rules
10	W. C. Chang [17]	Applicable to only short simple open ended questions only.
11	18	Large amount of human intervention is required

IV. CONCLUSION

In this paper, overview of detail review is presented for question categorisation. It is observed that most of

the research has considered only surface features instead of deep syntactic and semantic features.

A conclusion is prepared that categorisation is not limited to semantic feature extraction only. However, it is observed that dynamic approach must be used for feature extraction, feature reduction and selection of classifiers for more accuracy in question categorisation. Usage of different domain also considered while choosing the dynamic selection approach.

REFERENCES

- [1] <http://www.ascd.org/publications/educational-leadership/mar97/vol54/num06/Seven-Strategies-That-Encourage-Neural-Branching.aspx>
- [2] Zoller, U. & Tsaparlis, G., "Higher and lower-order cognitive skills: The case of chemistry." *Research in Science Education*, 27, 1997, 117- 130.
- [3] Bloom (Ed.), "Taxonomy of Educational Objectives: The Classification of Educational Goals"; pp. 201-207; B. S. Susan Fauer Company, Inc. 1956.
- [4] Karl O. Jones¹, Janice Harland², Juliet M.V. Reid and Rebecca Bartlett, "Relationship between examination questions and Bloom's taxonomy," 39th ASEE/IEEE Frontiers in Education Conference, 2009.
- [5] Anderson, L.W" & Krathwohl (Eds.). "A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives," New York: Longman, 2001.
- [6] N. Yusof and C. J. Hui, "Determination of Bloom's cognitive level of question items using artificial neural network," *Proc. of 10th International Conference on Intelligent Systems Design and Applications (ISDA)*, IEEE; doi:10.1109/ISDA.2010.5687152, 2010.
- [7] Li Xin, HUANG Xuan-Jing, and WU Li-de. "Question classification using multiple classifiers", In *Proceedings of the 5th Workshop on Asian Language Resources and First Symposium on Asian Language Resources Network*, 2005.
- [8] J Silva, L Coheur, Ana Mendes, and Andreas Wichert. "From symbolic to subsymbolic Information in question classification", *Artificial Intelligence Review*, 35(2): 137 - 154, February 2011.
- [9] D. Zhang, W. S. Lee, "Question classification using support vector machines," *Proc. Of 26th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 26-32, 2003; doi:10.1145/860435.860443
- [10] V. Krishnan, S. Das, and S. Chakrabarti, "Enhanced Answer Type Inference from Questions using Sequential Models," *Proc. of Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 2005.
- [11] Z. Huang, M. Thint, Z. Qin, "Question classification using head words and their hypernyms," *Proc. of the Conference on Empirical Methods in Natural Language Processing*, ACM, pp. 927-936, 2008.
- [12] D. Metzler, W. B. Croft, "Analysis of Statistical Question Classification for Factbased Questions," *Journal of Information Retrieval*, vol. 8, no. 3, 2005, pp. 481-504.
- [13] J. Bullington, L. Endres, M. Rahman, "Open ended question classification using support vector Machines," MAICS, 2007.
- [14] T. Fei, W. J. Heng, K. C. Toh, and T. Qi, "Question Classification for E-learning by Artificial Neural Network", *Information, Communications and Signal Processing*, 2003 and Fourth Pacific Rim Conference on Multimedia. *Proceedings of the 2003 Joint Conference of the Fourth International Conference*, 2003, pp. 1757-1761.
- [15] A. A. Yahya and A. Osman, "Automatic classification of questions into Bloom's cognitive levels using support vector machines," *Proc. The International Arab Conference on Information Technology*, Naif Arab University for Security Science (NAUSS), pp. 1-6, 2011.
- [16] S. S. Haris, N. Omar, "A rule-based approach in Bloom's Taxonomy question classification through natural language processing," *Proc. of 7th International Conference on Computing and Convergence Technology (ICCCT)*, 2012, pp. 410-414, Dec. 3-5, 2012.
- [17] W. C. Chang and M. S. Chung, "Automatic applying Bloom's taxonomy to classify and analysis the cognition level of English question items," *Proc. Joint Conferences on Pervasive Computing (JCPC)*, pp. 727 - 734, 2009.
- [18] Nazlia Omar , Syahidah Sufi Haris , Rosilah Hassan , Haslina Arshad , Masura Rahmat , Noor Faridatul Ainun Zainal & Rozli Zulkifli, "Automated analysis of exam questions according to bloom's taxonomy", Published by Elsevier Ltd. Selection and/or peer reviewed under responsibility of the UKM Teaching and Learning Congress, 2011.