Comparative Study of Different Approaches for Measuring Difficulty Level of Question

Ayesha Pathan, Dr. Pravin Futane

Department of Computer Engineering, PCCOE, Pune, Maharashtra, India

ABSTRACT

Semantics-based information representations such as ontologies are found to be very useful in repeatedly generating important factual questions. Formative the difficulty-level Of these system generated questions is helpful to successfully make use of them in various learning and specialized applications. The accessible approaches for result the difficulty-level of factual questions are very simple and are limited to a few basic principles. We suggest a new tactic for this problem by considering an edifying theory called Item Response Theory (IRT).

In the IRT, facts skill of end users (learners) are considered for assigning difficulty levels, because of the assumptions that a given question is apparent differently by learners of various proficiencies. We have done a detailed study on the features/factors of a question statement which could perhaps determine its difficulty-level for three learner categories (experts, intermediates, and easy).

KEYWORDS*: Difficulty level estimation, Item response theory, Question generation, Automatic Quiz Generation, Difficulty Ranking, Semantic Similarity*

I. INTRODUCTION

Question difficulty is important in test creation and question analysis. Nowadays, it is widely recognized that test construction is really time-consuming for teachers. The use of Computer Assisted Assessment reduces considerably the time spent By teachers by Constructing examinations paper.

There are many types of assessment or 'testing' to access student's learning curves.

However, written examination is the most common approach used by any higher education institutions for students' assessment. Question is an element that is intertwined with the examination. Questions raised in the paper plays an important role in e orts to test the students' overall cognitive levels held each semester. Effective style of questioning as described by Swart is always an issue to help students attend to the desired learning outcome. Furthermore, to make it effective, balancing between lower and higher-level question is a must Swart Bloom's Taxonomy, created by Bloom has been widely accepted as a guideline in designing reasonable examination questions belonging to various cognitive levels. The hierarchical models of Bloom's are widely used in education fields constructing questions to ensure balancing and student cognitive mastery.

II. LITERATURE REVIEW

1. Hochschule fur Technik Stuttgart Schellingstr (2017):

Empirically verify that Bloom's taxonomy ,a standard tool for difficulty estimation during question creation. Question difficulty estimates guide test creation, but are too costly for small-scale testing. We empirically verify that Bloom's Taxonomy, a standard tool for difficulty estimation during question creation, reliably predicts question difficulty observed after testing in a short-answer corpus. We also find that difficulty can be approximated by the amount of variation in student answers, which can be computed before grading. We showthat question difficulty and its approximations are useful for automated grading, allowing us to identify the optimal feature set for grading each question even in an unseen-question setting. Testing is a core component of teaching, and many tasks in NLP for education are concerned with creating good questions and correctly grading the answers. We look at how to estimate question difficulty from question wording as a link between the two tasks. From a test creation point of view, knowing question difficulty levels is imperative: Too many easy questions, and the test will be unable to distinguish between the more able test-takers, who all achieve equally good results. Too many hard questions, and only the most able test-takers will be clearly distinguishable from the (low performance result.)

2. Neung Viriyadamrongkij and Twittie Senivongse (2017):

Online inquiry communities such as Question-Answer Communities (QAC) have captured interest of online users since they can share and search for any information from any place in the world. The number of questions and answers submitted to a popular community can increase rapidly, and that can make it difficult for users who look for the \right" questions to answer. That is, from the view of knowledgeable experienced users, they tend to look for hard challenging questions as an opportunity to share their knowledge and to build respect with the community. Hence it is desirable to distinguish difficult questions from easy ones. Current researches estimate complexity of questions based on the analysis of the features of the QAC without considering the contents of the questions. This paper presents a method to measure question difficulty levels based directly on the question contents. In particular, we analyze the difficulty of terms that appear in a JavaScript-related question, based on the proposed JavaScript concept hierarchy. In an evaluation of the performance of the question difficulty estimation, our concept based measure givessimilar performance to that of the existing measure based on the features of the QAC.

3. Sasitorn Nuthong, Suntorn Witosurapot (2017):

This Automatic Quiz Generation system is utterly handy for reducing teachers' workloads in quiz creation. Nevertheless, by exploiting a coarse-granular concern inside difficulty ranking mechanism, only a few number of automatic generated quizzes can be obtained. In order to increase the number of usable quizzes, we suggest how a 5-level difficulty ranking score using a hybrid similarity measurement approach together with property filtering of the key data can be potential for serving this propose. Based on experiment results, our proposed similarity measure outperforms three other candidates. Enabling users with finer options of making sensible quiz generation. Hence, this mechanism can be regarded as a synergistic technology for improving teachers' quality of life for the future.

4. Surbhi Choudhary, Abdul Rais Abdul Waheed, Shrutika Gawandi, Kavita Joshi (2015):

In this modern world e-book has become a basic requirement for the candidates to appear and prepare for their competitive exams within college premises. In this paper we are proposing a replica system for smart question paper generation of universities. The mechanism behind this system is that many random question papers are generated along with the complexity level of the questions in terms of percentage.[4] After generation that particular question is then mailed to the respective university. In this system administration of the database inputs set of question paper with an option of check box to tick the accurate answer. More ever weightage of the particular question in terms of marks and hours and the complexity of the question is determined. After this course whole question paper all along with the weightage is stored in the database. Inside order to make question paper for 100 marks, admin sets all the weightage and difficulties to solve the problem. When the difficulty and weightage is specified a pre doc le as per selected format will be downloaded to the admin and an electronic mail will be triggered.

Variety of difficulty may vary from easy, Medium and hard.

5. Pawel Jurczyk, Eugene Agichtein (2007):

Question-Answer portals such as Naver and Yahoo! Answers are rapidly fetching rich sources of information on a lot of topics which are not well served by general web search engines. Unluckily, the quality of the submitted answers is uneven, ranging from excellent detailed answers to snappy and insulting remarks or even advertisements for commercial content. Furthermore, user feed-back for many topics is sparse, and can be insufficient to reliably identify good answers from the bad ones. Hence, estimating the ability of users is a critical task for this rising domain, with potential applications to answer ranking, spam detection, and incentive mechanism design. We present an analysis of the link structure of a general-purpose question answering community to discover authoritative users, and capable experimental results over a dataset of more than 3 million answers from a popular community QA site. We also explain structural differences between question topics that correlate with the success of link analysis for authority discovery.

Existing system:

There are very few such system previous available like a method was proposed based on Measuring Difficulty Levels of JavaScript online Questions in Question-Answer Community Based on Concept Hierarchy.

III. PROPOSE SYSTEM

We propose a system to automate paper setting and to measure the difficulty levels of questions in an Engineering subjects by analyzing the Terminology that appears in the questions.
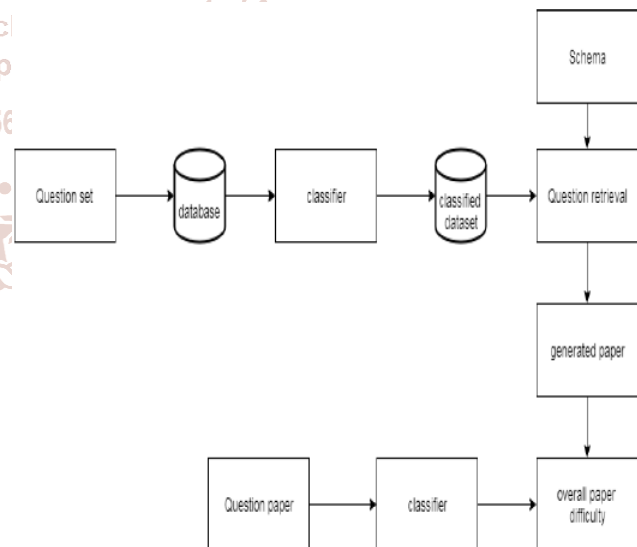
IV. PROPOSE SYSTEM ARCHITECTURE

Figure4.1. System Architecture

V. CONCLUSION AND FUTURE SCOPE

Establishing mechanisms to control and predict the difficulty of assessment questions is clearly a big gap in existing question generation literature. Our contributions have covered the deeper aspects of the problem, and proposed strategies, that exploit ontologies and associated measures, to provide a better difficulty-level predicting model, that can address this gap. We developed the difficulty level model (DLM) by introducing three learner specific logistic regression models for predicting the difficulty of a given question for three categories of learners. The output of these three models was then interpreted using the Item Response

Theory to assign high, medium or low difficulty level. The overall performance of the DLM and the individual performance of the three regression models based on cross-validation were reported and they are found to be satisfactory. Comparison with the state-of-the-art method shows an improvement of 8.5 difficulty-levels of benchmark questions. The model proposed in this paper for predicting the difficulty-level of questions is limited to A Box based factual questions. It would be interesting to extend this model to questions that are generated using the T Box-based approaches. However, the challenges to be ad-dressed would be much more, since, in the T Box-based methods, we have to deal with many complex restriction types (unlike in the case of A Box-based methods) and their influence on the difficulty-level of the question framed out of them needs a detailed investigation. For establishing the propositions and techniques stated in this paper, we have implemented a system which demonstrates the feasibility of the methods on medium sized ontologies. It would be interesting to investigate the working of the system on large ontologies.

REFERENCES

[1] UlrikePad o \Question Difficulty How to Estimate Without Norming, How to Use for Automated Grading" Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications, Copenhagen, Denmark, September 8, 2017.

[2] Neung Viriyadamrongkij and Twittie Senivongse \Measuring Difficulty Levels of JavaScript Questions in Question-Answer Community Based on Concept Hierarchy"978-1-5090-4834-2/17/ 31.00 c 2017 IEEE

[3] Itziar Aldabe Edurne Martinez \ArikIturri: an Automatic Question Generator Based on Corpora and NLP Techniques" Conference Paper in Lecture Notes in Computer Science June 2006

[4] Sarah K. \Two Methods for Measuring Question Difficulty and Discrimination in Incomplete Crowd sourced Data "Association for the Advancement of Artificial Intelligence 2013

[5] Asma Ben Abacha, Marcos Da Silveira, and Cedric Pruski. Medical ontology validation through question answering. In AIME, pages 196{205, 2013. 10.1007/978-3-642-38326-730.

[6] Maha Al-Yahya. Ontology-based multiple choice question generation. The Scientific World Journal, Vol 2014, page 9, ID: 10.1155/2014/274949, 2014.

[7] T. Alsubait, B. Parsia, and U. Sattler. A similarity based theory of controlling mcq difficulty. In eLearning and e-Technologies in Education (ICEEE), 2013 Second International Conference on, pages 283{ 288, Sept 2013.

[8] T. Alsubait, B. Parsia, and U. Sattler. Generating multiple choice questions from ontologies: Lessons learnt. In Proceedings of the 11th Inter-national Workshop on OWL: Experiences and Directions (OWLED 2014), volume 1265, pages 73{84, Oct 2014.

[9] Tahani Alsubait. Ontology-based multiple-choice question generation. PhD thesis, School of Computer Science, The University of Manchester, 2015.

[10] Tahani Alsubait, Bijan Parsia, and Ulrike Sattler. Mining ontologies for Analogy questions: A similarity-based approach. Volume 849 of CEUR Workshop Proceedings. OWL Experiences and Directions, 2012.

[11] Tahani Alsubait, Bijan Parsia, and Ulrike Sattler. Generating multiple choice questions from ontologies: Lessons learnt. Volume1265 of CEUR Workshop Proceedings.OWL Experiences and Directions, 2014.

[12] Xinming An and Yiu-Fai Yung. Item response theory: What it is and how you can use the irt procedure to apply it. In SAS Global Forum, 2014.