Survey On: Theory Online Examination with Short Text Matching

[1]Pooja Kudi
[2]Tejaswini Dhatrak
[3]Kavita Daware
[4]Amitkumar Manekar

Abstract— In traditional Online Examination System, only objective type questions are assessed and according to that marks are given to the student. However, this technique lacks the capability of evaluating descriptive answers. In university examinations, there are many types of question included for evaluation of the students. Therefore, the automated system must be capable of evaluating the descriptive answers. The online examination system checks the student answer by matching the answer with predefined set of answer. The predefined answers are saved on the server and evaluation is done automatically using the automatic assessment tools. Here the machine learning approach is used to solve this problem using text mining. Measuring the similarity between, sentences, words, documents and paragraphs is an important component in various tasks such as text summarization, information retrieval, automatic essay scoring, document clustering, and machine translation and word-sense disambiguation. In this system JSON is used for transferring data between web application and server, serving as an alternative to XML.

Index Terms— Text mining, automatic assessment tools, Machine learning, JSON, XML, IndusMarker.

———————————— ◆ ————————————

1 INTRODUCTION

Large number of Candidates attends the exams for that evaluation is required manual effort and in some cases student having the poor handwriting which is not clearly understandable by evaluator. Sometimes quality of evaluation may change according to mood of evaluator. The evaluation work is very lengthy and time consuming. So, to avoid such problems the automated examination environment is developed. Hence, the solution made for automating the work using computers. The novel approach is to describe an automated descriptive answer marking system that can be utilized to improve teaching and learning of the technical subjects. Question paper is prepared by the expert teacher in consultation with the technical staff with the system who assists in uploading the same in the desired format. Candidates appear for the test online within the exam centre. User authentication has been provided. On successful login, each student gets the question paper and his/her answer book. They are required to answer the questions by typing in the blank space provided for the answer. The software facilitates saving of the typed work as per student convenience. In case of a system failure, the saved work is available to the student for continuation. The problems like document representation, classifier construction and classifier evaluation are solved by using machine learning approach. In Online Examination System *JSON* is used for transferring data between web application and server, serving as an alternative to *XML [9]*. Data mining consists of many techniques such as clustering, neural networks and classification and decision trees. Data mining is a process, before applying any technique preprocessing of data is required. All textual based information are stored, may be in computers or on web. Any computer or laptop can easily stored large amount of data because of advance hardware storage devices. Accumulating information is easy but finding related information on demand can be difficult.

2 RELATED WORK

Text Mining

Text mining is used to extract important information or data or pattern or knowledge from various sources which are in the unstructured form [6]. Unstructured-form means which is not in predefined form if the data or information is in an unstructured form then it is difficult to handle them. The main purpose of text mining is to find valuable information from natural language text. The text mining consists of two components.

A. TEXT MINING COMPONENTS

1. Text refining:
Text refining translates free-form text data into an intermediate form [3].

2. Knowledge distillation:
Knowledge distillations retrieve the knowledge or patterns from intermediate form [3]. Intermediate form can be converted into struc-

ture like graph and relational data representation. The IF is based on the documents where each entity of document is concept based. Each entity represents the concept or object in particular interested domain.

Stage-I: Pre-processing Text

Pre-processed text is easy to mining as compare to natural languages documents. So, before applying any text mining technique pre-processing of documents that are from various sources is an important task during text mining process.

Stage II- Text Mining Technique is applied.

At this stage to process the text one of selected algorithm is applied on answer text. The different algorithms are used such as clustering, summarization, classification, information extractions and visualizations.

Stage III - Analysis of Text

At this stage the outputs are analyzed for knowing the status of student's knowledge.

B. TEXT MINING TECHNIQUES

1) Summarization:
   Summarization is the process which can be apply on the single document or multiple documents. It summarize the whole document in such a way that meaning of data will not be changed but length of the of the document will be reduced.

2) Categorization
   It is a supervised technique which is based on the set of input and output. Number of classification techniques is used to categorize text such as Naive Bayesian classifier and Nearest Neighbour classifier.

3) Naive Bayesian Classifier
   If there is graph then child node is connected to the parent node so, there will be no other connections are possible. According to this technique all words of document are not dependent to each other.

4) Nearest Neighbour Classifier
   It calculated similarity factor of unknown document with all known documents. If K similar documents are considered then it is called as K nearest neighbour classifier.

5) Clustering
   This technique is differs from the categorization which is also used for group similar document. It divides same document into same cluster then cluster is divided into two categories: hierarchical clustering and portioning clustering.

6) Hierarchical Clustering
   The result of hierarchical clustering is stored on a single clustered tree. It can be divided into two categories: Bottom up hierarchical clustering method and top down hierarchical clustering method.

7) Bottom Up Hierarchical Clustering
   An individual cluster consider as single document. It checks similarity of the cluster it will be combined, this will be done repeatedly until it forms the single cluster.

8) Top Down Hierarchical Clustering Method
   Based on the similarity factor of single cluster it will be spiltted. In this technique work starts from single cluster.

- Information Extraction
   This information extraction software identifies relationships with key phrases in the text. Pattern matching is process for predefined sequences in a text. Firstly analyze unstructured text for information extraction. The information in the Natural language text document cannot be used for mining.

9) Visualization
   This method provides better and faster understandable Information. Which help us to mine the large documents collection.. It is purely text based Visual text mining or information visualization puts large textual sources in a visual hierarchy or map and it provides browsing capabilities in addition to simple searching. By using this users can differentiate the colors, relationships, distance etc. The collection of documents can be represented as a structured format using in-

dexing or vector space model etc.

C. XML AND JSON

For data exchanging in the context of web service applications two formats area used that is XML and JSON. XML is the de facto standard format. The data in XML format has to be parsed before being processed at client side and server side also. XML is quilt time and memory consuming, such as DOM needs to load the whole XML document into memory before processing the data. As compare with XML, JSON is a light-weight key-value style data exchanging format. As we know, the efficiency of mapping data between different data models is the key point to improve the performance of web service applications. A data model mapping approach, named as Dynamic Advanced Binding (DAB), is proposed for processing JSON data.

XML

XML stands for Extensible markup language which is derived from standard generalized markup language. XML is a Markup Language having format that contains set of rules for the encoding the documents which is readable for both human & machine. XML does not provide any data type so needs to be parsed into particular data type. No direct support for array also. XML supports Namespaces. XML can get support of objects through mixed use of attributes & elements. XML is document oriented and needs more effort for mapping. XML needs XML Document Object Model (DOM) implementation & with that additional code for mapping text back to the JavaScript objects. XML is a user-defined hierarchical data format. XML is a language used for creating user-defined markups to documents and encoding schemes. XML does not have predefined tag sets and each valid tag is defined by either a user or through another automated scheme.

JSON

JSON is derived from Javascript Object Notation. JSON is one type of text-based format or standard for interchanging data i.e. human readable. JSON syntax is lighter than XML as JSON has serialized format of data having less redundancy. JSON does not contain start and end tags. JSON is light – weighted in compare to XML as it has serialized format and so faster also. JSON supports data type including integer and strings, JSON also supports array. JSON has support of native objects. JSON does not have support for Namespaces. JSON does not support Comments. JSON is data oriented and can be mapped more easily. JSON uses only evel() for parsing i.e. for interpreting the JavaScript code & returns the result. It does not need any additional code for parsing.

1) AUTOMATIC ASSESSMENT TOOLS

Project Essay Grade (PEG)
PEG [12] is one of the earliest and longest-lived implementations of automated essay grading. The design approach for PEG is based on the concept of *"proxes"*,i.e. computer approximations or measures of *trins*, in*trins*ic variables of interest within the essay (what a human grader would look for but the computer can't directly measure) to simulate human rater grading.

Intelligent Essay Assessor (IEA)
IEA is based on the Latent Semantic Analysis (LSA) technique that was originally designed for indexing documents and text retrieval [13]. LSA represents documents and their word content in a large two-dimensional matrix semantic space.

Educational Testing Service (ETS I)
The technique uses lexical-semantic techniques to build a scoring system, based on small data sets [14]. It uses a domain- specific, concept-based lexicon and a concept grammar, both built from training data.

Electronic Essay Rater (E-Rater)
E-Rater uses a combination of statistical and NLP techniques to extract linguistic features from the essays to be graded [15]. Essays are evaluated against a benchmark set of human graded essays. E-Rater adopts a corpus-based approach to model building by using actual essay data to analyze the features of a sample of essay responses.

C -rater
C rater [1] is educational testing service which is technique used for the content scoring. This content scoring is based on model building which makes various answer model for candidate's short answer [7]. The ETS that is Education Testing Service used for checking short answer of candidate in range of 100 words approximately [8]. It uses analytical approach & rubrics item which specify correct & important terms which should be present in candidate's answer. The problem of different answer structure of candidate is solved using c-rater. It saves 0-12 hour work of humans.

a. Steps of c-rater model are [1]:
1 .Model building
2. C Rater automatically processes
3. Maintaining algorithm gold map
4. Apply student answer

b. Task performs by c-rater [1]:
4   Processes for spelling correction
5   Part of speech tagging & parsing
6   Parse tree passed through feature extractor

Bayesian Essay Test Scoring sYstem (BETSY)

BETSY [16] is a program that classifies text based on trained material. The goal of the system is to determine the most likely classification of an essay into a four point nominal scale (e.g. extensive, essential, partial, unsatisfactory) using a large set of features including both content and style specific issues.

Intelligent Essay Marking Systems (IEMS)

IEMS is based on the Pattern Indexing Neural Network [17]. The system can be used both as an assessment tools and for diagnostic and tutoring purposes in many content-based subjects. Students can be given immediate feedback and can learn where and why they had done well or not made the grade. Thus it can be embedded in an intelligent tutoring system that will help students to write better by grading papers fast and providing the feedback quickly.

Schema Extract Analyze and Report (SEAR)

This system provides an expandable, flexible method for the automated marking of the essay content and provides a method for the automated marking of the essay style too [18].

Automark

Automark [6] system is developed to provide short answer marking. The system incorporates a number of processing modules specifically aimed at providing tough marking in the face of errors in spelling, typing, syntax and semantics. Automark looks for specific content within answers, the content being specified in the form of a number of marking scheme designed [7]. Each template represents one form of a valid or a specifically invalid answer. Using an offline custom written configuration interface, templates are developed. The marking process progresses through a number of stages. First, the incoming answer text is preprocessed to standardize the input in terms of punctuation and spelling. Then, a sentence analyzer identifies the main syntactic constituents of the text and how they are related. The pattern-matching module searches for matches between the marking scheme templates and the syntactic constituents of the student text. The Automark system was able to correctly mark most of the answers containing spelling, syntax, and semantic errors

Paperless School free-text Marking Engine (PS-ME)

PS-ME [19] is designed as an integrated component of a Web-based Learning Management System. This system applies NLP techniques to assess student essays in order to reveal their level of competencies as for knowledge, understanding and evaluation. The student essay is submitted to the server, together with information about the task in order to identify the correct master texts for comparison. Each task is defined via a number of master texts that are relevant to the question to be answered.

IndusMarker

*IndusMarke*r exploits structure matching, i.e., matching a prespecified structure, developed via a purpose-built structure editor, with the content of the student's answer text. The examiner specifies the required structure of an answer in a simple purpose designed language. The language was initially called *QAL [7]* but later on it is redefined it as a sublanguage of *XML* and named it Question Answer Markup Language *(QAML)[6].*

3 ANALYSIS

Compared with XML, JSON can be parsed efficiently. Researchers have already shown that JSON has been able to take the place of XML as data exchanging format in web services. This was verified with simulating test on the performance of JSON and XML [20]. A JSON-based object serialization algorithm was proposed for presenting the navigator for any Java object and producing a collection of JSON expressions according to the navigator used Java reflection to realize mapping data between JSON and Java [21]. However, the performance of web services will be decreased with exploiting the technique to process complicated application data [22].

Demerits of Text mining:

The information needed is no where written.
To mine the text for information or knowledge no program is made in order to analyze the unstructured text directly.

Merits of Text mining:

1) Database cannot store large amount of information this problem has been solved through Text Mining
2) Using number of techniques of text mining like information extraction, relationship between different entities can be easily found from the document set.
3) Extracting the patterns from large amount of unstructured information easily Text mining is used; otherwise it would have been a great challenge.

4 CONCLUSION

This review paper is on Online examination with short text matching. The traditional technologies were based on XML. But as per analysis chapter JSON is more effective than XML. There are many techniques included in text mining which helps to extract the relevant information from large amount of data. Various automatic assessment tools are available which are discussed in related work.

REFERENCES

[1] C. Leacock and M. Chodorow, "C-Rater: Automated Scoring of Short-Answer Question," Computers and the Humanities, vol. 37,no. 4, pp. 389-405, 2003.

[2] Mahesh T R, Suresh M B, M Vinayababu "Text Mining: Advancements, Challenges and Future Directions" International Journalof Reviews in Computing 2009-2010.

[3] Vishal Gupta, Gurpreet S. Lehal "A Survey of Text Mining Techniques and Applications" Journal of Emerging Technologies in Web Intelligence, vo!.1, no. 1, august 2009.

[4] C.Cui and H.Ni. Optimized simulation on XML with JSON. Communication Technology (Chinese) 42(8) pages 108-111. 2009 08, 212

[5] T. Zhang, Q. Huang, Y. Mao and X. Gao. Algorithm of object serialization based on JSON. Computer Engineering and Applications(Chinese) 43(15), pages 98-100, 2007.

[6] Z. Li Transition from Java to JSON data with Java reflection. Academic journal of Chengde Petroleum College (Chinese) 3(12), pages 36-39 , 2010

[7] RaheelSiddiqi, Christopher J. Harrison, and RosheenaSiddiqi"Improving Teaching and Learning
Through Automated Short-Answer Marking" IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 3, NO. 3, JULY-SEPTEMBER 2010.

[8] Salvatore Valenti, Francesca Neri and Alessandro Cucchiarelli," An Overview of Current Research on Automated Essay Grading" DIIGA - Universita' Politecnicadelle Marche, Ancona, Italy.Journal of Information Technology Education Volume 2, 2003.

[9]   NurzhanNurseitov, Michael Paulson, Randall Reynolds, Clemente Izurieta "Comparison of JSON and XML Data Interchange Formats", A Case Study Department of Computer ScienceMontana State University –Bozeman, Montana, 59715, USA..

[10]   J.H. McMillan, Classroom Assessment: Principles and Practice forEffective Instruction, pp. 40 and 117. Allyn and Bacon, 1997.

[11]   Zhao Qiao-fang , Li Yong-fei"Research and Development of Online Examination System" North China Institute of Science and Technology Beijing, China,2012.

[12]   Hearst, M. (2000). The debate on automated essay grading. *IEEE Intelligent Systems, 15*(5), 22-37, IEEE CS Press.

[13]   Jerrams-Smith, J., Soh, V., & Callear D. (2001). Bridging gaps in computerized assessment of texts. *Proceedings of the International Conference on Advanced Learning Technologies, 139-140,* IEEE.

[14]   Whittington, D. & Hunt, H. (1999). Approaches to the computerized assessment of free text responses. In M. Danson (Ed.),*Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborough University, UK.

[15]   Burstein, J., Leacock, C., & Swartz, R. (2001). Automated evaluation of essay and short answers. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborough University, Loughborough, UK.

[16]   Rudner, L.M. & Liang, T. (2002). Automated essay scoring using Bayes' Theorem. *The Journal of Technology, Learning And Assessment,1*(2),3-21.

[17]   Ming, P.Y., Mikhailov, A.A., & Kuan, T.L. (2000). Intelligent essay marking system. In C. Cheers (Ed.), *Learners Together, Feb. 2000*, NgeeANN Polytechnic, Singapore.

[18]   Christie, J. R. (1999). Automated essay marking-for both style and content. In M. Danson (Ed.), *Proceedings of the Third Annual Computer Assisted Assessment Conference*, Loughborough University, Loughborough, UK.

[19]   Mason, O. & Grove-Stephenson, I. (2002). Automated free text marking with paperless school. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborough University, Loughborough, UK.

[20]   C.Cui and H.Ni. Optimized simulation on XML with JSON. Communication Technology (Chinese) 42(8) pages 108-111. 2009 08, 212

[21]   T. Zhang, Q. Huang, Y. Mao and X. Gao. Algorithm of object serialization based on JSON. Computer Engineering and Applications(Chinese) 43(15), pages 98-100, 2007.

[22]   Z. Li Transition from Java to JSON data with Java reflection. Academic journal of Chengde Petroleum College (Chinese) 3(12), pages 36-39 , 2010.