iCCE Mumbai | India | 2016

THE 24th INTERNATIONAL CONFERENCE ON COMPUTERS IN EDUCATION

Nov. 28th - Dec. 2nd, 2016

"Think Global Act Local"

Work in Progress Posters Proceedings

Sponsors from Govt. of India

Sponsors from Indian Industries

ISBN 978-986-84735-9-1

Publisher

Asia-Pacific Society for Computers in Education

No.300, Jungda Road, Jhongli District, Taoyuan City 320, Taiwan

Editors

Bo JIANG

Chengjiu YIN

Jayakrishnan WARRIEM

Weiqin CHEN

Jie-Chi YANG

Sahana MURTHY

Su Luan WONG

Sridhar IYER

PREFACE

This volume contains the Work-In-Progress Poster (WIPP) Proceedings of the 24th International Conference on Computers in Education (ICCE 2016). The WIPP session will provide opportunities for poster presenters to showcase well-formulated and innovative ongoing work or late-breaking results.

All the selected paper went through a rigorous blind review by independent peer reviewers to ensure high quality work. We hope that the papers in the proceedings will stimulate more research ideas and discussions among the young researchers.

On behalf of editors
Bo JIANG
Chengjiu YIN
Jayakrishnan WARRIEM

TABLE OF CONTENT

Chen, W. et al. (Eds.) (2016). Proceedings of the 24th International Conference on Computers in Education. India: Asia-Pacific Society for Computers in Education

Automatic Assessment of Reading with Speech Recognition Technology

Preeti RAO*, Prakhar SWARUP, Ankita PASAD, Hitesh TULSIANI and Gargi Ghosh DAS

Department of Electrical Engg., I.I.T. Bombay, India
*prao@ee.iitb.ac.in

Abstract: In this paper, we describe ongoing research towards building an automatic reading assessment system that emulates a human expert in a spoken language learning scenario. Audio recordings of read aloud English stories by children of grades 6-8 are acquired on an available tablet application that facilitates guided oral reading and recording. The created recordings, uploaded to a web-based ratings panel, are currently evaluated by human experts on four relevant dimensions. Observations of typical learner progress patterns will form the bases of a system that applies Automatic Speech Recognition (ASR) techniques to obtain robust automatic predictions of reading fluency and word decoding accuracy.

Keywords: oral reading assessment, tablet application, automatic speech recognition

1. Introduction

It is well known that in India's large rural population, millions of children complete primary school every year without achieving even basic reading standards (ASER, 2012). Since reading competence enhances overall learning by enabling the child to self-learn various subject material from the vast available text resources, the importance of imparting reading skills in early school cannot be over stated. Technology holds the promise of scalable solutions to alleviate the literacy problem. At least one recent effort that has gained visibility is the introduction of a feature known as same-language subtitling (SLS) in which synchronized text subtitles are incorporated in popular TV programs, including songs, so that viewers benefit from exposure to the script while simultaneously listening to the audio (Kothari et al. 2002). A further step in literacy training at the school level would be to facilitate oral reading. Reading aloud has traditionally been an important instructional component in the school system in many countries (Dowhower, 1994). Reading research articles over the decades have presented empirical evidence that assisted oral reading, while simultaneously listening to a fluent reader, is very effective in improving a student's reading skills. Repeated readings of a passage, whether assisted or unassisted, have been shown to lead to improvements both in word decoding and in comprehension. Further, these benefits carry over to new unpracticed texts (Dowhower, 1994; Rasinski, 2003). Given the importance of inculcating oral reading skills, specific scoring rubrics have been developed by educators to evaluate reading fluency in terms of accuracy, rate and expressiveness.

It is the goal of the present work to consider scalable technology solutions that facilitate oral reading practice and assessment in situations where access to language teachers is limited. We choose the specific context of L2 English which is a curriculum subject across schools in rural India where the medium of instruction is primarily the regional language. Qualified English teachers are scarce and even if the books are available, the students' exposure to communicating via speaking is severely limited. We describe a tablet based app that facilitates oral reading practice and present research that targets the automatic assessment of recorded speech according to accepted norms of reading proficiency. The goal is to achieve reliable means of objective feedback to (i) the student as a motivational component, and (ii) higher authorities in the school system as a meaningful student progress tracking. The proposed solution, developed for English, can be extended to any language.

Automatic speech recognition (ASR) is a technology that converts an acoustic speech signal to text using language and other constraints. It can therefore be applied, in principle, to evaluate the accuracy of the read speech of a student with reference to the known text. Further speech analysis techniques are available to derive other aspects of speech delivery such as expression. However, ASR is most successful in applications where the variability in the acoustics due to speaker and environment conditions are controlled. In the school reading scenario, we expect significant diversity in speaking accents and pronunciation, and the ubiquitous presence of background noise. To obtain robust ASR in this case, we need a system that is built specifically for the task and anticipated use case. A thorough understanding of the feedback and evaluation requirements and of the type of variations that occur in practice will be useful in addressing the mentioned challenges.

In the next section, we describe the tablet application we employ that serves to facilitate oral reading and recording. This is followed by a discussion on known attributes of oral reading proficiency which serve to define the scope of the automatic assessment. Methodologies for collecting and labeling audio data to train the machine learning algorithms of the ASR system are discussed.

2. The tablet reading application

Mobile tablets provide for a low cost, portable devices that can be easily handled by children. The screen space of a 7 inch tablet is sufficient for the convenient display of text and pictures in story reading. Interactivity can be easily incorporated via touch. Additionally, a microphone and camera are always available on device. Sufficient memory and internet connectivity ensure that the tablet can be embedded in a larger connected system where content delivery and transfer of data are easily achieved. There are a number of Android OS based tablets in the market that satisfy these basic requirements. There are a few apps available for such tablets that facilitate guided reading accompanied by a narrator voice and the display of text with word-by-word highlighting at a suitable reading pace. For the proposed work, we adopt the SensiBol Reading Tutor app (2016) for Android tablets due to the availability of customization for classroom use with multiple separate child accounts. The app allows a child to listen to a narrator reading out the story in a listening mode. The child can use the record mode while reading aloud herself. Both unassisted and assisted (i.e. shadowing the narrator) modes are available in the read aloud mode. The stored child recording synchronized with the video is available on the tablet for listening which feature encourages self-assessment and more practice. In order to make it even more interesting, the child's recording can be enhanced by adding audio effects and by mixing in any background track available in the original story video. Further, the child recording is also available to the teacher for review at any time. All recordings are made with a headset mic to minimize background noise which can be very deleterious for ASR. The content is a selection of stories from BookBox (2016), a readily available rich resource of illustrated text designed for child readers.

The SensiBol RT app also provides backend support where every registered child's audio recordings, together with metadata information such as child name, story name, date and time can be archived. A web-based ratings panel displays the audio at the sentence level together with the expected story text. This is obtained by segmenting the full story recording based on a combination of information from the video timings combined with the detection of long silences in the audio. The sentence-level audio can then be rated by a human expert on various dimensions in comparison with the corresponding narrator audio. This facility is vital for the process of creation of the labeled data resources required for ASR based system development. The labeling exercise is underway based on the field testing deployment of the reading app in a tribal school near Mumbai involving children in grades 6-8 where tablet based story reading is a scheduled and supervised activity conducted in the school hours (LETS, 2016).

3. Automatic prediction of reading skill with ASR

Proficient readers organize the text into meaningful phrases and read with appropriate prosody and pace apart from the correct pronunciation of words. Comprehension has been shown to be predictable by the prosody of the student's reading, i.e. its pauses and intonation (Miller, 2008). Thus prosodic oral reading can signal that children have achieved fluency and are more capable of understanding what they read. This suggests that performance in different dimensions, word decoding and prosody related, must be considered by an automatic assessment scheme. Our observations on field recordings of repeated readings (interspersed with listening to the narrator) of a single story collected from a group of 5 children (in grades 6-8) over a span of several days indicate the following:

1. The child typically starts out with reading out word by word in list style including disfluencies comprising of long pauses, hesitations and incomplete words.
2. Improvements are always noted in one or more dimensions from reading to reading (where there is a separation of at least one day). The earliest to improve is phrasing (pauses reduce, words become more smoothly connected). The reading pace gradually increases.
3. After 2 or 3 readings, the volume and intonation start varying (this also conveys improved confidence). Eventually, the correct intonation, matching that of the narrator, is achieved.
4. Some word pronunciations get gradually clearer. But certain wrong pronunciations, especially those linked to grapheme-to-phoneme confusions and L1 influence, persist.

We can conclude from the above that evaluation should target phrasing and reading pace in the first instance, followed by prosody. Finally, fine feedback on word pronunciations can be provided. Thus the level of feedback is adapted to the child's proficiency level. In order to build training datasets for the different stages of machine based evaluation, we are in the process of obtaining expert ratings of word accuracy (word substitution, insertion and deletion) and three prosodic attributes, viz. phrasing, volume and intonation and reading pace, on a 4-level scale at the sentence level. These ratings are obtained for each of several repeated readings of 18 distinct stories by 100 children from the selected school and grades. The word level transcriptions will be used to build acoustic models and language models for the ASR system that incorporate typical pronunciation variations and observed disfluencies. The prosody ratings will be used to train classifiers on selected acoustic features to detect proper phrasing and sentence intonation. All this functionality further needs to be robust to at least the low levels of background noise and interference expected in the audio.

While ASR has been used previously in objective assessment of language skills of children, the present work is targeted towards the more challenging scenario of continuous speech (rather than isolated words as in the extensive work by Alwan et al. (2007)). Achieving robust automatic evaluation will also help us to introduce more sophisticated testing for comprehension such as story retelling exercises.

Acknowledgements

We would like to thank Nagesh Nayak and Sujeet Kini of SensiBol Audio Technologies for help with the customization of the Reading Tutor App for this study. We thank Prof. Alka Hingorani and her team at IDC, I.I.T. Bombay for initiating the LETS project and organizing the field study.

References

ASER, The Annual Status of Education Report (Rural) (2012) url: http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER_2012/fullaser2012report.pdf

Kothari, B., Takeda, J., Joshi, A., & Pandey, A. (2002). Same language subtitling: a butterfly for literacy?. *International Journal of Lifelong Education*,*21*(1), 55-66.

Dowhower, S.L. (1994). Repeated reading revisited: Research into practice. *Reading and Writing Quarterly*, 10, 343-358.

Rasinski, T. V., & Hoffman, J. V. (2003). Oral reading in the school literacy curriculum. *Reading Research Quarterly, 38*(4), 510-522.

SensiBol Reading Tutor App, SensiBol Audio Technologies Pvt. Ltd. (2016), url: http://sensibol.com/readingtutor.html

BookBox, A book for every child in her language. (2016) url: www.bookbox.com

LETS (Learn English Through Stories), Tata Centre for Technology and Design at I.I.T. Bombay (2016). url: http://www.tatacentre.iitb.ac.in/15mobitech.php

Miller, J., & Schwanenflugel, P. J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly*, *43*(4).

Alwan, A., Bai, Y., Black, M., Casey, L., Gerosa, M., Heritage, M., & Narayanan, S. (2007). A system for technology based assessment of language and literacy in young children: the role of multiple information sources. In *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop*.

Chen, W. et al. (Eds.) (2016). Proceedings of the 24th International Conference on Computers in Education. India: Asia-Pacific Society for Computers in Education

Comparison of Concept Map Evaluation between Kit-Build Method and Handmade Method

Warunya WUNNASRI*, Jaruwat PAILAI, Yusuke HAYASHI & Tsukasa HIRASHIMA
Graduate school of Engineering, Hiroshima University, Japan
* warunya@lel.hiroshima-u.ac.jp

Abstract: This paper describes the reliability of our automatic concept map evaluation framework that is called Kit-Build concept map. This framework is developed based on concept map that is used for organizing and representing knowledge. It is enhanced to support an education. However, the automatic concept map evaluation has not investigated about the reliability. To confirm the reliability of Kit-Build concept map, we compare Kit-Build concept map with two handmade concept map evaluation methods that contain the structural scoring and the propositional scoring. These handmade methods can evaluate concept map flexibly because the human can understand the meaning of each proposition in concept map even the words of proposition do not appear in a learning material. So the handmade methods are claimed that they have the reliability enough for evaluating concept map. To reduce time cost and human workload, researchers propose the automatic concept map evaluation that becomes an important and useful for a classroom situation. It can evaluate a lot of concept maps in a short time, but it still needs to be examined the reliability. We designed preliminary experiment in two learning situations that are teaching and reading situations and compared the correlation between the handmade methods and Kit-Build concept map. Even though these are preliminary results, they suggest that using Kit-Build concept map in teaching situation gets acceptable reliability when it is compared the correlation with two handmade concept map evaluation methods.

Keywords: Concept Map Evaluation, Kit-Build Concept Map

1. The Concept Map Evaluation Method

Joseph D. Novak developed concept map as a graphical tool for representing and organizing knowledge, and it is also used to evaluate learners understanding in classes. So a lot of handmade concept map evaluation methods are proposed and get the reliability widely because their procedures require the human for evaluating. Nevertheless, these methods take cost such as time cost and human workload too much, so it is not convenient to use in classes that have a lot of learners. Thus the automatic concept map evaluations are proposed, but they still have to be investigated the reliability.

1.1 The Handmade Concept Map Evaluation Method

From our investigation, the handmade concept map evaluation methods are categorized by using the precedence of each method. One of a typical method is the Novak and Gowin structural method (Novak & Gowin, 1984), which is grouped in the structural scoring. They give high scores the precedence on a level of the hierarchy and the number of crosslink in a concept map while the valid proposition can get only one score per proposition. The significant meaning of proposition may be neglected. That is different from the McClure and Bell relational method (McClure et.al, 1999). This method is grouped in the propositional scoring because it pays attention to the meaning of proposition precedence. The procedure investigates the suitability of meaning of each proposition. If the linking word is appropriate with concepts clearly, that proposition will get three scores as a perfect score. The score will be depreciated depending on the meaning of linking word. It is reasonable for evaluating understanding from concept map. But these handmade methods have to use an expert for evaluating and require a long time for scoring each concept map.

1.2 The Automatic Concept Map Evaluation Method

Because of the time consuming and workload that has to pay for the handmade evaluation, the automatic concept map evaluation methods are proposed. Most of them use the criteria map as the target of learning. They compare the learner map with criteria map to evaluate learners' understanding. In this study, we call the automatic comparison concept map evaluation method. This comparison inherits the property from the human method that is the structural scoring and propositional meaning scoring. If learner maps are same as criteria map, it shows that learners can understand in instructor's objective well, which includes the understanding of structure and meaning of the proposition.

Our framework, Kit-Build concept map (Hirashima et al., 2015) is an automatic concept map assessment that uses the exact matching in propositional level for evaluating concept map. It has been already employed in classrooms practically and confirmed that the framework and results of the diagnosis were useful to support teachers in science learning in elementary school. That proves that it is suitable for using in teaching situation that instructor gives the direction following instructor's interpretation. However, we have not examined the quality of the propositional exact matching evaluation. And we have not studied our framework in reading the situation that learners have to interpret material by themselves. So we produce the experiment to investigate the reliability of Kit-Build concept map by comparing well-known handmade evaluation methods. For using Kit-Build concept map, the instructor has to prepare the criteria map, which is called the goal map in our framework. It is constructed as the informal concept map because it should follow the instructor's objective that requires learners to understand that is not the universe context. After that, the goal map is extracted to the kit that contains a list of concepts and relationships. This kit that is provided to learners can help learners to reduce their cognitive load more than the traditional concept map, which they must create all components by themselves. After that, learners are requested to reconstruct concept map by using the kit, and it is called the learner map. The framework will check leaner maps by exact matching on each learner's proposition with goal map's proposition and generates a similarity score. The instructor can investigate learners' misunderstanding individually and can find the overview of all learners by overlaying concept map as the group map and the group-goal difference map immediately. After result analyzing, the instructor can adjust the goal map or teach learners about leaky content again.

2. Research Methodology

To confirm the reliability, we produce the preliminary experiment to compare the correlation between the handmade concept map evaluation method and Kit-Build concept map. For the handmade evaluation method, we chose the Novak and Bell structural concept map evaluation and the McClure relational propositional method that they are the typical traditional method.

In this preliminary experiment, ten university students were requested to read the article that described "Introduction of concept map." After that, they had to construct concept maps following their reading interpretation by using 21 provided concepts on CmapTools application. It means they must create linking word by themselves. Two handmade evaluation methods evaluated these concept maps, and the raw scores of each method are normalized by using their perfect score. Then, they had to use Kit-Build concept map to reconstruct concept map by using kit. The kit contained 21 concepts that are same as provided concepts in CmapTools and additional 22 relationships. These concept maps were evaluated by our automatic evaluation method. The score is represented as the similarity score when learner map was compared with the goal map. After reading situation, the instructor taught the participants about the same article following instructor's interpretation. And the participants were requested to construct the concept map as same as steps in the reading situation. They had to create linking words by themselves and Kit-Build concept map.

3. Results of Preliminary Experiment and Conclusions

From the preliminary experiment procedure, the average scores of each evaluation method in reading and teaching situation are shown in Table 1. All scores of the reading situation are lower than teaching situation because participants used their interpretation to create concept maps that are the different way from the instructor's interpretation. And the score of the propositional scoring is higher than the structural scoring because the evaluator tried to understand the meaning of each proposition and tried to give the score to participants as much as possible. That is different from the structural scoring that forces evaluator to give the precedence to the structure of concept maps too much. It can make some leaky essential meaning. So the McClure and Bell relational method can show learners understanding more meaningful than the Novak and Gowin structural method.

Table 1: The result of experiment (*n*=10)

	Structural Scoring (Novak and Gowin's)	Propositional scoring (McClure and Bell's)	Automatic comparison (Kit-Build concept map)
Reading	24.56	35.09	30.30
Teaching	28.99	53.41	50.00

While the score of Kit-Build concept map is in the middle area, it is lower than the McClure and Bell propositional scoring but is higher than the Novak and Gowin structural scoring. Even though Kit-Build concept map uses the exact matching to score each proposition of learner maps, the scores are acceptable when are compared with the handmade concept map evaluations, which use very flexible matching for scoring the maps. Table 2 shows the correlation between the score of handmade methods and Kit-Build concept map. The p-values show we cannot discuss the correlation between the handmade methods with Kit-Build concept map in reading situations. Because, when the participants read the article, they interpreted the article by themselves and it is possible to be various ways. While the result from teaching situation has a marginal medium correlation between both handmade evaluation method and Kit-Build concept map, it shows the lecture from instructors can make an agreement on the article by teaching and guides the learners to understand in the same direction with the instructor. From the assumption that the handmade evaluation methods are reliable, we conclude Kit-Build concept map has reliability enough for evaluating concept map in teaching situation because it is not much different when it is compared with reliable handmade concept map evaluation methods.

Table 2. The correlation between handmade evaluation method and Kit-Build concept map

	Kit-Build in reading situation	Kit-Build in teaching situation
Novak and Gowin's structural scoring method	0.1406 (p-value=0.6984)	0.6209 (p-value=0.0553)
McClure and Bell's relational method	0.2702 (p-value=0.4503)	0.5520 (p-value=0.0980)

For the future work, we plan to make the full experiment and compare the stability of handmade evaluation and Kit-Build framework. Because different evaluators may score concept map by various interpretation individually, the score of handmade evaluation will depend on evaluators' interpretation that affects the stability of the assessment. While Kit-Build concept map uses the exact matching for comparing between goal map and learner map, it will return the same result in anytime. From this assumption, we try to confirm Kit-Build concept map has ability enough for evaluating concept maps.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 15H02931.

References

Hirashima, T., Yamasaki, K., Fukuda, H., & Funaoi, H. (2015). Framework of kit-build concept map for automatic diagnosis and its preliminary use. *Research and Practice in Technology Enhanced Learning*, 10(1), 1-21.

McClure, J.R., Sonak, B. & Suen, H.K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 36(4), 475–492.

Novak, J. D., & Gowin, D.B. (1984). Learning how to learn, New York: Cambridge University Press.

Chen, W. et al. (Eds.) (2016). Proceedings of the 24th International Conference on Computers in Education. India: Asia-Pacific Society for Computers in Education

Understanding and Analyzing Students Frustration Level During Programming

Prathmi KURTIKER[a*], Ramrao WAGH[a]
[a]*Department of Computer Science and Technology , Goa University, India*
*kurtikerprathmi@gmail.com

Abstract: We present our proposed and ongoing approach to detect the frustration among novice Java learners. We use the logs from the worldwide source code repository Blackbox that captures data from BlueJ IDE environment of our learners. We describe some factors which are the cause of frustration to novice programmer and conclude that it is possible to model an individual student level of frustration in laboratory settings using derived factors; and also possible to get an average frustration across all lab sessions. The real-time capturing, analysis and detection of frustration will result in timely intervention by instructors to ensure that learners regain their interest.

Keywords: Affective state, Frustration, Integrated Development Environment, Blackbox project, Intelligent Tutoring System

1. Introduction and Related Work

It is a common situation that during first programming courses, inability to perform the given programming exercises due to various reasons leads to frustration that causes students to disengage from the programming task and hamper further learning. If this state is not detected on time and attended to it may lead to more serious consequences. Our research focuses on detecting learners' affective state of frustration while learning to program in a controlled computer laboratory set up so that real time intervention by instructors is possible especially in a large class.

Identification of a learners affective state by Facial Expressions, Paralinguistic Feature (voice), Body Language and Posture, Physiology, Brain Imaging and EEG, Text, Multimodality are widely researched (Calvo & D'Mello, 2010). But these methods are not yet feasible in real-world lab setting due to the cost involved as well as the intrusion it will result in the laboratory set up. It has been shown that in a computer lab setting where learners program on any Integrated Development Environment (IDE), it is possible to detect frustration by log data (Rodrigo & Baker, 2009; Faw, 2015). Our approach is to use logged data generated while a student works on the lab exercises using BlueJ IDE. This can be adapted to other environments with suitable modification. The prior work on coarse grain detection of frustration (Rodrigo & baker, 2009) detect an average students' frustration across all five lab exercises using compilation logs, but failed to detect individual students' frustration between each lab session. In other similar work, (Faw, 2015) captured contextual and keystroke logs of students learning in Intelligent Tutoring System (ITS). This study presented and compared two models – model 1 that uses contextual data whereas model 2 that used both contextual and keystroke data and concluded that the model 2 gives better accuracy than model 1. (Altadmri & Brown, 2015) analyzed a novice programmer mistakes in large scale data from Blackbox repository of BlueJ users all over the world and found top 18 common and frequent mistakes made by programmer.

Our approach is to use the world-wide large-scale repository but use the data captured of only our students by providing identification markers so a individual frustration level can be computed using the captured data and demographic data known to us.

2. Method and Task Design

2.1 Environment and Task Description

Our ongoing study is conducted at the Department of Computer Science and Technology of Goa University on first degree Computer Science students learning Object Oriented Programming in Semester III. The class strength is 54. There are two courses related to Object Oriented Technology , one being a theory course discussing the OO Programming and modeling (using UML) while the lab course deals with Programming in Java and use of UML for group projects. Both the courses are taught by second author of this paper. The study is conducted during the lab hours when students engage in Programming using Java. The entire class is divided in two batches. Both the batches are of approximately equal capability as alternate students have been assigned to each batch. The recording of data is done for only one batch having 23 students (16 female & 7 male). This batch is asked to program using BlueJ IDE (www.bluej.org). The other batch is using Eclipse IDE (www.eclipse.org). The reason for capturing data for only one batch is that BlueJ IDE on every computer in the laboratory is configured with participant ID which can be associated with only one user. We will use other protocols such as observation protocols with respect to Eclipse batch to compare the performance of both the batches.

The BlueJ IDE captures and sends the contextual data and source code to Blackbox repository maintained by Blackbox team. So as the students complete their exercises in BlueJ, the Blackbox data repository project captures a live interaction data about each students programming task which is perpetual (Brown, Kölling, McCall, & Utting, 2014) and can be queried.

2.2 Procedure and Feature Extraction

As the conducted study is local to our students and in computer laboratory, the demographics information about each participant and the programming lab exercise they are working on is already known. In order to find individual student frustration across all lab sessions, our first step is to create student profile and a frustration profile for each BlueJ user. Blackbox captures data about each student in their own profile, so all the data about each individual will be captured in particular profile.

Our major challenge is to find only those factors that are intrinsic to Programming that lead to frustration. There may be other factors that lead to frustration among students. The second challenge is differentiating between confusion and frustration. It is noticed that, in the beginning of any lab session, students are confused but most of them are able to find their way and achieve the desired task without causing frustration but in some cases confusion persists and lead to frustration. We are interested in finding the point when confusion and other such factors turn into frustration so some corrective action can be taken including providing manual intervention by the instructor at that point.

Till date we have captured data of this batch for all sessions related to learning various basic OO features of Java programming. The more structured data capture will be undertaken in the later part of the course where simple design problems will be given to the students and they are expected to program the solution. We plan to also observe the sessions using systematic observation protocols so as to correlate the data captured with observed data.

3. Experimental Analysis

3.1 Feature Selection

In our model, Frustration is considered as cumulative in nature. At the beginning student is not frustrated or very less frustrated as the motivation towards goal is high. The individual students' frustration across all lab sessions given by $FS = FS_i + FS_{i+1} + FS_{i+2} + \ldots + FS_{i+n}$ where i (session) range from 1 to n. Each session will be evaluated to find a frustration based on the factors listed in **Table 1** and an aggregate score of frustration across each session for each individual will be found. The factors are explained below.

F1: (Error Quotient (EQ)) that ranges from 0 to 1 is introduced by (Tabanao, Rodrigo, & Jadud, 2008) and represents the students' struggling with syntax error. Given two compile event, it checks whether the error message, the line number and the source code is the same. Where 0 means no two successive compile events have same syntax error in program and 1 means every compile event result into same syntax error all time.

F2: Every invoked compile event results into success (true) or failure (false). If the number of compile events in a sequence that are false is more, this will cause frustration.

F3: If the self-perception of student is highly positive towards the goal of completion, he/she will create more edit event followed by compile event. If the errors still persist this leads to false state which may lead to frustration.

F4: Sometimes there is frequent invocation of compile event with/without edit event in a short time span which is followed by error message. This is sure sign of frustration.

F5: The time taken between two successful compile events in sequence is important factor, if time taken is large with respect to particular task and if the task is highly important rise in frustration will be high.

These factors can be computed from the data captured in Blackbox data repository.

Table 1: Factors that Cause Frustration During Programming.

Factors	**Cause**
F1	Error quotient (EQ)
F2	The number of false compile events in sequence
F3	The number of edit event follow compile event : error / false
F4	Frequent invocation of compile event with/without edit event in short time span
F5	Time between two consecutive successful compile event

4. Conclusion and Future Direction

The main purpose of this study is to find frustration among learners of a programming language such as Java. We have outlined some of the factor which cause and indicate frustration in novice Java learners in laboratory environment. We conclude that it is possible to model frustration by carefully drawing and combining features from Blackbox data repository project. In future we would derive and prove more factors which cause frustration in novice programmer and find individual student frustration in lab settings.

Acknowledgements

The authors would like to thank all the participants who agree to take part in the study. And also thank Blackbox project team University of Kent for their support to conduct our research.

References

Calvo, R. A., & D'Mello S. (2010). Affect Detection: An Interdisciplinary review of models, methods, and their applications. *IEEE Trans. Affect. Comput.*, 1:18-37.

Rodrigo, M. M. T., & Baker, R. S. J. d. (2009). Coarse-grained detection of student frustration in an introductory programming course. *In International workshop on Computing Education Research*, pages 75-80.

Fwa, H. L. (2015). Automatic detection of frustration of novice programmers through contextual and keystroke logs. *ICCSE conference.*

Brown, N. C. C., Kölling, M., McCall, D., & Utting, I. (2014). Blackbox: A large scale repository of novice programmers activity. *In Proceedings of the 45th ACM Technical Symposium on Computer Science Education, SIGCSE,* pages 223-228.

Altadmri, A. & Brown, N. C. C. (2015). 37 Million Compilations: Investigating Novice Programming Mistakes in Large-Scale Student Data. *In Proceedings of the 46th ACM Technical Symposium on Computer Science Education – SIGCSE,* pages 522-527.

Tabanao, E. S., Rodrigo, M. M. T., & Jadud, M. C. (2008). Identifying At-Risk Novice Java programmers Through the Analysis of Online protocols. *Philippine Computing Society Congress.*