T. Hirashima et al. (Eds.) (2011). Proceedings of the 19th International Conference on Computers in Education. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education

Statistical Level Checker with Personalised English Passage Suggestion

Wasan NA CHAI[ab] **Taneth RUANGRAJITPAKORN[b]**
Nualsawat HIRANSAKOLWONG[a] **Thepchai SUPNITHI[b]**
[a]Mathematics and Computer Science department,
King Mongkut's Institute of Technology Ladkrabang, Thailand
[b]Human Language Technology Laboratory,
National Electronic and Computer Technology Center, Thailand
wasannachai@gmail.com, khnualsa@kmitl.ac.th
{taneth.ruangrajitpakorn, thepchai.supnithi}@nectec.or.th

Abstract: In this paper, a system to classify a readability level of English reading passage and to match student personal interest is purposed. Student model is applied to collect student information for selecting their preferable passage topic. Statistical passage level checker is implemented to match student readability level with passage difficulty by using neural network. Three linguistic features, syllable, vocabulary and sentence complexity, are chosen to distinguish a difficulty difference among passage level. The best accuracy gained by the system is 86.25% and the constantly reliable feature for this task is a sentence complexity of the passage.

Keywords: Readability level checker, English reading passage, English learning, personalisation, neural network

1. Introduction

English class is one of the most boredom subject for Thai students since Thai children are not familiar with inflection, syntactic word order, and grammar learning therefore they become idle and inactive in class. Furthermore, their reading passages become tiresome since each student has individually preferable topic and they tend to lose their learning motivation to read an assigned non-interested passage. Letting students choose reading-passage by themselves also leads to significant burden for instructors to scope an appropriate level of those passages.

In language learning, readability checker tool or passage grading system are one of the important application that assists students and instructors in terms of reducing a load to select a proper readability level of reading passages. The major issue is that most of them was implemented on rule-based approach and the designed rules are reckoned for an English native speaker. The rule-based systems and methods including Flesch reading easy formula [1], Kincaid formula [2], SMOG-grading [3] and Fox index [4] are rigid and can hardly be applied to students who study an English language as their foreigner language since the level of English understanding and skill are rather different based on each country standard of language learning.

Recently, the framework for passage grading system using statistical approach was purposed [5]. However, the mentioned system was a framework which exploits three linguistic features; syllable, vocabulary and sentence complexity score, along with a conditional random field (CRF) as their machine learning to create a model of the passage level. Although CRF is reliable one among other machine learning techniques for its discriminative training, it has a specification to number management since CRF recognises an input as string, not an integer. So far, an experiment result of the framework has not been published. Another study on a passage grading system using supervised learning by a neural network [6] was later reported. It applied the same linguistic features as the above mentioned framework but the machine learning was altered to a neural network. They applied a neural network to generate three models based on each linguistic feature and exploited those models separately to level a reading passage based on an academic level. The limitation of the system is not much sufficient accuracy as around 80%. The statistical systems work properly in practical but the load falls to students who have to search a reading passage by themselves and they occasionally conduct a searching again if the system returns an unsuitable readability level to them.

The question to be solved in this paper is to find a solution for matching student readability with their preferable topic. Furthermore, to improve an accuracy of passage level checker, we extend the existing statistical passage level checker system using neural-network by integrating the three features into a single model and compare the result between those two methods. Last, each feature is focused to compare the efficiency and reliability among them.

2. Statistical Level Checker with Personalised English Passage Suggestion

Statistical Level Checker with Personalised English Passage Suggestion is an automatic system for matching student's readability and a difficulty level of an English reading passage with student personalisation. The system consists of two main parts; student model and passage level checker. Student model represents student information for selecting interesting passage and improving English skill for individual student while passage level checker is a tool to examine a compatible difficulty of reading passage to student. An overview of the tool is sketched in Figure 1.

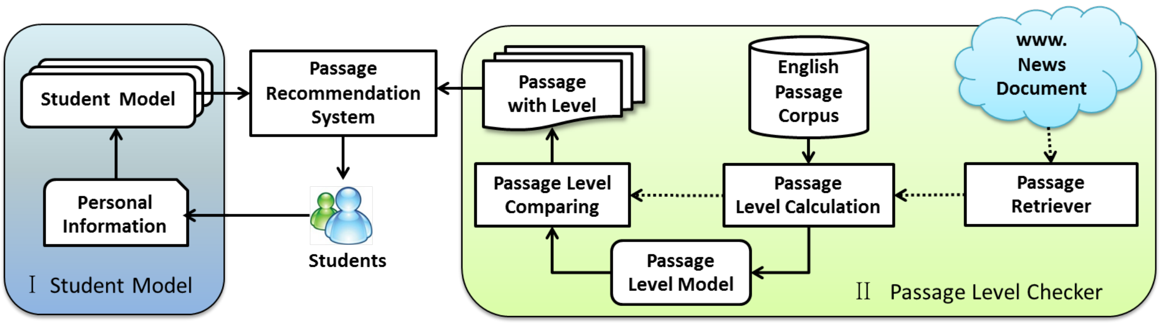

Figure 1. An overview of the system

2.1 Student Model

Students model is a representation of student personal information in several aspects. To recognise students' interest and performance in English learning, student model is designed to consist of three parts; 1) profile information, 2) interest information and 3) competency information.

The profile information is a personal profile that collects details about student name, gender, age and current academic level. The interest information gathers student interested topics and activities including their desired occupation, hobby, sport, favourite food and beverage, favourite song genre and movie, etc. The competency information is a collection of student competence from test results. Competency tests are designed to seek out student strong and weak fluency of English reading skills. The result will help on selecting the passage full of their weak linguistic feature which will improve such skill to individual. An example of student model is shown in Table 1. With the information, a passage retriever and a similar topic recommendation has been applied to select an appropriate passage to satisfy student interested topic.

Table 1. Examples of four sentence types with underlined criteria

Type	Subject		Student A			Student B		
profile	name gender age academic level		Peera Chareounsap male 15 level 9			Chutima Lakprae female 14 level 8		
interest	desired occupation hobby sport favourite food favourite beverage favourite music genre favourite movie		veterinarian, astronaut collecting stamp football, tennis pizza, noodle soda rock, pop thrill, sci-fi			dentist, scientist drawing badminton, swimming noodle, ice-cream fruit juice pop, r&b romantic, comedy		
competency	test result	time	vocab	grammar	summary	vocab	grammar	summary
		1st	17/30	12/30	18/30	17/30	12/30	18/30
		2nd	19/30	15/30	17/30	19/30	15/30	17/30
		3rd	20/30	14/30	20/30	20/30	14/30	20/30

2.2 Passage Level Checker

Passage level checker is a tool that automatically identifies a readability level of an English reading passage by comparing to a passage corpus. Three linguistic features; vocabulary, syllable and sentence complexity, are exploited to distinguish the differences among passage-level. In training process, level models based on the number of levels from a passage corpus is generated. The models are afterwards used to determine a level of a target passage and the tool returns its grade level as a result. A diagram of the passage level checker is demonstrated in Figure 2.

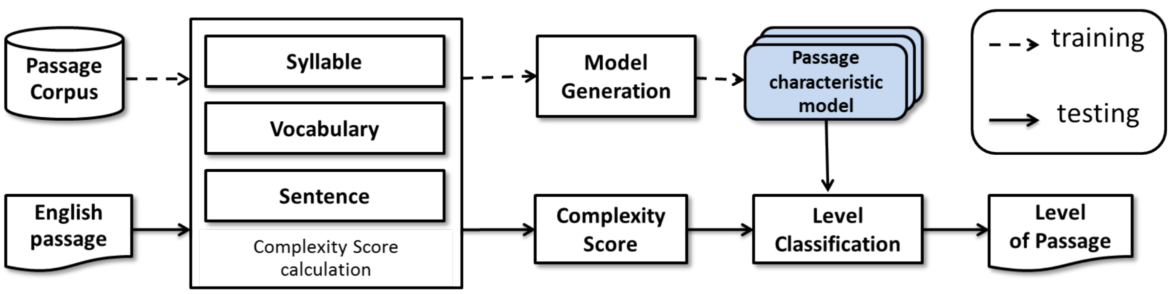

Figure 2. An overview of the passage level checker tool

In the pre-process for constructing an English reading passage corpus, Englsih passages are word-segmented by white space and sentence are divided by full-stop, question mark, and exclamation mark punctuation. An abbreviated form of auxiliary verb is expanded fully, for instance, "*he'll*" is transformed to "*he will*".

2.2.1 Syllable Complexity Score Calculation

A syllable complexity is a quality of average difficulty of words existing in the reading passage. AJAX syllable counter [7] is applied to count a syllable amount of each word. An average number of syllable for each passage is calculated by (1)

$$F1 = \frac{\sum_{i=1}^{n} (n_{syl_i})}{W} \qquad (1)$$

where n_{syl_i} is the number of syllable of word *i*th and *W* is the total number of words in a passage.

2.2.2 Vocabulary Complexity Score Calculation

A vocabulary complexity is a measurement of a lexical meaning difficulty in the passage context. To calculate vocabulary complexity, word classes which are content and function word are separately concerned because of their different significance.

A content word shows a stable lexical meaning and it is an open-class word which opens to possibilities for expansion. On the other hand, a function word is a word that contains little lexical meaning, but instead serves to express grammatical relationships with other words and function words are relatively small number of items. Moreover, content words are variable in form due to inflection. Therefore, words in a passage are split into two classes and handled separately in word level examination. For content words, lemmas[1] are extracted by Morpha [8][9], a lemmatisation tool, to prevent non-matching inflected forms. To some degree, content words are recognised to be more difficult than function words thus a parameter for a function word is set to 1.0 while a parameter for a content word is set greater to 1.5.

Beside of named-entities, unknown words that do not match the reference level word list are treated as the highest level since they are inclined to be a domain-specific word or specialised technical term. A process of vocabulary complexity score calculation is sketched in Figure 3.

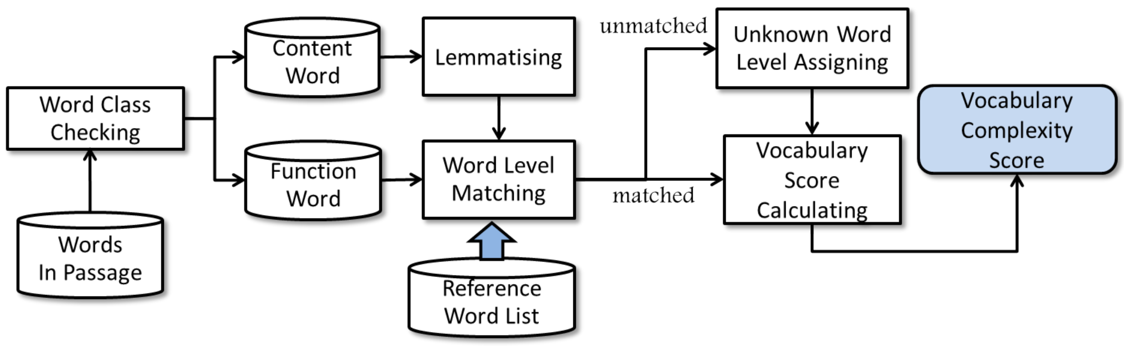

Figure 3. A process of vocabulary complexity score calculation

With above mentioned criteria, a vocabulary complexity score is computed by (2).

1 A lemma refers to the canonical form, dictionary form, or citation form of a word, e.g., in English, die, dies, died, and dying are forms of the same lexeme, with "die" as their lemma. It is different from a word stem which all affixes are removed.

$$F2 = \frac{\Sigma_{i=1}^{n}\left[\left(Lv_{c_i} \cdot 1 \cdot n_c\right) + \left(Lv_{f_i} \cdot 1.5 \cdot n_f\right)\right]}{\Sigma_{i=1}^{n}\left(Lv_{w_i} \cdot W_i\right)} \quad (2)$$

where Lv refers to a level of a word in reference list, c indicates a content word, f is a function word, n is a frequency, w refers to any kind of word and W_i is a frequency of i^{th} word.

2.2.3 Sentence Complexity Score Calculation

Sentence complexity is a difficulty of sentences in a passage. Apparently, the more clauses a sentence has, the more difficulty it gains. Basically, a sentence type in English falls into four types which are simple sentence (S), compound sentence (CP), complex sentence (CX) and compound-complex sentence (CPX). The main criteria used in this process is to capture an existence and a type of conjunction and clause marker within the sentence with co-occurring punctuation(s). The complexity score of sentence type is computed based on an amount of complexness of clause(s) by using (3).

$$F3 = \frac{\Sigma_{i=1}^{n}\left(S \cdot CP^{N_{CP}} \cdot CX^{N_{CX}}\right)}{Total_{sentence}} \quad (3)$$

where S refers to a simple sentence, CX indicates a complex sentence and CP is a compound sentence. N_{CX} is a number of a literation of a complex sentence and N_{CP} is a literation number of compound sentence. In case of *CPX*, it is counted if both *CP* and *CX* exist at least one. The examples of sentence types within the corpus are shown in Table 2.

Table 2. Examples of four sentence types with underlined criteria

Type	Examples
S	- The boy bought that books from the store. - Coral provides good hiding places for fish.
CP	- He sees the recycling truck, and he also sees Janey behind it. - It is really dark in the art gallery, but Harry has a light
CX	- Dinosaurs lived on Earth long before there were any people. - When he pushed the last sign into place, the seven lines of light shone into the middle where Henry used to stand on.
CPX	- The man wants the painting, but Harry doesn't have it since the painting has been stolen. - The two friends silently agreed, but their faces showed no fear though the beast ran toward them.

2.2.4 Model Generation and Level Classification

In the former implementation of statistical passage grading system [6], Neural-network [10] is applied to generate three models to determine the level of a passage. Currently, we alter the model generation in two steps by creating a model of features and apply the obtained model into vector to generate a level model by neural-network again. Figure 4 shows a comparison between former model generation method and the new method.

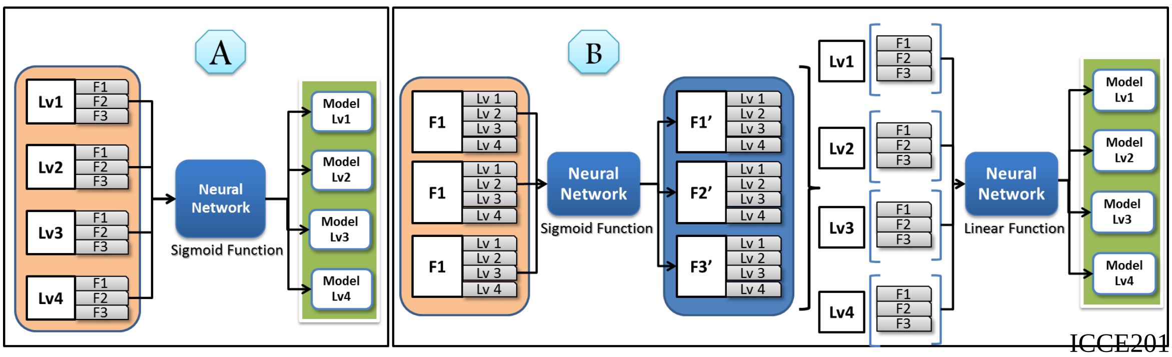

Figure 4. Flows between former model generation and the vector model generation

The difference of Figure 4A and Figure 4B is that the proposed vector method in figure 4B could explicitly return an available method to examine feature impact for each level. Once a passage model is obtained, it is used to classify a level of the target reading passage by the use of neural-network to determine probability as shown in Figure 5.

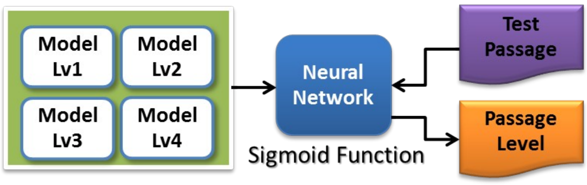

Figure 5. A level classification process

2.3 Integration of Student Model and Passage Level Checker

Once student model and result of passage level checker are gained, the recommendation seeks keywords and a domain topic from the retrieved passage and matches the found information with the interested item of individual student. The final result of the system is an English reading passage with preferable topic and suitable readability level.

There are two beneficial aspects for incorporating student model and passage level checker. The former is to help students to enjoy their reading with their interested topic passage. The latter is to select a reading passage with an appropriate level that suits student readability. Moreover, the system has an optional function to choose a passage which is compatible to student weak skills by observing the competency information to especially improve such skill. This will improve Thai student in English learning with their better attention and motivation and it also becomes a student oriented learning.

3. Passage Level Checker Experiment

3.1 Passage Corpus and Reference Word List

Reading passages used for training and testing were collected from reading passages and supplementary reading passage assigned in school class approved by the Ministry of Education of Thailand. The chosen passages are for Thai students who learn English as a foreigner language. The reading passages are divided into four grade-levels based on an academic grade of Thailand; junior primary school (grade 1-3), primary school (grade 4-6), junior high school (grade 7-9) and high school (grade 10-12). The number of passages for each level is 200 reading passages and the total number is 800.

To examine word level, reference word list is collected from vocabulary list approved by the Ministry of Education of Thailand. Table 3 shows a statistic of the reference word list of each grade-level.

Table 3. A statistic of reference word list from the Ministry of Education of Thailand

grade level	content word	function word	total
junior primary school	560	33	593
primary school	2,111	71	2,182
junior high school	3,566	82	3,648
high school	3,802	57	3,859
total	10,039	243	10,282

3.2 Experiment Setting and Result

To estimate an accuracy, 5-fold cross-validation is applied. Two methods of model generation; three model method and single metric model method, are tested separately. The comparison result between former model generation method and our method is shown in Table 4. To compare efficiency of single feature, accurate result of each feature and combination of features separated by level are shown in Table 5 where *Sy*, *Vo* and *Se* stand for syllable complexity, vocabulary complexity and sentence complexity respectively. A total amount and percentage of accuracy gained from each feature focused only from the correct results are given in Table 6.

Table 4. A result between former model generation method and the purposed method

	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Avg.
Former method	86.72%	84.38%	83.59%	80.47%	82.03%	83.44%
Metric method	83.59%	87.50%	88.28%	82.81%	85.78%	**86.25%**

Table 5. An accurate result of each feature and combination of features

		Sy	Vo	Se	Sy+Vo	Sy+Se	Vo+SE	All	Sum
Amount of Correct passage	Level 1	2	6	4	23	20	37	47	139
	Level 2	2	6	2	20	22	41	46	139
	Level 3	2	3	6	16	38	36	36	137
	Level 4	1	1	9	15	36	35	37	134
	Sum	7	16	21	74	116	149	166	549

Table 6. A total amount and percentage of accuracy gained from each feature

	Syllable (amount|%)		Vocabulary (amount|%)		Sentence (amount|%)	
Level 1	92	66.19%	113	81.29%	108	77.70%
Level 2	90	64.75%	113	81.29%	111	79.86%
Level 3	92	67.15%	91	66.42%	116	84.67%
Level 4	89	66.42%	88	65.67%	117	87.31%
Sum	363	66.12%	405	73.77%	452	82.33%

4. Discussion

The proposed model generation method does not give sufficiently higher accuracy than the former method. However, it allows us to directly investigate the impact of each feature rather than the former one which is hard to access to feature tuning.

From comparison of each feature, the sentence complexity constantly shows reliability for classifying passage level and the syllable complexity is a moderate feature in this task. For the vocabulary complexity, its accuracy obviously depends on a level of a passage. The vocabulary complexity performs greatly for lower levels while sentence complexity shows decent potential on higher level. Since the lower level passages (level 1 and 2) contain reoccurred simple vocabularies in the easy conversation style and the number of lexicons is small, vocabulary complexity can capture them simply and returns the most accurate result. In the other hand, the higher level passages (level 3 and 4) contain several lexical meaning words and the variety of them based on different domain cause the performance of vocabulary complexity to certainly drop. Moreover from error observation, we found two major issues which are unknown word issue and multi-sense word issue (polysemy) to emphatically lessen accuracy of vocabulary complexity. The former issue is caused by the missing word from reference word list. Many simple and general words are absent from the list especially a noun, for instance, “dragonfly”, “coral”, “glove”, “dinosaur”, “motive”, “helmet”, etc. These unknown words are ranked to highest level and cause the system to determine a passage containing them to higher level than its realistic level. Since the reference word list from the Ministry of Education of Thailand is not reliable because of non-coverage issue, the solution will fall to garner the words from the passage corpus itself and rank them by existence frequency in each level. The latter issue is a word with multiple meanings. They cannot be handled in the system efficiently since the system recognised them as they are the same and treat them as its lowest level in the reference word list. This causes the vocabulary complexity score calculation to give a lower level to a passage than it should be.

For syllable complexity, the length of syllable is not a certain measurement for reading difficulty since some short word can be more difficult than the longer syllable word, for instance, the word “woe” which possesses one syllable is definitely harder to understand for English learners than the word “butterfly” which counts as three syllables. Therefore, the performance of the feature is not much reliable for passage level determining task.

5. Conclusion and Future Work

In this paper, we purpose a statistical passage level classification system which provide an English reading passage with proper readability level and preferable topic for Thai student. Student model is applied to detect student interested topic of reading and their readability whilst passage level checker provides a level approval to filter a reading passage that does not suit student readability. The statistical passage level checker distinguishes a readability difficulty from different level by calculating three linguistic features which are syllable complexity, vocabulary complexity and sentence complexity. Neural-network is exploited to generate a level characteristic model based on three above-mentioned features from a reading passage corpus to prevent a rigidity of inflexible criteria for different English learning standard. From the experiment result, the average accuracy is 86.25% while the sentence complexity score shows a potential on passage level determination for a single feature. The variable accuracy depending on passage level falls to vocabulary complexity score which encounters a matching issue from reference word list in terms of polysemous ambiguity and non-coverage lexical entry. To improve the passage level checker, we plan to add word sense disambiguation to solve polysemous word issue. A better method to gather reference word list will be researched for better vocabulary complexity score calculation. Furthermore, new linguistic features such as speech type (direct speech and indirect speech) and idiom usage will be attached to specify more accurate readability difficulty. For recommending interested passage, a topic selection will concern with more implicit personal information, such as parent marriage status or their relationship with community, to prevent suggesting a non-suitable passage.

References

[1] Flesch, Rudolf . "A New Readability Yardstick," Journal of Applied Psychology, 32, 1948, 221-233.

[2] Luo Si and Jamie Callan.2001. A Statistical Model for Scientific Readability. In Proceedings of the tenth international conference on Information and knowledge management.

[3] McLaughlin, G. Harry (May 1969). "SMOG Grading — a New Readability Formula". Journal of Reading 12 (8): 639–646.

[4] Stephen Fuller, PharmD, Cheryl Horlen, PharmD, Robert Cisneros, PhD, and Tonja Merz, PharmD. 2007. "Pharmacy Students' Reading Ability and the Readability of Required Reading Materials". American Journal of Pharmaceutical Education 2007; 71 (6) Article 111.
[5] Wasan Na Chai, Taneth Ruangrajitpakorn, Thepchai Supnithi. 2010. A Statistical Approach on Automatic Passage Level Checking Framework for English Learner. In Proceedings of ICCE 2010.
[6] Wasan Na Chai, Naulsawat Hiransakolwong, Taneth Ruangrajitpakorn, Thepchai Supnithi. 2010. Passage Grading System Using Supervised Learning. In Proceedings of KICSS 2010.
[7] AJAX Syllable Counter : available at http://www.russellmcveigh.info/content/html/syllablecounter.php
[8] G. Minnen, J. Carroll and D. Pearce. (2000). 'Robust, applied morphological generation'. In Proceedings of the 1st International Natural Language Generation Conference, Mitzpe Ramon, Israel. 201-208.
[9] Morphological and Orthographic Tools for English (morpha): available at http://www.informatics.sussex.ac.uk/research/groups/nlp/carroll/morph.html
[10] Hornik, K., Stinchcombe, M., White, H., 1989, "Multilayer feedforward networks are Universal Approximators," Neural Networks 2: 359-366.