Image Recommendation for Informal Vocabulary Learning in a Context-aware Learning Environment

Mohammad Nehal HASNINE[a*], Kousuke MOURI[b], Brendan FLANAGAN[a], Gokhan AKCAPINAR[a, c], Noriko UOSAKI[d] & Hiroaki OGATA[a]
[a]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
[b]*Institute of Engineering, Tokyo University of Agriculture and Technology, Japan*
[c]*Department of Computer Education & Instructional Technology, Hacettepe University, Turkey*
[d]*Center for International Education and Exchange, Osaka University, Japan*
*hasnine.mohammadnehal.5z@kyoto-u.ac.jp

Abstract: This paper aims to develop an image recommendation system to support informal vocabulary learning for a context-aware learning environment called SCROLL (System for Capturing and Reminding of Learning Log). The idea is based on ubiquitous learning logs analysis, image (contextual and non-contextual) feature analysis, and information visualization techniques. We designed a model based on which the proposed Feature-based Context-specific Appropriate Image (FCAI) recommendation system will be developed. The purpose of recommending FCAI images is to assist language learners in informal learning of foreign vocabulary. Those images are recommended by analyzing 25350 ubiquitous learning logs (i.e. a learner's ethnographic information, learning location, learning time, learning context, image information etc.) at the right time, right place and in the right learning context-basis. This paper also articulates the approaches for ubiquitous learning logs analysis, word-bank formation, image analysis methods, and algorithm development.

Keywords: Informal Learning, Image Recommendation, Image-based Vocabulary Learning, Image Analytics, Ubiquitous Learning Analytics.

1. Introduction

EFL (English as a Foreign Language) learners' reported vocabulary learning in formal educational settings to be difficult and stressful (Turgut et al., 2009) and boring learning activity (Chen et al., 2008). These perceptions often affect the vocabulary learning interest and learning goal of learners (Wu et al., 2017). As a result, research attention is noticeably given to developing informal vocabulary learning systems. Language learning applications such as duolingo, rosetta stones, rakuten lingvist and iKnow etc. already drawn great attention to the motivated learners for informal learning of foreign vocabulary. Another inevitable problem with newly memorized words is quick forgetting because it is said that what is hard to memorize is often easy to forget. One solution to this problem is using visuals particularly still images along with texts. Because the power of visuals in vocabulary learning cannot be ignored. Pedagogical studies such as imagery techniques (Kellogg, 1971), the pictorial superiority effect and the dual coding theory (Paivio A. R., 1968; Paivio A., 1973) have already stated the cognitive benefits of images in the human brain while memorizing a new word or recalling a memorized word. However, the problem remained is determining an appropriate image at a right time, in a right place, and in a right learning context. Generally speaking, this task (determination of an appropriate image to represent a word) is exceedingly challenging for both human and computers. Now that we stepped into data-centered Society 5.0, this problem can be solved by leveraging the power of educational big data, image analytics, and information visualization.

This paper introduces the idea of Feature-based Context-specific Appropriate Images (FCAI) to support the informal vocabulary learning in SCROLL system (Ogata et al., 2011). The proposed system would be able to recommend context-specific appropriate images for a word to be learned at the right time, right place and under the right learning context.

2. Literature Review

Contextual image recommendation in the context of informal vocabulary learning is considered to be a young subdomain of context-aware learning environments. A contextual image search scheme is proposed by Lu et al. (2011) for textual context re-ranking and visual context re-ranking. This is claimed to be the first model that used both contextual data and visual context together to build a contextual image search scheme. This contextual image search scheme is based on the context capturing, contextual query augmentation, image search by text, and contextual re-ranking. However, the efficacy of this system is not tested for vocabulary learning. Therefore, we could not determine the learning effect of their system-suggested contextual images. Ubiquitous English Vocabulary Learning (UEVL) is a framework that assists learners in experiencing systematic vocabulary learning process (Huang et al., 2012). However, the system has no functionality to recommend contextual images or ubiquitous images. WordNet (Miller et al., 1990), a popular image database, contains a large number of images. However, it is unable to provide automatic recommendation of the contextual images by analyzing right time, right place and right context. Besides the above-mentioned research systems, commercial software like duolingo, rosetta stones, rakuten lingvist and iKnow can be found in the marketplace. However, these systems are also unable to recommend contextual images to represent a word. For informal vocabulary learning, we developed AIVAS (Appropriate Image-based Vocabulary Acquisition System) system that allows the on-demand creation of learning materials utilizing appropriate image (Hasnine et., al., 2017). AIVAS is supported by an Image Reranking Algorithm called AIVAS-IRA that is able to determine the most appropriate image to represent a concrete word to be memorized. Our previous study revealed a significant role of AIVAS-IRA suggested appropriate images for concrete noun acquisition. Additionally, for abstract nouns, our study revealed that AIVAS-IRA recommended appropriate images are more liked by foreign language learners over to the Yahoo suggested images (Hasnine, M. N., 2016). However, our studies were carried out with non-contextual images (i.e. the images found in widely used commercial image search engines).

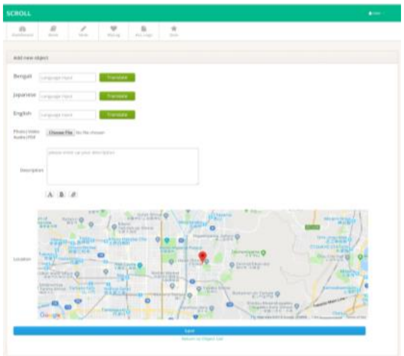
This paper presents the idea of Feature-based Context-specific Appropriate Images (FCAI) for informal vocabulary learning. In order to implement the system, we leverage the power of ubiquitous learning logs, image analytics, and information visualization techniques. A model for precise analysis of data collected from different educational environments is designed to support this research.

3. The Framework

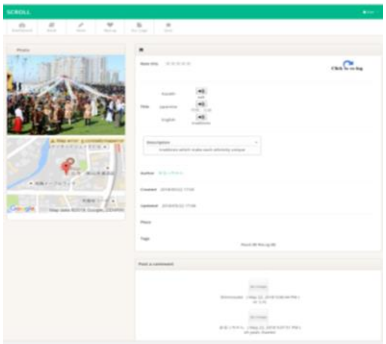
At present, in Kyoto University, we are developing a learning analytics framework for seamless learning. The LA framework offers the opportunity to learn in various environments regardless of the learning context, location or time (Flanagan et al., 2017). The framework is designed in the way that it provides an interface between integrated production and research systems to allow user authentication, information, and learning analytics results to be seamlessly transferred between systems (Flanagan et al., 2018). SCROLL system is a key component of this framework. Because, the system collects multi-dimensional ubiquitous learning logs as educational big data (Mouri et al., 2014).

3.1 SCROLL: A Context-aware Learning Environment

SCROLL, a multi-linguistic language learning tool that offers informal learning of foreign vocabulary. This client-server tool runs on different platforms including Android mobile phones, PC and general mobile phones (Mouri et al., 2014). A foreign language learner can create (as shown in Figure 1(a)) his/her own vocabulary learning materials using this system. A learning material (as shown in Figure 1(b)) in SCROLL system consists of a contextual image, the translation data, and the pronunciation. Moreover, the system is capable of recording a learner's vocabulary learning experience (such as the geolocation information, vocabulary knowledge, quiz, learning context, contextual image information etc.) as Ubiquitous Learning Logs (hereafter, ULLs) into its server. In SCROLL, a learning context is determined based on a learner's learning location and the textual description provided by the learner him/herself. A contextual image is the one that a learner uploads to create a vocabulary learning material. The role of a contextual image is that it helps a learner to recall his/her memory when taking part in quiz session. SCROLL[1] system is currently available for public use. A learner is able to create an account for free.



(a) Learning Material Creation Interface



(b) Sample of a Learning Material

Figure 1: SCROLL System

3.2 ULLs Analysis & Problem Identification

Our recent analysis indicates that, at present, SCROLL server contains 25350 vocabulary learning logs from 1684 learners. Most of the logs are created for either Japanese or English language's vocabulary learning. SCROLL ULLs are generated to record learners' learning experiences with images, audios, videos, location information, QR-code, RFID tag and sensor data etc. (Ogata et al., 2011, Mouri et al., 2014). Our recent analysis indicates that those ULLs are created for different parts of speech (such as nouns, verbs, adjectives etc.), sentences, phrases, numbers etc. Another analysis indicates that only 4720 contextual images are uploaded for 25350 logs. Which mean, only 18.6% vocabulary learning logs are created with contextual images and rest 81.4% learning materials are created without uploading any contextual images. This, in future, may lead SCROLL to an un-interactive vocabulary learning environment. We consider this as a problem for our learning analytics framework which needs to be solute. Additionally, the current system it is unable to recognize different ways of looking and categorizing learning objects (such as words, images etc.) based on the nationalities of the learners. Besides, at present, SCROLL only creates vocabulary learning materials with the contextual images that are uploaded by learners themselves. At present, there is no mechanism, except for using 'search' option, to recommend or lookup to others' uploaded images while creating a learning material. In order to provide a solution to this problem, we proposed a model for FCAI image recommendation system to support vocabulary learning using SCROLL system. By implementing this model, we plan to overcome the current limitations that the learners of foreign languages are experiencing while using SCROLL system.

3.3 The Proposed Solution: Feature-based Context-Specific Image Recommendation System

Learning analytics is not a genuine new research field rather it actually borrows from different related fields and synthesizes several existing techniques (Chatti et al., 2012). Learning analytics gives researchers the freedom to analyze and utilize big data collected from different educational environments. In this study, we designed a method by merging three techniques, namely ULLs analysis, image feature analysis, and information visualization. Based on this method, we propose Feature-based Contextual Appropriate Images (FCAI) for informal vocabulary learning. FCAI

[1] *https://scroll.let.media.kyoto-u.ac.jp/*

images are those that describe unique image features and the learning context based on the current learning location. Figure 2 displays the overview of the model that is designed and proposed to support this study.

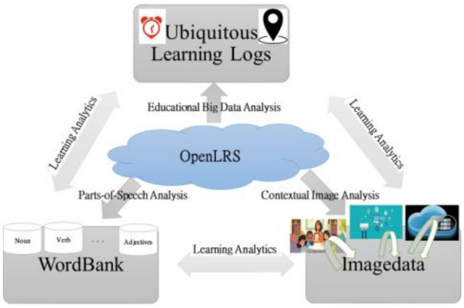


Figure 2: The Proposed Model

3.3.1 Model for Ubiquitous Learning Logs Analysis

In the central database, learning activities from various learning environments are storing on a regular basis. In this study, we extract and analyze the following logs from the central database:

- Vocabulary Information: Vocabulary that a learner wishes to learn in a specific context.
- Learner Information: Profile information of learners (i.e. name, age, gender, education etc.).
- Cultural Information: Information about nationality, social interaction etc.
- Study Place, Time & Location: Geo-locational information, place-details, and study time etc.
- Past Knowledge: Vocabulary that learners have previously acquired (i.e. learning history).
- Contextual Image Information: Unique image features (color, shape, object etc.) that may describe the learning context and/or the word itself.

Based on the above mentioned information, the model finds the learning patterns (synonymously, learners' learning behaviors) from the ULLs. By employing machine learning algorithms, the patterns between the unique image feature(s) will be determined.

3.3.2 Word-bank Formation

This study plans to form a new word-bank from the previous vocabulary information of SCROLL system. Because, SCROLL server not only contains words but also contains sentences, phrases, random number etc. For association analysis using text mining algorithms, the raw vocabularies need to be pre-processed utilizing text preprocessing techniques. Also for image recommendation, a word's multidimensionality and homonomy are challenges in any learning context, hence raw vocabularies need to be preprocessed and defined properly. In preliminary text analysis, we used Mecab, an open-source Natural Language Processing (NLP) tool that segments Japanese sentences into its Parts of Speech (POS). For English sentences, we used another NLP-based tool for annotating text with part-of-speech called TreeTagger. The abstractness-concreteness of an English word is determined based on two previous studies (Paivio et al., 1968, Brysbaert et al., 2014). It can be added that, abstractness-concreteness of a word refers to- whether a word is a noun (abstract/concrete etc.), verb, adjective or anything else.

3.3.3 Contextual and Non-Contextual Image Feature Analysis Techniques

In order to extract and analyze the unique contextual information, we plan to employ both hand-crafted and deep CNN-based unsupervised learning feature extraction methods. For the first prototype of the system, we plan to employ FFT-based feature extraction method in power spectrum and/or pre-trained AlextNet as the feature extractor to analyze image features. To begin with, we plan to analyze the contextual images that are stored in SCROLL server. In this study, a contextual image is the one that describes a learners learning context and location (Figure 2(a)). When taking part into quiz session, the contextual images help learners in recalling their memories. However, SCROLL server contains roughly 4720 contextual images. The number is not sufficient for a machine learning algorithm to predict on the most appropriate FCAI image. Therefore, we plan to accumulate more contextual images using crowdsourcing tools. Crowd4u[2], a micro-volunteering and crowdsourcing platform is going to be used for our crowdsourcing task. With this tool, we plan to accumulate contextual data from different sub-continents of the world.

Together with contextual images, we also plan to accumulate non-contextual images. Non-contextual images are those that found in widely used commercial image search engines. We plan to accumulate non-contextual data from Google image search engine by using Fatkun batch image downloader. Additionally, we plan to use AIVAS image sets (Hasnine et al., 2016, 2017). With the help of contextual and non-contextual images, we plan to establish a benchmark dataset for our algorithm.

(a) Contextual Image (extracted from SCROLL)   (b) Non-contextual Image (taken from AIVAS)

Figure 3: Example Contextual Image and Non-Contextual Image

3.3.4 Algorithmic Steps

The key algorithmic steps are shown in Figure 2.

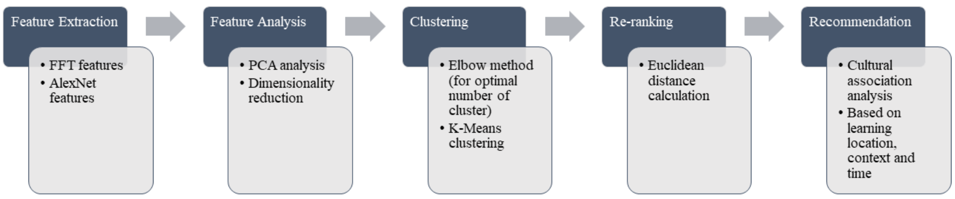


Figure 2: Algorithmic Steps of FCAI Recommendation System

For implementation of the algorithm, extracted features will be analyzed. Then, the optimal numbers of clusters have been determined from the dataset using clustering algorithm. After that, centroid of each cluster will be determined. The centroid of each cluster is the scale of appropriateness for that particular cluster. Next, based on the scales of appropriateness, FCAI images will be re-ranked. Finally, FCAI images will recommend in multiple categories and the most appropriate images in each category will be ranked.

4. Discussion and Future Directions

This paper presents the idea of recommending FCAI images for informal vocabulary learning. Our recent analysis (refer to 3.2) on over 25000 ubiquitous learning logs led us to propose this study. An FCAI image contains unique visual features that describe the word most appropriately in a specific learning context and location-wise. In order to build an environment for next-generation ubiquitous vocabulary learning, we plan to integrate SCROLL with AIVAS (Hasnine et al., 2017) system that is already capable of recommending ‘appropriate images’ for concrete and abstract nouns. By doing this, we would be able to interpret images and analyze their features to identify unique attributes that common among learners bearing similar cultural background. In the context of vocabulary learning, word-learning mechanisms may be different for nouns and verbs. Different types of visual aids have different effects on our cognitive processes. It is said that verbs are generally harder to learn than nouns and that moving objects have more impact on human cognitive processes compared to still

[2] *https://crowd4u.org/en/*

images. Moreover, many adjectives, adverbs, prepositions, and conjugations are very difficult to express with contextual/non-contextual images. Therefore, initially, the proposed FCAI system will be for English nouns. The development of the first prototype of the FCAI recommendation system is underway. In future, we plan to evaluate the output by analyzing learners' interaction with FCAI images and measuring short/mid/long-term memory retention rates. Moreover, we plan for interaction and cultural association analysis using information visualization technique.

Acknowledgements

This research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304.

References

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, 46(3), 904–911.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6), 318–331.

Chen, C. M., & Chung, C. J. (2008). Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education*, 51(2), 624–645.

Flanagan, B., & Ogata, H. (2017). Integration of Learning Analytics Research and Production Systems While Protecting Privacy. *In Proceedings of the 25th International Conference on Computers in Education (ICCE 2017)* pp. 333–338.

Flanagan, B., & Ogata, H. (2018). Learning Analytics Infrastructure for Seamless Learning, Companion Proceedings of the 8th International Conference on Learning Analytics and Knowledge (DesignLAK2018).

Hasnine, M. N., Ishikawa, M., Hirai, Y., Miyakoda, H., & Kaneko, K. (2017). An Algorithm to Evaluate Appropriateness of Still Images for Learning Concrete Nouns of a New Foreign Language. *IEICE Transactions on Information and Systems*, 100(9), 2156–2164.

Hasnine, M. N., Ishikawa, M., Hirai, Y., Miyakoda, H., Pemberton, L., & Kaneko, K. (2016). An Image Recommender System that Suggests Appropriate Images in Creation of Self-Learning Items for Abstract Nouns, International Journal of Management and Applied Science, 2(5), 38–44.

Huang, Y. M., Huang, Y. M., Huang, S. H., & Lin, Y. T. (2012). A ubiquitous English vocabulary learning system: Evidence of active/passive attitudes vs. usefulness/ease-of-use. *Computers & Education*, 58(1), 273–282.

Kellogg, G. S., & Howe, M. J. (1971). Using Words and Pictures in Foreign Language Learning. *Alberta Journal of Educational Research*, 17(2), 89–94.

Lu, W., Wang, J., Hua, X. S., Wang, S., & Li, S. (2011, November). Contextual image search. *In Proceedings of the 19th ACM international conference on Multimedia* (pp. 513–522). ACM.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International journal of lexicography*, 3(4), 235–244.

Mouri, K., Ogata, H., Uosaki, N., & Liu, S. (2014). Visualization for analyzing ubiquitous learning logs. *In Proceedings of the 22nd International Conference on Computers in Education (ICCE 2014)* (pp. 461–470).

Ogata, H., Li, M., Hou, B., Uosaki, N, El-Bishouty, M. M., & Yano, Y. O. N. E. O. (2011). SCROLL: Supporting to share and reuse ubiquitous learning log in the context of language learning. *Research & Practice in Technology Enhanced Learning*, 6(2).

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology*, 76(1p2), 1.

Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, 11(4), 137–138.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive psychology*, 5(2), 176–206.

Serrano-Laguna, Á., Martínez-Ortiz, I., Haag, J., Regan, D., Johnson, A., & Fernández-Manjón, B. (2017). Applying standards to systematize learning analytics in serious games. *Computer Standards & Interfaces*, 50, 116–123.

Turgut, Y., & Irgin, P. (2009). Young learners' language learning via computer games. *Procedia-Social and Behavioral Sciences*, 1(1), 760–764.

Wu, T. T., & Huang, Y. M. (2017). A mobile game-based English vocabulary practice system based on portfolio analysis. *Journal of Educational Technology & Society*, 20(2), 265–277.