Online English Vocabularies Autonomous Learning Model Based on EM Algorithm

Bo Zhang1, Ji Fang 2* and Zhan Xiong 3
1 School of computer science, Wuhan Donghu University, Wuhan,Hubei, 430212, China
2 Library, Huazhong University of Science and Technology, Wuhan,Hubei, 430074, China
3 School of Electrical Engineering and Telecommunication, Faculty of Engineering,University of New South Wales, Sydney, NSW, 2032, Australia
*Corresponding author’s e-mail: fangji@hust.edu.cn

Abstract. Under the background of the rapid spread of the Internet, a variety of English vocabularies learning models have emerged in mobile Internet terminals. On the one hand, its development inherits the traditional English learning theory. On the other hand, it also reveals its own uniqueness. The acquisition of human information is time-sensitive and the generalization of information is full of the horizons of language learners. The problem is how to effectively reduce negative impacts and improve learning efficiency. The EM algorithm can effectively filter English vocabularies and generate a mean value from a mixture of k different normal distributions, which makes English vocabulary online learning more efficient.

1. Introduction
With the continuous updating of the Internet and communication methods, English for children, which is aimed at children aged 2-15 years, is increasingly being valued by parents. The demand for English language learning is showing a trend of aging and the form of online education is gradually becoming more mature. From traditional offline one-on-one teaching and one-to-many teaching in the classrooms to online language education and the combination of online and offline, there are a variety of educational forms.

Compared with passive one-way traditional English teaching, the use of language teaching has greatly enriched the forms of language teaching such as big data, speech recognition, live interaction, remote evaluation and other new technologies. The diversity of English education forms is redefining language education which enables parents and children to have more and more personalized choices. At present, the main form of learning for children is to learn only through offline training institutions, accounting for 60.0%. In the future, the English learning methods of children will be changed from pure offline to the combination of online and offline, which will account for 74.1%. The 69.1% of the parents believe that the effect of children’s English learning is mainly detected by the communication ability during their daily life and work. The second one is the test scores of school or progression and the children’s devotion of school. The content of the course is not wonderful thus children dislike
these courses which is the main reason why the course can't be continued. The parents of children in the age of 2-6 are especially prominent, accounting for 36.6%.

The mobile Internet English vocabularies professional application mode is a huge subversion for traditional English vocabularies acquisition. It breaks the time and space restrictions of traditional education, which allows learners to learn repeatedly at an appropriate time. The level of intelligence integration is extremely high which can eliminate the process of summary, induction and search by themselves. What is more, as long as it is connected with the Internet, you can use it. Compared with the traditional English vocabularies acquisition mode, the mobile Internet terminal fully mobilizes the visual and auditory of the vocabulary learners at the same time, which is convenient for efficient memory. This acquisition model is completely self-service so that the effect of teachers has decreased. The use of various favorite network materials to actively learn has become the best practice mode of the autonomous learning concept. The intelligent program design of the network can effectively help the learners to complete the independent planning, independent monitoring and independent evaluation in the process of independent learning.

The mobile Internet English vocabularies autonomous learning model has the following characteristics: high connectivity. The biggest difference between the mobile Internet and the traditional Internet is that it uses the "cloud" computing to connect traditional PCs and wireless mobile terminals. The process of transferring the resources on the traditional Internet and PC to the mobile terminal is realized, which makes the learning life more convenient. The learners can complete the learning process only on the small mobile devices instead of the PCs. Free learning, storage and switching between mobile devices make the autonomous learning process of English vocabularies highly connected. This high degree of connectivity is only between English vocabularies learners. The connectivity between teachers who provide online instruction also increases. This powerful connection has completely jumped out of the traditional learner interpersonal circle, network and equipped with high-quality and excellent English vocabulary education resources over the world.

This mode has the infinity of time and space. The learning of English vocabularies is not logically systematic because there is less connection on the system. Therefore, vocabularies can be memorized in groups by groups. Such memorizing method does not require a large period of time so that fragmented time can be fully utilized to learn. For the space, the mobile network becomes a fixed and moving air classroom, which allows English learners to gather in a platform to share resources that suit them.

Another characteristic is the generalization of information intensity. English learners face a variety of vocabulary learning network materials. The high-density exposure of information fragments for English vocabulary learners has made them to select the most useful learning information which is more appropriate. Actually most of this information cannot cause learners to pay enough attention thus it becomes generalized. Such generalization is negative for autonomous vocabulary learners because the high-intensity generalization distracts the learner's attention and wastes time. This is not conducive for vocabularies’ learning.

2. optimization of EM algorithm
The EM algorithm is a method for parameter maximum likelihood estimation proposed by Dempster, Laird and Rubin in 1977. It can estimate the maximum likelihood of parameters from a complete data set for processing incomplete data such as defect data, truncated data, and noisy data.

English vocabularies data D is considered as a set generated from a mixture of k different normal distributions. It has the following characteristics: they are randomly selected from the normal distribution; xi is distributed in this order. The simple case of children's online English vocabularies autonomous learning mode is taken into consideration. The choice of a single normal distribution is based on a uniform probability and the k normal distributions have the same variance $\sigma^2$ which is known. A hypothesis $h = (\mu_1 ... \mu_k)$ is the input, that describes the mean of each of the k distributions and finds a maximum likelihood hypothesis for the mean, which means that a hypothesis h maximizes p(D/K).
When the data $x_1, x_2, \ldots, x_m$ are extracted from the normal distribution, the maximum likelihood hypothesis of the mean $\mu_{ML}$ of the distribution can be obtained:

$$\mu_{ML} = \arg\min_{\mu} \sum_{i=1}^{m} (x_i - \mu)^2$$

(1)

In this case, the maximum likelihood hypothesis $\mu_{ML}$ is equal to the sample mean:

$$\mu_{ML} = \frac{1}{m} \sum_{i=1}^{m} x_i$$

(2)

A mixture of $k$ different normal distributions, each of which is considered to be a triple $(x_i, z_{i1}, z_{i2})$, where $x_i$ is the observed value of the $i$th data and $z_{i1}, z_{i2}$ indicates which of the two normal distributions is used to generate the value $x_i$. Specifically, the value of $z_{ij}$ is 1 when $x_i$ is generated by the $j$th normal distribution, otherwise it is 0. $x_i$ is the observed value in the description and $z_{i1}, z_{i2}$ are hidden variables. If the values of $z_{i1}, z_{i2}$ are determined, the mean $\mu_1$ and $\mu_2$ can be obtained.

The EM algorithm is applied to the mean of $k$ normal distributions, the purpose of which is to search for a maximum likelihood hypothesis by continuously re-evaluating the expected values of hidden variables $z_{ij}$ based on current assumptions $(\mu_1, \mu_2)$. The maximum likelihood hypothesis is recalculated using the expected values of the hidden variables. The hypothesis is initialized to $h = (\mu_1, \mu_2)$, where $\mu_1$ and $\mu_2$ are the arbitrary initial values. Repeat the following steps to estimate $h$ until the process converges to a stable value. Assuming the current assumption $h = (\mu_1, \mu_2)$ is true, calculate the expected value $E[z_{ij}]$ of each hidden variable $z_{ij}$. $E[z_{ij}]$ is the probability that $x_i$ is produced by the $j$th normal distribution, $(\mu_1, \mu_2)$ and $x_i$ are the determined values:

$$E[z_{ij}] = \frac{P(x = x_i | \mu = \mu_j)}{\sum_{j=1}^{k} P(x = x_i | \mu = \mu_j)} = \frac{e^{-\frac{1}{2} \sigma^2 (x_i - \mu_j)^2}}{\sum_{j=1}^{k} e^{-\frac{1}{2} \sigma^2 (x_i - \mu_j)^2}}$$

(3)

Assuming that each hidden variable $z_{ij}$ takes the expected value $E[z_{ij}]$ obtained in step 1, a new maximum likelihood hypothesis is calculated and this hypothesis $h = (\mu_1, \mu_2)$ is replaced with a new hypothesis $h' = (\mu_1', \mu_2')$. A new maximum likelihood hypothesis $h' = (\mu_1', \mu_2')$ will be derived by $E[z_{ij}]$ and the sample mean $\mu_j$ will be derived and estimated from a single normal distribution. The new expression is simply the weighted sample mean of $\mu_j$ and the expected value $E[z_{ij}]$ produced by the $j$th normal distribution is weighed.

$$\mu_j = \frac{\sum_{i=1}^{m} E[z_{ij}] x_i}{\sum_{i=1}^{m} E[z_{ij}]}$$

(4)

3. the conclusion

There are many coincidences between the mobile Internet English vocabulary autonomous learning model and the traditional English autonomy learning theory, which are isomorphic. The autonomy of English vocabularies autonomous learning mode based on the mobile Internet breaks the category of autonomy in traditional theory. This autonomy depends entirely on the learner's individual learning habits, without the direct intervention of teachers and the direct influence of other learners. However this kind of autonomy is too free and random for students to learn, which reduces the attention to the learning object and thus affects the learning effect.

Acknowledgments

This paper was co-supported by the Natural Science Foundation of Hubei Province under grant 2016CKC777, the Foundation for Young Scholars in Wuhan Donghu University under grant 2018dhzk006. Thanks for the support of Hubei provincial science, Technology Department and Wuhan Donghu University. Thank you for providing the experimental platform in Wuhan Donghu
University. We have completed the implementation of the algorithm on this platform and the experimental data is presented.

Reference

[1] Mohsen. M. A, Balakumar. M. A review of multimedia glosses and their effects on L2 vocabulary acquisition in CALL literature. Recall. 2011, 23: 135-159.
[2] Wu .W-H, Jim. W Y-C and Chen C-Y. Review of trends from mobile learning studies: A metal-analysis. Computers & Education, 2012, 59(02): 817-827.
[3] CANTONI. V, CINQUE. L and GUERRA. C. 2-D object recognition by multiscale tree matching. Pattern Recognition, 1998, 31(10): 1443-1454.
[4] OFLAZER. K. Error-tolerant retrieval of trees. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(12): 1376-1380.
[5] PELILLO. M, KALEMM. S and ZUCKER. S. W. Matching hierarchical structures using association graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(11): 1105-1120.
[6] WANG. T. L, SHAPIRO. B. A and SHASHA. D. An algorithm for finding the largest approximately common substructures of two trees. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(8): 889-895.
[7] MESSMER. B. T. Efficient graph matching algorithms for preprocessed model graphs. Switzerland: University of Bern, 1996.
[8] MESSMER. B. T and BUNKE. H. A new algorithm for error-tolerant subgraph isomorphism detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(5): 493-504.
[9] MESSMER. B. T and BUNKE. H. A decision tree approach to graph and subgraph isomorphism detection. Pattern Recognition, 1999, 32(12): 1979-1998.
[10] MESSMER. B. T and BUNKE. H. Efficient subgraph isomorphism detection: A decomposition approach. IEEE Transactions on Knowledge and Data Engineering, 2000, 12(2): 307-323.
[11] CORDELLA. L. P, FOZZIA. P and SANSONE. C. A (sub)graph isomorphism algorithm for matching large graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(10): 1367-1372.
[12] MCKAY. B. D. Practical graph isomorphism. Journal of Symbolic Computation, 2014, 0(1): 94-112.
[13] MESSMER. B. T and BUNKE. H. Error-correcting graph isomorphism using decision trees. International Journal of Pattern Recognition and Artificial Intelligence, 1998, 12(6): 721-742.
[14] DIAS. J. R and MILNE. G. W. A. Chemical applications of graph theory. Journal of Chemical Information & Modeling, 1976, 32(1): 210-242.
[15] POOLE. J. Similarity in legal case-based reasoning as degree of matching between conceptual graphs: Work in progress. Proceedings First European Workshop on Case-Based Reasoning, 1993.
[16] B. RNER. K, PIPPIG. E and TAMMER. E. C. Structural similarity and adaptation. European Workshop on Advances in Case-Based Reasoning. Springer Berlin Heidelberg, 1996.
[17] FISHER D H. Knowledge acquisition via incremental conceptual clustering [J]. Machine Learning, 1987, 2(2):139-172.
[18] LEAKE. D. B, PLAZA. E and INTELLIGEN. Z. Case-Based Reasoning Research and Development. Second International Conference on Case-Based Reasoning, 1997.
[19] EHRIG H, HABEL A, KREOWSKI H. J. Introduction to graph grammars with applications to semantic networks [J]. Computers & Mathematics with Applications, 1992, 23(6-9):557-572.
[20] MAHER, P. E. A similarity measure for conceptual graphs. International Journal of Intelligent Systems, 1993, 8(8): 819-837.
[21] SHOWBRIDGE, P., KRAETZL, M and RAY, D. Detection of abnormal change in dynamic networks. Information, Decision and Control, 1999(IDC 99). IEEE, 1999: 557-562.
[22] WANG, Y. K., FAN, K. C and HORNG, J. T. Genetic-based search for error-correcting graph isomorphism. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1997, 27(4): 588-597.