Curriculum Q-Learning for Visual Vocabulary Acquisition

A. H. Zaidi, R. J. Moore & E. J. Briscoe
University of Cambridge, Computer Laboratory
ahmed.zaidi@cl.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
Faculty of Computer Science and Techology

Task

Learning an optimal **curriculum** for vocabulary acquisition using reinforcement learning and visual prompts.

Motivation

- Factors that teachers need to consider when constructing a curriculum: **difficulty** and **appropriateness** of content.
- Difficulty is measured relative to the *zone of proximal development (ZDP)*, introduced by Vygotsky, which is a representation of what a learner is capable of achieving without help, with *some* help, and of concepts that are beyond the learner's current ability.
- Appropriateness is a measure of whether content being presented is within the ZPD or, in the case of scaffolding, comprises material from within the ZPD.
- Determining difficulty and appropriateness is laborious and resource intensive task that involves experts conducting focus groups and analysis. Current methods of curriculum design are inefficient and assumes a static curriculum for all students.

Proposal

We propose the use of reinforcement learning (RL) in order to learn an optimal curriculum (policy) for each student for the task of visual vocabulary acquisition. We evaluate our models by simulating three types of student at different levels of proficiency *(beginner, intermediate, and advanced)*.

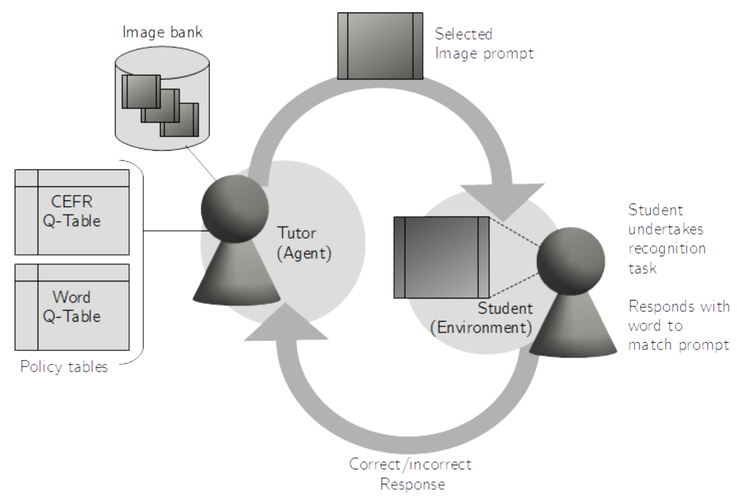

Figure 1: Overview of the system. A simulated student takes the place of a human actor in our study.

Curriculum Q-Learning

In order to automate the process of curriculum learning for visual vocabulary acquisition, we must first identify the key components of our RL system.

- The agent is the automated tutor that must learn what information to present to the student.
- The environment is the student with whom the agent is interacting.

The RL algorithm used by our proposed system is Q-Learning, an *off-policy* algorithm for Temporal Difference (TD) Learning. Q-Learning can be defined as follows:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q^{\pi}(s',a') - Q(s,a)] \quad (1)$$

We incorporate two models into our system, the *Common European Framework of Reference (CEFR) level model* and *word level model*. CEFR is an international level for language ability. The CEFR level model has 6 states which are defined by the 6 CEFR levels. The actions are whether the student should progress to the next level, stay in the current level, or go back a level. The word level model has two states: active (show the word), inactive (hide the word). The actions are remain in the current state or toggle state.

CEFR	Back	Remain	Forward
A1	0	1	0
A2	0	1	0
B1	0	1	0
B2	0	1	0
C1	0	1	0
C2	0	1	0

Figure 2: Q-Table for CEFR Level model. The table is biased towards the remain state at initialisation to ensure that the student remains in the current CEFR level until it is no longer beneficial to do so.

Status	Remain	Toggle
Active	0.1	0
Inactive	0	0

Figure 3: Q-Table for Word Level model. The table is biased towards the remain state at initialisation to ensure that the word is always seen at least once for each student.

To evaluate the students' understanding, we present a word in the form of an image. The objective for the students is to describe the image, and based on their response, the Q-Learning algorithm and thus the policy is updated. A valid response is defined by the target word associated with the image or a near synonym of that target word, which is automatically generated by looking at the top 10 nearest words to the target word in a pre-trained *word2vec* model [1].

At each interaction with the tutor, the student is rewarded in the following way:

- Correct answer is rewarded negatively (-1)
- Incorrect answer is rewarded positively (+1)

Simulations

To evaluate the performance of our system, we simulated three types of students at varying levels of proficiency: *beginner, intermediate and advanced*. In this case, we modelled the student's probability of getting a question correct as a negated Gompertz [3] distribution:

$$P(success \mid u,q) = 1 - \exp(-b\ \exp(-c(l(q) - l(u)))) \quad (2)$$

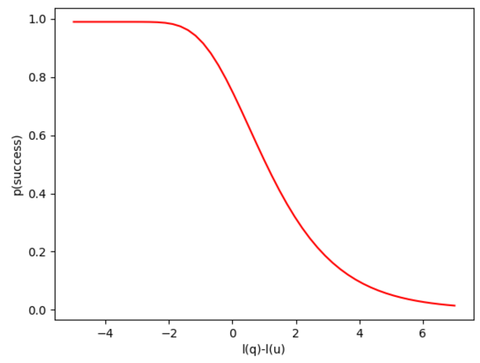

Figure 4: Gompertz curve used as a model to simulate student success probabilities.

Results

- The beginner student remains around A1 and A2 which is reflective of the student's current level.
- The intermediate student increases in CEFR level until level 3 (B2).
- The advanced student reaches an advanced or higher CEFR level.
- Agent tutor pushes the student to edge of their ZDP.
- Reward experiences a downward slope as the students reach their current level of vocabulary and are now being pushed to understand more advanced material.

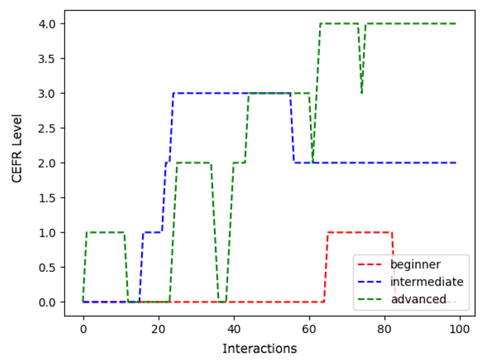

Figure 5: CEFR levels determined by the agent for students of varying levels of proficiency over 100 interactions

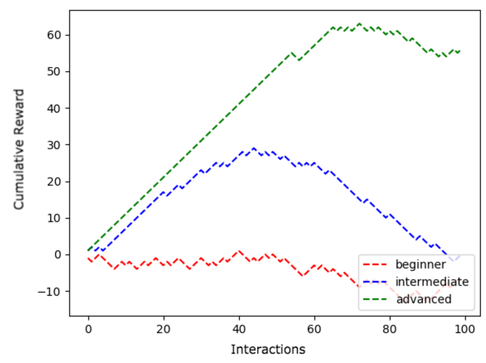

Figure 6: Cumulative reward earned by students of varying levels of proficiency from the agent over 100 interactions

Conclusions

- Effectively model a personalised curriculum for vocabulary acquisition using Q-Learning.
- Framework to extrapolate the difficulty and appropriateness of new material.
- Lays foundations for future pedagogically inspired RL architectures.
- Parallels can be drawn between the concept of ϵ-greedy and Krashen's Input Hypothesis or the $i+1$.
- The interactions between the agent and the environment in RL is analogous to the social interaction approach to language acquisition, specifically the equal importance of input and output.

Future Work

- Deploying the system on-line in order to collect user data will allow us to validate and improve our existing models.
- Incorporating memory and spaced repetition learning [2], a phenomenon initially documented by Ebbinghaus (1885), in order to optimise the policy and emulate cognitive processes.
- Deep learning models to approximate the Q-value will allow the system to capture additional signals pertinent to language acquisition.
- Adaptive reward model that reflects difficulty to encourage memory retention.
- Cognitively grounded models can also be applied to agents instead of students.
- Creating a dynamic environment guided by a curriculum grounded in pedagogically inspired RL may result in improved learning rates for the agent.

References

[1] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[2] Burr Settles and Brendan Meeder. A trainable spaced repetition model for language learning. In *ACL (1)*, 2016.

[3] Charles P Winsor. The gompertz curve as a growth curve. *Proceedings of the national academy of sciences*, 18(1):1–8, 1932.