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 (ZPD), 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 a 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).

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:

\[
Q(s, a) = Q(s, a) + \alpha[r + \gamma \max_a Q'(s',a') - Q(s, a)]
\]

(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 model level 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).

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:

\[
P(\text{success} | u,q) = 1 - \exp(-6 \exp((-u/q - l(u)(l(q) - l(u)))))
\]

(2)

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 ZPD.
- Reward experiences a downward slope as the students reach their current level of vocabulary and are now being pushed to understand more advanced material.

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