Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 2734–2743

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Detection of Reading Impairment from Eye-Gaze Behaviour using Reinforcement Learning

Harshitha Nagarajan[1], Vishnu Sai Inakollu[1], Punitha Vancha[1], Amudha J[1] *

[1]*Department of Computer Science and Engineering, Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham, India*

Abstract

Experimental psychology and neuroscience reveal that decision-behavior plays a dominant role in human-selective-attention when it comes to reading, object and scene detection. Difficulties in reading are easily reflected by eye-movement patterns. Hence, modelling eye-gaze behaviour for normal readers and people with reading impairments can greatly help in contrasting reading strategies used, which can in turn help in early identification and diagnosis of impairments such as dyslexia. This paper introduces a novel method of formulating a reinforcement learning model that is explainable, and can obtain the sequence of gaze targets based on recorded observations of dyslexic and non-dyslexic children. Results reveal that despite being a less sophisticated model, it is able to obtain the optimal reading policy of the ideal reader, from a set of good and poor readers with the help of a strong reward system and Q-Learning agent.

Keywords: reinforcement learning; eye-gaze behaviour; eye-tracking; modelling gaze behaviour; dyslexia;

1. Introduction

Dyslexia is a lifelong impairment that can impact children and adults alike. Studies [1] suggest that dyslexic youngsters confront numerous challenges in their educational interactions and social environments. They also have disappointments and low self-esteem as a result of their lack of accomplishments, notably in academics, which may have an impact on their long-term life prospects. It is critical for parents to be aware of dyslexia and its effects on their children in order to ensure their children's long-term development.

* Corresponding authors
E-mail addresses: harshithan1301@gmail.com, ivsvishnusai@gmail.com, punithareddy14@gmail.com, j_amudha@blr.amrita.edu

1877-0509
10.1016/j.procs.2023.01.245

1.1. Reinforcement learning for analyzing eye-gaze behavior

Reinforcement learning (RL) is a common Artificial Intelligence (AI) mechanism that is incorporated in many cognitive models. It is a computer approach to understanding and automating goal-directed learning and decision-making in dynamic tasks. The RL mechanism has lately grown in popularity and success as a tool for analysing human behaviour in a number of dynamic activities. The solution to an RL problem can either follow a model-based or a model-free approach. A model-based approach, also known as "planning", is followed when the internal operations of the environment is known. Algorithms under this category include dynamic programming policy evaluation, value iteration, policy iteration etc. Model-free approaches are used when the internal dynamics of the environment, such as the state transitions, are not known. The algorithms obtain "training data" from observing interactions between the agent and its environment. Popular algorithms under this approach include Monte Carlo methods, Temporal difference prediction, Sarsa, Deep Q Networks, Actor-critic architectures, Q learning and much more.

Apart from its success in popular gaming industries [2], [3], finance sector [4], healthcare [5], and a plethora of other applications, exploring the utility of RL in helping to achieve human-like intelligence has only recently begun. Gathering food, avoiding predators, manufacturing tools, and social interaction all rely on visual information for life, thus actively determining where to direct our eyes is a necessary skill. Eye tracking can be used to analyse human eye gaze behaviour and help interpret a plethora of information about people and their surroundings. Reference [6] shows how eye gaze behaviour can be used to gauge a person's degree of stress. Eye tracking can be used to measure parameters like blink frequency, pupil size, and fixation qualitative score of mental stress. Another study[7] employed eye movement data to forecast the image's valence and divide it into three categories: pleasant, neutral, and unpleasant. Features like fixation count and number of blinks were used to characterise the scene. When it comes to the way humans read, everybody has their own strategy to interpret the text in a way that works best for them. An adult may be able to skim through a document faster than a child owing to experience. Perceptual grouping of words would help an experienced reader "skip" adjacent words, whereas an inexperienced reader would try and focus on every word. Studies have shown that eye tracking movements is an effective way to identify people with reading difficulty.

Analysing eye gaze behaviour can aid with the identification of dyslexia. Visual attention span is low for the dyslexic. These characteristics can help identify the onset or presence of reading impairment, and eye-tracking data helps achieve the same. Hence, building an explainable model for eye-gaze behaviour using Reinforcement Learning, for reading tasks can prove to be very useful. Finding the respective optimal policies for "good" and "poor" readers can help describe the differences in reading strategies followed, thereby help in diagnosis and treatment of any traces of reading impairments.

Therefore, this paper introduces a Reinforcement Learning (RL) framework to model human gaze behaviour for the task of reading. The person reading a piece of text can be considered a goal-directed agent, with their fixation locations with respect to time, can be considered as a sequential decision process of the agent. The agent must successfully reach the last word of this text stimulus, considering that a good amount of words have been read. At each state, observable parameters with respect to the environment can include the current gaze location, target gaze location, and the duration of fixation. It is a proven fact that a major difference between dyslexic and non-dyslexic children was that the fixation durations were higher in number and longer for dyslexic readers. The definitive **goal** of the agent is to *maximize internal rewards* through *changes in gaze fixation*. A strong reward system is defined based on what the agent should and should not do.

1.2. Objective

The aim is to determine an optimal strategy used by "good readers" for reading a passage. Eye-tracking data from a custom dataset collected from 7-year old children with and without dyslexia, is used to determine gaze behaviour. Subjects were highly suitable for examining gaze strategies, as children at the age of 7 tend to read "more carefully" than adults, by reading each word in a sentence without skipping any. The use-case can be categorized as a model-free control problem, where- given a set of observations, a deterministic optimal policy is obtained. Therefore, a Q-Learning agent is used to obtain the optimal policy for a set of good readers and poor readers.

This paper sheds critical light on the capacity of reinforcement learning as an effective method to model and

quantify eye-gaze movement behaviour for reading tasks. Our main contributions include:

1) To provide a simple, efficient and scalable framework for sequential word-level reading using RL
2) Formulating an RL agent specifically for the task of reading, such that it can help identify visual impairments. Finding such an optimal policy can further be used to classify readers as "effective" or "non-effective" (good or poor) in its deployment phase.

Apart from this section, the rest of the paper has been organised as follows- Section 2 describes relevant previous works, Section 3 explains the thought process behind formulating the use-case as an RL problem, Section 4 describes implementation details, Section 5 displays the results obtained, Section 6 discusses limitations and improvements and Section 7 concludes the paper.

2. Related works

Raw data from any eye-tracking device typically goes through pre-processing to extract meaningful features, such as fixation points, duration, saccades, areas of interest etc. Algorithms based on vector quantization can be used for this purpose, as highlighted by [8,9]. A detailed account on dataset creation for recording eye-gaze metrics for dyslexic and non-dyslexic students has been provided in [10].

Previously, there have been algorithmic and machine learning approaches to distinguish between good and poor readers using eye-tracking data. For example, the dispersion threshold algorithm can be used to identify fixation locations of the eye [11]. The paper discusses a correlation study that was done between the two categories of reading abilities, and the significant differences in reading behaviours are recorded and displayed. Feature extraction translates the raw eye movements data into fixation and saccades, and this information is perceived by the brain only during gaze fixation. Alternatively, a simple KNN-classifier can be used to classify a student as a good or a poor reader, based on recorded eye-metrics[12]. It is proved that difficulties in reading are reflected in eye movements.

Reinforcement Learning can be used to model human behaviour for dynamic tasks in particular, those that are temporal in nature. And it can be modelled close to the way humans learn, as exhibited by [13], where an RL model learns a popular dynamic control task called Dynamic Stock and Flows (DSF).

In relevance to RL being deployed for eye-gaze movement tracking in visual tasks, two essential questions for which answers need to be confirmed are- (1) How does a visual system select sequential decisions for gaze actions in tasks of reading? (2) Is visual attention allocated in a strictly serial manner with respect to reading sentences? The next two paragraphs provide answers to these questions.

How visual systems use perception and belief states to select gaze targets in a sequential manner when it comes to search tasks has been unclear. However, it has been established that building a quantifiable, computational model can help understand and study how subsequent actions are selected to achieve a goal [14]. The work presented derives policies for 'myopic' (greedy search) and 'planned'(ideal) observers and the scanpaths for the same are predicted, all formalized within the framework of a Markov Decision Process (MDP). This study confirms that the selection of gaze shifts are in-fact, best described by a probabilistic planning strategy.

The study done in [15] proves that reinforcement learning agents strongly prefer serial word processing. Such AI agents can learn fairly sophisticated eye-movement behaviours in order to perform the task of identifying the next sequence of words to read, or the next set of actions to take. These behaviours include directing the eyes toward the centres of words because this viewing location permits the most rapid identification of words.

For an MDP without a reward function and transition probabilities defined, learning a policy from recorded sequences of fixations can be done in multiple ways. One such method is highlighted in [16], where the optimal policy is learned through least-squares policy iteration (LSPI) by using linear projections of state-action descriptors. Features for state descriptions are generated with the assumption that humans tend to fixate their eye gazes at the centre of the object.

Another interesting thread of research done in this field is to do with the concept of Inverse Reinforcement Learning (IRL). The biggest advantage of IRL is that it offers a substitution for manual specifications of reward functions, for solving an MDP. The IRL algorithms learn reward functions, given optimal policies [17].

Widespread research has been done with respect to IRL being used to analyse eye-tracking data. A novel IRL model for predicting search fixation scanpaths for visual search tasks has been proposed in [18]. IRL is used to jointly

recover the reward function and policy used by people during visual search, trained on the COCO-Search18 dataset. States are represented as dynamic contextual beliefs (DCBs), which update beliefs about objects to obtain an object context that changes dynamically with each new fixation. Modelling the eye gaze behaviour of wheelchair drivers can also be done using IRL, as proposed by [19]. The work depends on risk factors to which drivers pay attention to, and these are represented as features for the model. Analysis of the learned models showed that experienced drivers tend to pay attention to blind corner more than novice drivers. Hence, similar contexts can be derived for visual reading tasks in the sense of modelling what and what not the reader pays attention to.

It was observed that most of the research involving forward reinforcement learning for eye-gaze behaviour is directed towards visual-search, object detection tasks and not many of them deal with the *analysis of reading tasks*. Although there are theoretical descriptions of problem formulation and state representations, there is a dearth of open source implementations of the works, which makes it harder for beginners to reproduce and improve upon. With respect to Inverse Reinforcement Learning, the lack of built-in libraries and frameworks make it difficult for rapid prototyping.

Hence, this paper proposes a simple yet effective method of formulating an MDP for the task of reading, such that the reading policy of the ideal reader can be obtained. In addition, optimal policies of both good and poor readers can be identified and analysed. This will not only help establish the capacity of RL as an early identifier of Reading Impairments, but will also provide a scalable direction for enhancing research in this field.

3. RL problem formulation

In this section, we discuss the considerations and assumptions taken into account, while modelling the problem as a Markov Decision Process (MDP).

For a quick overview on the basics, an MDP is formally represented as a 4-tuple (S, A, r, T). For an RL agent interacting with its environment, S denotes the state space, A denotes the action space, r stands for the immediate reward awarded for taking an action, and T represents the transition probability of the agent moving from one state to the next, upon taking an action. It is crucial to have a strong reward function r, as the reward r(s, a); where s ∈ S and a ∈ A; helps establish the importance or unimportance of an action, which ultimately helps the agent learn an optimal policy that will direct its course of actions in an idealistic manner.

The formulation of an MDP requires the understanding and definition three important components of an MDP (Fig. 1), the ***Environment***, ***States and Actions***, and the ***Reward System***.

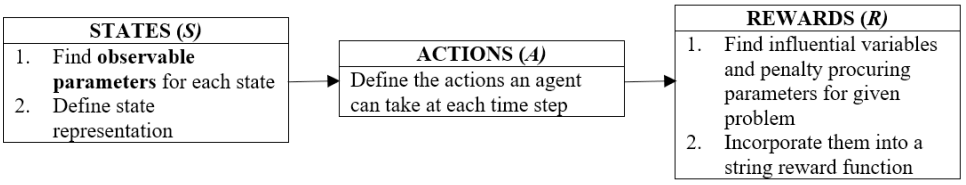


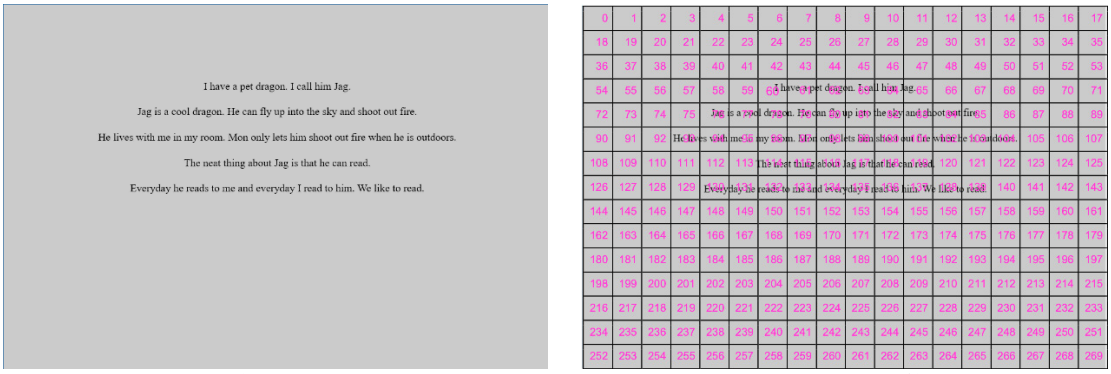
Fig. 1. **General workflow** for formulating a simple RL problem

3.1. The environment

For a person reading a piece of text on screen, visual cues of the presence or absence of a word are captured through the eyes and interpreted by the brain. Hence, the *text stimulus* on the screen is to be considered as the *environment*.

Human vision is hierarchical in nature. The **automatic perceptual grouping** performed by the human brain is what allows us to identify and detect areas of interested objects as a whole, rather than a pixel-by-pixel identification approach. In a similar sense, visual attention during the task of reading is not attracted by a single fixed point, but rather by an *area of interest*. Owing to this, the environment can be divided into a grid of Text-Cells (TC) and Non-

Text Cells (NTC) as shown in Fig. 2, where TCs are cells containing textual words and NTCs are empty spaces.



(a) (b)

Fig. 2. A depiction of the stimuli and environment used in this study (a) The text stimulus provided to the subjects [*image credit to [8]*] (b) The text stimulus divided into 15×18 grids, and numbered sequentially

3.2. States and actions

At every moment of reading, observable parameters for the agent are the *current gaze location* and features describing the response of the individual to the stimulus, such as *fixation duration*. Therefore, the state representation can be modelled as a combination of current gaze location and fixation duration. The action must be able to depict where the reader directs his/her gaze next, hence, the action in this case is the *next location of gaze* in the image(stimulus).

3.3. Designing the reward system

The attempt at defining a reward function for an agent reading text stimuli in this study has been made by considering distinguishing parameters between good and poor readers, and with an assumption about an idealistic reading strategy, i.e. the scenario of reading one single line of text progressively from left to right.

Deciding on the next course of action should ideally prompt the agent to fixate gaze locations more towards the right. The agent must also move progressively towards the end of the sentence, making sure to read all the words in the sentence one after the other. Positive rewards must be awarded in these cases.

Penalty procuring situations would occur when the next gaze locations are targeted towards the left, which would mean that the agent is traversing to the left to revisit a word. It is also not ideal in cases where the agent fixates gaze for too long, wanders off gaze attention to empty cells (NTCs) and has a longer scan path length in general.

Another variable to account for is the cell visitation frequency. Once an NTC is visited, a “good” reader would perceive that the cell is empty, hence avoid revisiting the area. However, in the contrary case, revisiting an NTC would mean that the reader/agent is unable to interpret the absence of text. Hence, the negative reward must be amplified by the cell visitation frequency. A detailed version of the reward system from an implementation perspective has been provided in the coming sections.

A visual representation of the formulated MDP has been provided in Fig. 3. The environment is the text stimuli grid, action is the next/target gaze location, and the state is represented by the current gaze location and fixation duration.

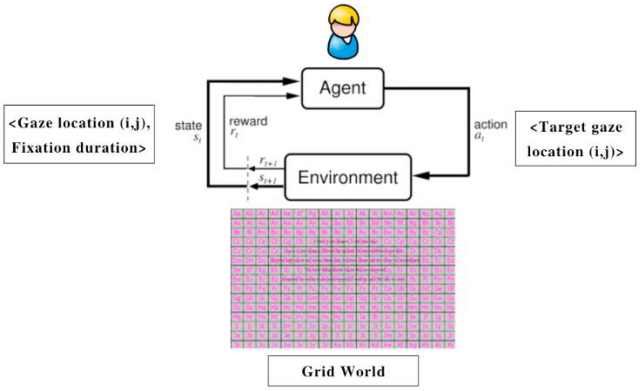
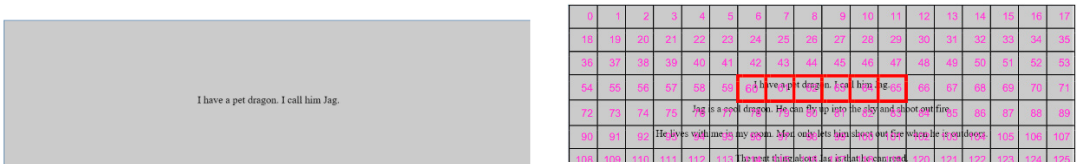


Fig. 3. A representation of the MDP that has been formulated

4. Implementation details

The RL model was built considering only the first line of text in the stimulus (Fig. 4) This was done to help explain the solution in a clearly defined manner.



(a) (b)

Fig. 4. (a) The first line of the text stimulus; (b) Corresponding grids to the first line highlighted in red [grids 60-65]

4.1 Data preprocessing

The dataset consists of eye movement metrics for reading a paragraph of text. Subjects belonging to the age group of 7 were given reading "stimuli", and their eye-gaze data was recorded [8]. These data samples are taken from 15 non-dyslexic children and 5 children with reading problems or dyslexia. Out of the 5 children with reading problems, there are 2 extreme cases of dyslexia and 3 children with moderate reading problems.

Out of the many metrics recorded, useful columns for the agent were extracted. These included the *participant ID, fixation duration, and gaze location* in the form of (x, y) pixel co-ordinates. The raw pixels were mapped to suitable grid indices using simple linear algebra.

The fixation duration was encoded as 0 if it was lesser than or equal to 200ms, 1 if it was between 200-1500 ms, and 2 if it was greater than 1500ms. 0,1 and 2 represent a low, normal, and high fixation duration respectively. Scan paths corresponding to the reading the first line of stimulus were manually extracted for each participant. Finally, to create a balanced dataset, more number of samples were generated for "poor" reader instances, making a total of 10 samples, as opposed to having 5 recorded ones earlier.

4.2 Creating the environment

An episodic environment simulator that can be started from the start after the end of each episodic event was created using Python's OpenAI Gym library. The state-action spaces are defined considering the entire grid environment, and the reward function is defined in section *4.3*. The episodic traversals occur within a *step()* function

described in the environment class. This function is responsible for calculating and returning the reward, along with updated state information for each action that is taken by the agent.

4.3 Defining the reward system

For every timestep in an episode, an agent is awarded a reward in three phases: (1) Based on the nature of Cell Visitation (2) Based on the fixation duration observed (3) Based on the scanpath length.

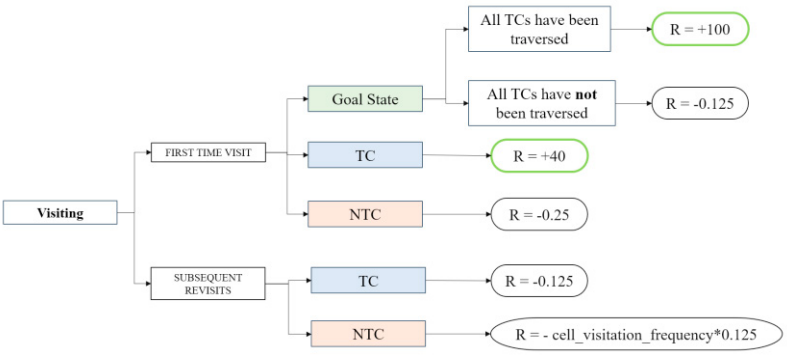


Fig. 5. Phase 1 of the reward system

An agent can visit a cell for the first time, or make subsequent revisits to the same cell. As described in section *3.3*, positive and negative rewards are awarded according to parameters considered. Therefore, in phase 1 of the reward system, a first-time visit can land the agent at the goal state, a text cell (TC), or a non-text cell (NTC). If the agent visits a goal state for the first time, it may or may not guarantee that the agent has reached the goal state progressively. Hence, a positive reward of 100 is awarded only when all TCs have been traversed. Else, the agent receives a negative reward of 0.125. If the agent visits a TC for the first time, a positive reward of 40 is awarded. If the agent visits an NTC, a penalty of -0.25 is procured. In the case where the agent is making revisits, there can be two sub-cases again- TCs or NTCs. If the agent is revisiting a TC, a negative reward of 0.125 is awarded. If the agent revisits an NTC, the negative reward of 0.125 is amplified by the ***cell visitation frequency*** variable, which keeps track of the count of visitation for each cell (Fig. 5).

The second phase depends on the fixation duration parameter (Fig. 6.a). The fixation duration can either be low (0), normal (1) or high (2). A higher negative reward of 0.25 is given when the fixation duration is high, -0.125 if it is normal and no reward if fixation duration is low.

Fig. 6. (a) Phase 2 of the reward system (b) Phase 3 of the reward system

Finally, the third phase accounts for the length of scanpath (Fig. 6.b). Considering fixation points for the first line of text only, the length of scanpath for good readers was observed to be 8-10 from preliminary data analysis. Therefore, a threshold of 15 is taken. Length lesser than 15 receives no reward. As the agent crosses the threshold mark, -0.01 gets added to the reward.

4.4 Training the Q-learning agent

During the training phase, the algorithm has access to the recorded sequences of eye-tracking results from reading the stimulus, which are used to learn a policy that is best followed by them while reading. Q-Learning is a model-free reinforcement learning algorithm that can find the optimal policy of actions for a given MDP based on maximizing the expected value of the total reward over any and all successive steps starting from the current state [14]. This decision-maker associates shift of attention-actions to cumulative rewards with respect to the task goal, which here is to progressively reach the end of the sentence.

Recalling the aim of this paper, an optimal reading policy is to be obtained from the samples recorded. Along with that, the local optimal policies for the set of good readers, and the set of poor readers can be obtained individual to observe differences. Therefore, 2 Q-Learning agents are trained to learn the behaviours individually and then compare. One agent learns the optimal policy based on the processed data from good readers, and the other agent learns the optimal policy of the poor readers.

Two important hyperparameters involved are the learning rate α, discount rate γ. The learning rate intuitively determines to what extent the agent will learn something new. Discount rate determines the importance of future rewards. After hyperparameter tuning, the learning and discount rates used for this study were $\alpha = 0.4$ and $\gamma = 0.99$ respectively.

5. Results and discussion

5.1 Policy for good readers

For the chosen hyperparameters, the optimal reading policy obtained for good readers was ***[198,60,61,62,63,64,65],*** with the maximum reward score of **298.625** . (The numbers represent the grid locations as shown in Fig 3.b). This is the most ideal, deterministic policy an agent could have learned, as the task is only to read a single line of sentence from left to right, traversing each word progressively in a sequential manner. The agent is able to learn this well, and hence, it also proves the fact that good readers follow an ideal reading strategy (Fig 7).

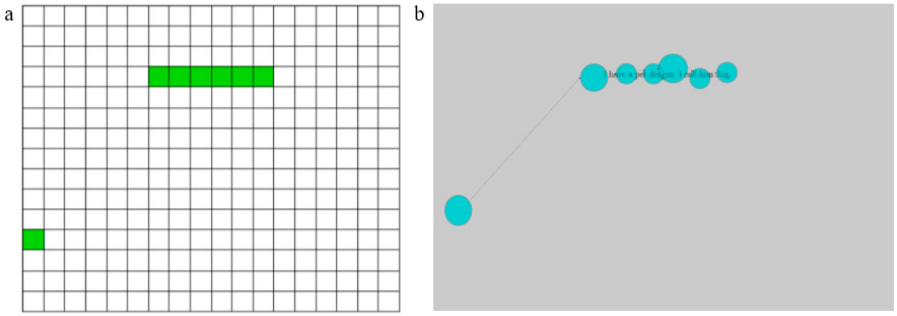


Fig. 7. Optimal policy for GOOD READER. The policy depicts that a TC is only visited once. (a)Grid view with TC cells highlighted in light green, visited once each (b) Gaye Fixation visualization on stimulus

5.2 Policy for poor readers

For the set of poor readers, the optimal policy is ***[198,60,61,63,64,65,62,63,64,65]***, with a score of **290.0**. It is clearly noticeable here that the poor reader policy is slightly different than that of the good reader, as more than one TC is being revisited multiple times(Fig.8). The subtle difference is indicative of the presence of reading impairments.

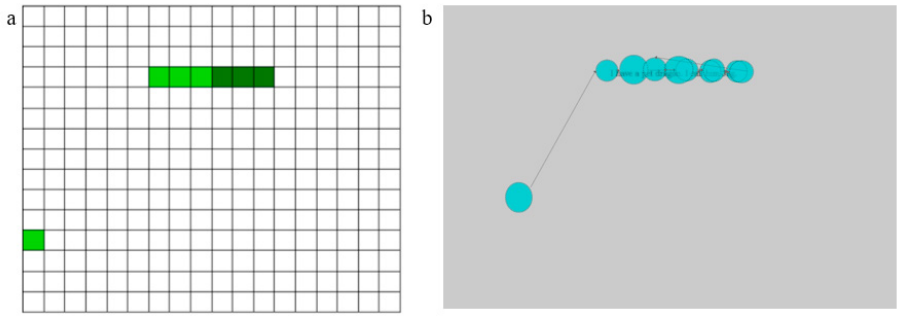


Fig. 8. Optimal policy for POOR READER. The policy depicts that TCs are visited multiple times sometimes. (a)Grid view with TC cells highlighted in light green (visited once) and dark green (visited more than once) (b) Gaze Fixation visualization on stimulus

5.3 Robustness of the Reward System

The proposed reward system reflects the differences in good and poor readers by observing individual scores obtained for each subject (Table 1). A reward **score >290** is obtained for **good** readers, and **<290** is obtained for **poor** readers. Fig. 9 shows the consistency of a high reward score awarded to good readers(depicted in green), and a fluctuation under the threshold of 295 for poor readers.

Table 1. Examples of the scanpaths of good and poor readers, along with associated scores from the reward system

Reader	Scanpath	Score
Good	[199, 97, 79, 60, 66, 156, 63, 61, 64, 63, 61, 62, 64, 65]	296.75
Good	[198, 96, 60, 61, 62, 62, 61, 61, 62, 62, 63, 64, 65]	297.125
Good	[236, 115, 79, 60, 61, 80, 62, 63, 64, 65]	297.75
Poor	[198, 60, 118, 78, 61, 83, 60, 76, 61, 118, 78, 60, 65, 61, 206, 206, 6, 80, 80, 60, 62, 61, 62, 44, 62, 62, 60, 84, 61, 216, 0, 17, 82, 62, 62, 80, 8, 63, 93, 79, 61, 62, 63, 60, 116, 99, 64, 180, 100, 222, 100, 0, 64, 61, 81, 101, 136, 80, 63, 64, 60, 77, 78, 97, 61, 98, 64, 65]	276.125
Poor	[198, 60, 60, 61, 76, 76, 65, 61, 61, 62, 44, 62, 44, 45, 45, 45, 45, 45, 45, 62, 45, 62, 63, 64, 45, 46, 47, 45, 28, 45, 63, 62, 114, 82, 63, 45, 45, 45, 45, 64, 63, 82, 83, 46, 46, 64, 65]	257.375
Poor	[198, 60, 80, 77, 101, 101, 77, 77, 101, 60, 62, 101, 101, 61, 63, 64, 64, 45, 45, 64, 65]	290

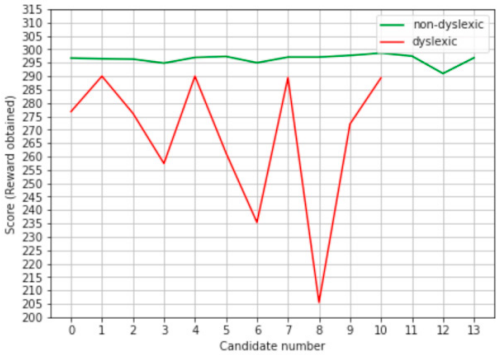


Fig. 9. The graph depicting the rewards obtained for good(non-dyslexic) and poor(dyslexic) subjects. The reward scores for poor readers do not exceed 290, and the lowest possible reward from the samples is 205.0.

6. Limitations and future directions

This work can be extended further in the future by improving the reward function in terms of generalization over multiple stimuli. This can also be attempted utilising IRL. The paper discusses a scenario where only the first line of

text is considered, this could be extended to reading the entire paragraph. That would in turn increase the state-action space and hence, suitable deep Q network strategies may be employed. Coming to the deployability aspect, it could be transformed into an online tool which could therefore help young children diagnose their reading conditions by creating classification modules based on reward scores.

The salient determining factor for distinguishing between good and poor readers is the *fixation duration*, which is represented in the state space of the model. This aspect can be improved upon to include multiple other visual attention parameters as well, such as saccade direction, fixation count etc. Rather than using TCs for all words, noteworthy words depicting an area of interest (AOI) can be found, and represented as TCs instead. The reward function, although strong, is specific to the text stimulus considered here. It is a stimuli-centric approach, wherein the positioning of text-cells and non-text cells with respect to this stimulus directs the agent to learn an optimal policy.

7. Conclusion

This work focuses on building a simple yet quantifiable RL model suitable for describing the behaviour of gaze and thereby, drawing conclusions of whether a person is a good or poor reader while making use of eye-tracking data from dyslexic vs non-dyslexic children. The intricate differences in reading policies between good and poor readers are successfully captured by the reward system. Although only one line of scanpaths is considered in this model, it effectively managed to differentiate between the two categories of readers, through the reward scores. The success of this model opens up research avenues on other similar use-cases such as the case a TV News reader recruitment process, as it heavily depends on speed and efficiency of the reader.

References

[1] Abd Rauf AA, Ismail MA, Balakrishnan V, Haruna K. Dyslexic Children: The Need for Parents Awareness. Journal of Education and Human Development 2018;7. https://doi.org/10.15640/jehd.v7n2a12.

[2] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing Atari with Deep Reinforcement Learning. n.d.

[3] Schrittwieser J, Antonoglou I, Hubert T, Simonyan K, Sifre L, Schmitt S, et al. Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model 2019. https://doi.org/10.1038/s41586-020-03051-4.

[4] Liu X-Y, Rui J, Gao J, Yang L, Yang H, Wang Z, et al. FinRL-Meta: A Universe of Near-Real Market Environments for Data-Driven Deep Reinforcement Learning in Quantitative Finance 2021.

[5] Yu C, Liu J, Nemati S. Reinforcement Learning in Healthcare: A Survey 2019.

[6] Jyotsna C, Amudha J. Eye Gaze as an Indicator for Stress Level Analysis in Students. 2018 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2018, Institute of Electrical and Electronics Engineers Inc.; 2018, p. 1588–93. https://doi.org/10.1109/ICACCI.2018.8554715.

[7] Tamuly S, Jyotsna C. Tracking Eye Movements To Predict The Valence of A Scene. n.d.

[8] Akshay S, Ashin VP. Saccade Calculation Algorithm Based On Vector Quantized Fixation. vol. 10. 2018.

[9] Akshay S, Adarsh M. Vector Quantization Based Algorithm to Calculate Fixation from Raw Eye Tracking Data. vol. 10. 2018.

[10] Navya Y, SriDevi S, Akhila P, Amudha J, Jyotsna C. Third Eye: Assistance for Reading Disability. Advances in Intelligent Systems and Computing, vol. 1118, Springer; 2020, p. 237–48. https://doi.org/10.1007/978-981-15-2475-2_22.

[11] Ramachandra CK, Joseph A. IEyeGASE: An intelligent eye gaze-based assessment system for deeper insights into learner performance. Sensors 2021;21. https://doi.org/10.3390/s21206783.

[12] Sowmyasri P, Ravalika R, Jyotsna C, Amudha J. An Online Platform for Diagnosis of Children with Reading Disability. n.d.

[13] Dutt V. Explaining Human Behavior in Dynamic Tasks through Reinforcement Learning. n.d.

[14] Hoppe D, Rothkopf CA. Multi-step planning of eye movements in visual search. Scientific Reports 2019;9. https://doi.org/10.1038/s41598-018-37536-0.

[15] Jiang M, Boix X, Roig G, Xu J, van Gool L, Zhao Q. Learning to Predict Sequences of Human Visual Fixations. IEEE Transactions on Neural Networks and Learning Systems 2016;27:1241–52. https://doi.org/10.1109/TNNLS.2015.2496306.

[16] Liu Y, Reichle ED, Gao DG. Using Reinforcement Learning to Examine Dynamic Attention Allocation During Reading. Cognitive Science 2013;37:1507–40. https://doi.org/10.1111/cogs.12027.

[17] Arora S, Doshi P. A survey of inverse reinforcement learning: Challenges, methods and progress. Artificial Intelligence 2021;297. https://doi.org/10.1016/j.artint.2021.103500.

[18] Yang Z, Huang L, Chen Y, Wei Z, Ahn S, Zelinsky G, et al. Predicting Goal-directed Human Attention Using Inverse Reinforcement Learning 2020.

[19] Maekawa Y, Akai N, Hirayama T, Morales LY, Deguchi D, Kawanishi Y, et al. Modeling Eye-Gaze Behavior of Electric Wheelchair Drivers via Inverse Reinforcement Learning. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020, Institute of Electrical and Electronics Engineers Inc.; 2020. https://doi.org/10.1109/ITSC45102.2020.9294255.

[20] Sutton RS, Barto AG. Reinforcement Learning: An Introduction Second edition, in progress. n.d.