

# A Cognitive Learning Model that Combines Feature Formation and Event Prediction

Eman Awad and Fintan Costello

School of Computer Science and Informatics, University College of Dublin  
Belfield, Dublin 4, Ireland

eman.awad@ucdconnect.ie , fintan.costello@ucd.ie

**Abstract.** This paper presents a computational model that addresses two central aspects of cognitive learning: feature formation and prediction. Most current cognitive models fail to account for these two aspects of learning; this model, based on frequentist probability theory, provides a unified account of feature formation and temporal prediction. We describe a computational model that learns categorical and temporal relationships between events to form features and make predictions about future events, with the aim of simulating human learning and prediction processes. With this novel learning mechanism, we aim to provide for further insight into the complex cognitive process of human reasoning.

## 1 Introduction

Human learning is a complex process involving different cognitive sub-processes. Some studies have demonstrated that the formation of categories based on the probabilistic occurrence of often complex features is central to human learning and inference [8], [26]; humans form features from perceived events and learn the categorical relationships between them, which are then used to make predictions [22], [24], [25], [9], [13]. Other studies have demonstrated that learning temporal relationships between events (associations between features perceived across time) is also vital to human learning and prediction [21], [20], [19], [23], [1].

With this paper we aim to present a computational model, which integrates categorisation and temporal prediction or conditioning processes as a unified account of learning. Although categorisation is key to the learning process, typical categorisation models are unable to realistically simulate the human temporal learning process as they only categorise static features [24], [25], [9] [13], [17], [22]. This is insufficient as humans continually observe a dynamic environment, thus a realistic model should incorporate the ability to learn temporal and categorical relationships.

Category formation and event prediction are both instances of probabilistic reasoning; there is evidence that probabilistic reasoning in humans follows frequentist probability theory [2], [3], a mathematical process involving the evaluation of past events in terms of their significance in predicting a future event.

The model in this paper follows the approach taken by Costello and Watts [3], [2]. The experimental results in these papers repeatedly showed that the biases observed in people’s estimated probability could be explained by a model where people’s reasoning follows frequentist probability theory but is subjected to random noise. This evidence suggests that people reason in a manner consistent with the frequentist probability theory and it is this theory that we will apply on our model.

The aim of this research is to create a model designed to form features, learn the categorical and temporal relationships between them, and predict future events following frequentist probability theory. The prediction mechanism in our model involves two stages: forming features and then calculating predictions. In the first stage, relationships between events or combinations of events will be evaluated in a process referred to as the model’s statistical evaluation process. The events that represent a reliable relationship are selected and formed into features. The temporal and categorical relationship between constituent events of the features is also learned. This process determines which past events are significant in predicting future events. In the second stage, the model combines predictions from relevant reliable features to calculate the overall predicted probability value for a given event.

The first section of this paper will delineate the evaluation process used to form reliable features; a description of the prediction mechanism will also be provided. The second section outlines an overview of the model design. Finally, we will test the model’s performance to evaluate the prediction mechanism and present the analysis of the results.

## 2 Reliability: Feature Formation

Our model is underpinned by the theoretical assumption that the human learning process follows frequentist probability theory; humans evaluate past experiences based on the statistical reliability of previously learned features to inform their predictions about what may occur next in the environment [12], [18]. We argue that this evaluation process is based on statistical reasoning to form predictive features, which will be used in the prediction calculation. We illustrate this formation of features using an example in Fig(1), where we aim to evaluate the reliability of the relationship between two events. This figure shows the occurrence of events  $A$  and  $S$  across time. After the first occurrence of  $A$  followed by  $S$ , an event  $N$  was created to represent the possibility of a relationship between  $A$  and  $S$ ; the probability of  $N$  will change as it occurs again. For example, consider time  $t=8$  with five events perceived in this period. Event  $A$  occurred three times in this period; in two of the three times, event  $A$  was followed by event  $S$  after one second. At  $t=8s$ , the overall probability of  $S$  occurring is two out of eight;  $p(S) = 0.25$ . The question is whether the occurrence of  $S$  after  $A$  can be explained by the base rate occurrence of  $S$  alone;  $P(S) = 0.25$ . If it cannot be explained by  $P(S)$ , we assume a reliable causal relationship between  $A$  and  $S$ . Otherwise, the relationship between the two events is not reliable and  $A$  does

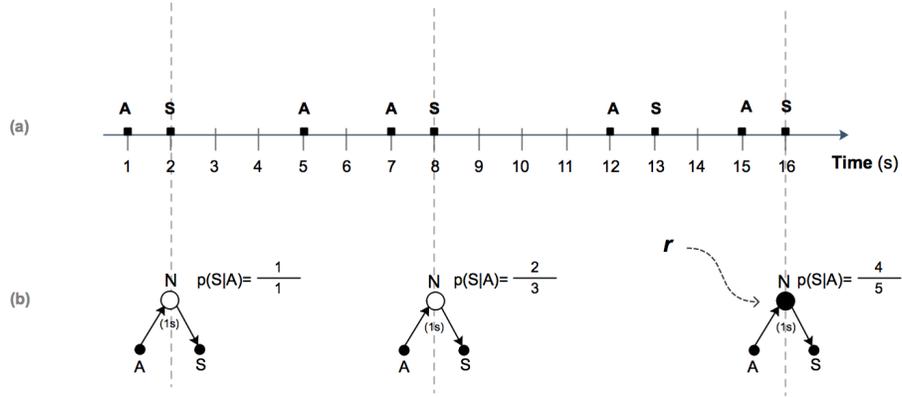


Fig. 1: (a) illustrates a series of events occurring across time. Time is measured in seconds. (b) shows the frequencies of events changing across time. At time 8s, the binomial result was higher than (0.05), indicating the unreliability of A in predicting S. However, at time 16s, the result was less than (0.05), indicating the reliability of feature A in predicting S. Filled circle represents reliable predictor N, indicated here as (r).

not predict S. Based on frequentist probability theory, the binomial probability function, is used to determine the reliability of the relationship [11]. We calculate this using the following equation:

$$Bin(k, x, p) = \binom{x}{k} p^k (1 - p)^{x-k} \quad (1)$$

Where  $x$  represents the number of times event A occurred (which may or may not be followed by event S),  $k$  represents the number of times that the consequent event S followed the antecedent event A, and  $p$  is the base rate probability of the consequent event  $P(S)$ . This function represents the chance of drawing a sample of  $x$  events from a population where  $p = P(S)$ , and of which  $k$  of them are instances of S.

Based on frequentist probability theory, if the binomial probability result  $Bin(k, x, p)$  is less than or equal to the statistical significance level of (0.05), it indicates that the occurrence of S after A cannot be explained by the base rate occurrence of S,  $P(S) = 0.25$ , and we can then deduce that the relationship between A and S is causal and reliable. However, if the binomial result is higher than 0.05, then there is insufficient evidence to prove that the relationship is reliable at that time.

At time = 8s in Fig(1), the binomial result  $Bin[2, 3, 0.25] = 0.14$  is higher than the significance level (0.05), indicating that there is not enough evidence to prove that there is a causal relationship between A and S. As such, event A is not a reliable predictor of S at this time.

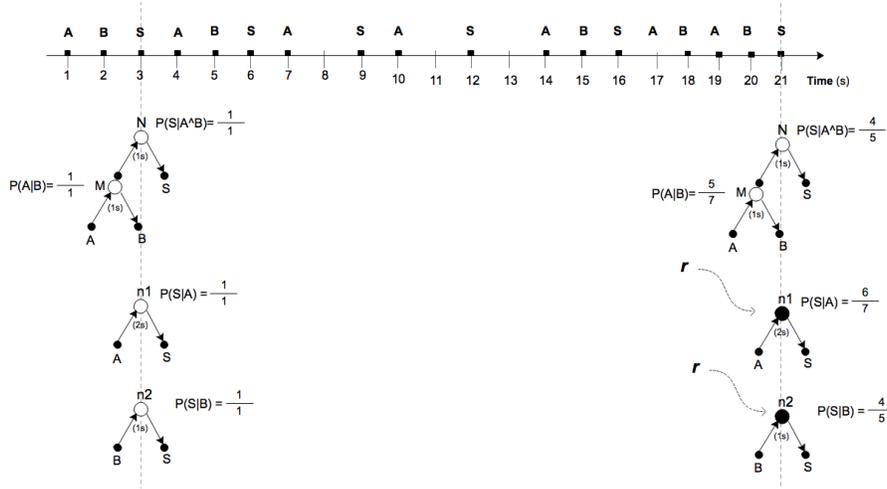


Fig. 2: An example of complex events occurring across time. Filled circles represent reliable predictors ( $n_1$  and  $n_2$ ), indicated here as ( $r$ ).

If we consider time  $t = 16s$ , event A occurred five times within this period; in four of these five times, A was followed by event S after one second. The question again is whether this occurrence of S after A can be explained by the base rate occurrence of S;  $P(S) = 0.25$ . Again, we calculate the binomial probability  $Bin[4, 5, 0.25] = 0.01$ , which is lower than the significance level (0.05), indicating that there is enough evidence to prove that the relationship is reliable. Thus, at this time feature A is considered a reliable predictor of S; we designate this as ( $r$ ) in Fig(1). This example illustrates the formation of a reliable feature N, which indicates that if A occurs, there is  $\frac{4}{5}$  chance of seeing S one second later. (In this example the time interval between the events is one second. This interval was chosen for illustrative purposes. The model uses various different intervals, which can be larger than 1s.)

**Complex Features:** The earlier example assumed that we only have a single event as an antecedent and a single event as a consequent. In reality, we can encounter more complicated events that have more than a single event as an antecedent. In the case of a complex feature, the complex event N holds a complex antecedent that includes two sub-events or more: event A as antecedent and another event B as a consequent. The two sub-events, A and B are the complex antecedent of N, and the consequent of that event is S (as shown in Fig.2).

To form complex features, we must first determine if the complex event represents a reliable relationship or not. To do this, we use Eq.(1) where  $x$  represents the number of times the complex antecedent (A then B) occurred and  $k$  represents the number of times that the complex antecedent was followed by the consequent S. In determining the relationship between complex features and S,

there are three probabilities that we must test the occurrences against: (1) the probability of  $S$ ,  $P(S)$  which was exactly as defined previously in simple events; (2) the probability of event  $n_1$  i.e.  $S$  following the antecedent  $A$  alone;  $P(S|A)$ , and (3) the probability of event  $n_2$  i.e.  $S$  following the antecedent  $B$  alone;  $P(S|B)$ . In this case, the question is whether the occurrence of  $S$  after ( $A$  then  $B$ ) can be explained by the base rate occurrence of  $S$ ,  $P(S) = 0.28$ , or by the observed rate occurrence of the simple features  $n_1$ ,  $P(S|A)$ , and  $n_2$ ,  $P(S|B)$ . If it cannot be explained by these probabilities i.e.  $P(S)$ ,  $P(S|A)$  and  $P(S|B)$ , we deduce that there is a reliable causal relationship between the complex antecedent and  $S$ . Otherwise, the relationship is not reliable and the complex antecedent ( $A$  and  $B$ ) together does not specifically predict the occurrence of  $S$ .

As a working example, see Fig(2). Here we consider three different features that potentially predict the occurrence of  $S$ : a simple feature  $n_1$  (which predicts that  $S$  will occur 2 seconds after the occurrence of  $A$ ), a simple feature  $n_2$  (which predicts that  $S$  will occur 1 second after the occurrence of  $B$ ), and a complex feature  $N$  (which predicts that  $S$  will occur 1 second after the occurrence of the complex antecedent ' $A$  then  $B$ '). At time  $t = 21s$ , we can ask whether  $n_1$  and  $n_2$  are reliable features (whether they represent statistically reliable relationships between  $A$  and  $S$  or  $B$  and  $S$ ) by using the binomial function as before, taking the base-rate occurrence of  $P(S) = 0.28$ . The binomial test results are  $Bin[6, 7, 0.28] = 0.001$  and  $Bin[4, 5, 0.28] = 0.019$  respectively, indicating that both  $n_1$  and  $n_2$  are reliable features:  $A$  reliably predicts  $S$  after 2 seconds (with probability 0.85) and, independently,  $B$  reliably predicts  $S$  after 1 second (with probability 0.8).

When testing the complex feature  $N$ , which represents the complex antecedent ( $A$  and  $B$ ) predicting the consequent  $S$ , we ask whether the observed rate of occurrence of  $S$  after this complex antecedent can be explained by the independent occurrence of these two predictors  $A$  and  $B$ . If the observed rate of occurrence of  $S$  after this complex feature cannot be explained by the independent occurrence of these two separate predictors, this indicates that there is some further causal relationship between the specific complex event ( $A$  and  $B$ ) and the occurrence of  $S$ , and the complex feature  $N$  will be marked as a reliable predictor. At time  $t = 21s$  feature  $n_1$  predicts  $S$  with probability  $P(S|A) = 0.85$ , and feature  $n_2$  predicts  $S$  with probability  $P(S|B) = 0.8$ . The complex antecedent ( $A$  then  $B$ ) has occurred five times by time  $t = 21s$ , with four of those occurrences followed by  $S$ . This rate of occurrence is explained both by the predictor  $n_1$  ( $Bin[4, 5, 0.85] = 0.39$ ) and separately by the predictor  $n_2$  ( $Bin[4, 5, 0.8] = 0.4$ ). This indicates that there is insufficient evidence to prove that the relationship is reliable between the specific complex antecedent ( $A$  and  $B$ ) and the occurrence of  $S$ ; the integrated event  $N$  is not a reliable predictor of  $S$  (i.e. not a reliable feature).

In other situations, the complex feature  $N$  might be reliable. For example, consider a different situation, where the complex antecedent ( $A$  and  $B$ ) occurred five times. In all of those times,  $S$  followed this complex antecedent, after one second (i.e. the integrated event,  $N$ ). However, event  $A$  and  $B$  occurred frequently

without being followed by  $S$  ( $A$  occurred ten times and  $B$  occurred fifteen times). Consider the base rate of the consequent and the probability of individual events;  $P(S) = 0.16$ ,  $P(S|A) = 0.5$  and  $P(S|B) = 0.33$  respectively, the binomial result for the complex event  $N$  are  $Bin[5, 5, 0.16] = 0.0001$ ,  $Bin[5, 5, 0.5] = 0.003$ , and  $Bin[5, 5, 0.33] = 0.004$ , which are all lower than the statistical significance level (0.05). In this case,  $A$  alone does not predict  $S$  and  $B$  alone does not predict  $S$ . However, both features  $A$  and  $B$  together will be a reliable predictor of  $S$ .

### 3 Predicted Probability

The set of independent predictive features identified by the evaluation process that was discussed earlier would subsequently be used in the prediction calculation. For the purpose of this discussion, we will consider the example in Fig(2) where the simple features  $n_1$  and  $n_2$  were both reliable predictors of  $S$ , and the integrated event  $N$  that connects these two predictors ( $A$  and  $B$ ) together, was not a reliable predictor. The aim of prediction calculation is to calculate the overall probability of  $S$  occurring next, after accounting for all recognised reliable predictors for this event. According to standard frequentist probability theory, if there is a set of independent features predicting an event, the overall predicted probability is calculated by OR-ing all the independent predictors using the following OR expression:

$$Pr(S_{i,t}) = 1 - \prod(1 - p_i) \quad (2)$$

where  $p_i$  is the probability of event  $S_i$  occurring, given by each independent predictor, at a given time ( $t$ ).

In this example, as previously explained, the two predictors  $A$  and  $B$  are independent from each other because the integrated event  $N$  that connects them together, does not reliably predict  $S$ . Since they are independent, we can combine the probabilities of the independent predictors, ( $P(S|A)$  and  $P(S|B)$ ) to calculate the overall prediction of  $S$ , using Eq.2 as follows:

$$Pr(S) = 1 - [(1 - (\frac{6}{7})) * (1 - (\frac{4}{5}))] = 0.97$$

This indicates that the overall probability of  $S$  occurring after the reliable predictor (simple feature  $n_1$  and  $n_2$  independently) at a given time is high. However, at some point in the future, event  $N$  may become reliable. In this case, the individual predictors ( $A$  followed by  $S$  and  $B$  followed by  $S$ ), are no longer independent predictors of  $S$ , and therefore, they will not be used in the calculation for the prediction of  $S$ . The integrated feature  $N$  will be used in the OR expression instead of  $n_1$  and  $n_2$ , to calculate the overall prediction of  $S$ .

### 4 Model Implementation

We utilise frequentist probability theory to build a unified prediction model that forms features; the prediction mechanism in this model uses these features to

make predictions. In our model, we divide the memory into short-term (*STM*) and long-term memory (*LTM*). The *LTM* holds a record of all the perceived events and the relationships between them. These events will be stored in the form of nodes. Each node contains an antecedent (earlier sub-event being perceived), a consequent (the later sub-event being perceived), the time interval between them ( $t$ ), the total number of times the antecedent was followed by the consequent, with the exact time interval  $t$  (represented as counter ( $k$ )), and the total number of times the antecedent was perceived (represented as counter ( $x$ )). On the other hand, the *STM* holds the most recent events only. Considering the limited capacity of *STM* in humans, we simulate this computational limitation by designing a limited *STM* in the models memory. The *STM* is also responsible for updating the counters ( $x$  and  $k$ ), for all related nodes, when new events are perceived.

In order to calculate the prediction of any given event  $S$ , the model selects all nodes that have  $S$  as a consequent and the antecedent of the node occurring in the *STM*. The binomial function will be applied on all the selected nodes, using Eq.2, to evaluate the relationship between all the antecedents and  $S$ ; this will enable the model to identify all the independent predictors that will reliably predict  $S$ . At the end, all the independent predictors will be combined to calculate the overall predicted probability of  $S$  using Eq.2.

## 5 Evaluating the Models Prediction Mechanism

We developed a simulation tool we refer to as the Random Generator (*RG*) to test this model. *RG* generates test data simulating real world environment by producing a sequence of events with various randomly generated temporal relationships linking the occurrence of events together. This test data generated by the *RG* will be used to evaluate our model's performance. In the initialisation phase of the *RG* tool, it assigns random probabilities to different temporal relationships between a large number of events and combinations of events. The next event is produced in a way that is dependent probabilistically on the sequence of prior events. The *RG* tool embodies a probabilistic context - sensitive grammar with long-range dependencies [14], [7]. Our model will perceive the sequence of events generated by the *RG* tool and learn the relationships between these events.

In this test, our model will be challenged to identify the various patterns within the sequence of events generated by the *RG* tool. The learning mechanism in our model aims to learn and recognise reliable relationships, to form features and estimate the predicted probability of future occurrences. If the learning mechanism functions effectively, it should recognise and successfully learn the hidden relationships between these events.

We will test the model by evaluating all the event predictions. Those predictions will be compared to the actual events that occurred succeeding the antecedents. For example, if the model predicts that event  $S_i$  will occur with a probability of 0.2,  $Pr(S_i, t) = 0.2$ , in the test, all the times the model makes

a prediction of an event occurring with this probability, will be counted. If the prediction matches the actual event i.e.  $S_i$  does actually occur with a probability of 0.2, then the model's prediction is considered to be accurate.

## 5.1 Material

The testing process involves two phases, the learning phase and the prediction phase. In the first phase, we run the model  $j$  number of times (where  $j = 10.000$ ), to observe and learn the relationships between the sequence of events generated by the RG tool. Events with reliable relationships between them are formed into a feature. In the second phase, we will run the model for  $N$  number of times (where  $N = 50.000$ ), to calculate and obtain the prediction  $Pr(S_{i,t})$  at each time-step.

To examine the efficiency of the prediction values, we calculate the proportion of the predicted events observed in the test data. First, we examine all previously perceived events and evaluate the prediction value given by the model for each of those events by counting the number of times that the predicted event actually happens. Subsequently, we compare the rate of occurrence of this real event (i.e. the proportion of its observed occurrence) with the prediction value  $Pr(S_{i,t})$  calculated by the model. The accuracy of the model's prediction mechanism can be assessed by comparing these two values.

## 5.2 Method

After observing the sequence of events generated by the RG tool, the model's prediction mechanism calculates the predicted probability for each event occurring at every time-step. This is represented by an overall predicted probability value  $Pr(S_{i,t})$  for each event  $S_i$  occurring at time  $t$ . We examine the value of  $Pr(S_{i,t})$  in different ranges. We separate the  $Pr(S_{i,t})$  values into range  $R$  specified in Table (1).

We then count the total number of occurrences that are predicted to occur with probability value  $Pr(S_{i,t})$  falling between a specific range  $R$ ; ( $R_{min} \leq Pr(S_{i,t}) \leq R_{max}$ ) i.e.  $\sum_j^N |Pr(S_{i,t}) \in R|$ . We calculate this value for all the  $R$  (see Table 1). For each range  $R$ , we identify all these occurrences and count the total number of times the predicted event actually does occur i.e.  $\sum_{t=j}^N |Pr(S_{i,t}) \in R \cap S_{i,t}|$ . Subsequently, we calculate proportion of those cases in which the predicted event actually occurs i.e.  $\mathcal{O}(S_{i,t}, R)$ . To calculate this, we divide the total number of times that the event does occur in each range  $\sum_{t=j}^N |Pr(S_{i,t}) \in R \cap S_{i,t}|$  by the total number of times the event is predicted to occur in the range  $\sum_j^N |Pr(S_{i,t}) \in R|$  as shown by the following formula:

$$\mathcal{O}(S_i, R) = \frac{|\sum_j^N |Pr(S_{i,t}) \in R \cap S_{i,t}|}{|\sum_j^N Pr(S_{i,t}) \in R|} \quad (3)$$

Where  $Pr(S_{i,t})$  is the predicted probability of event  $i$  at time  $t$ ;  $j$  is the starting time (measured in seconds) and  $N$  is ending time (measured in seconds).

For example, for range  $R = 0.20 - 0.25$  the total number of times the event was predicted to occur with this probability ( $(0.20 \leq Pr(S_{i,t}) \leq 0.25)$ ) was 3016 events i.e.  $\sum_j^N |Pr(S_{i,t}) \in R| = 3016$ . Out of these 3016 cases, only 643 actually occurred i.e.  $\sum_{t=j}^N |Pr(S_{i,t}) \in R \cap S_{i,t}| = 643$ , therefore the proportion of occurrences for the predicted events  $\mathcal{O}(S_i, R)$  was 0.21.

We assess the accuracy of the model's prediction by comparing this proportion of occurrences for the predicted events  $\mathcal{O}(S_i, R)$  to the range it is supposed to fall within ( $R$ ). If the model predicts that an event will occur with a probability of 0.15, and the proportion (given by  $\mathcal{O}(S_i, R)$  for the predicted events) is close to 0.15 (e.g. 0.18), this indicates that the model is predicting the next event accurately. The proportion should match or be close to the range R-value range.

Since the accuracy assessment for the model's prediction requires comparison between  $\mathcal{O}(S_{i,t}, R)$  and the range  $R$ , we used the absolute difference function as a measure of the comparison between the two. We measure the absolute difference between the mid point of the range  $M(R)$  and the proportion of observations  $\mathcal{O}(S_{i,t}, R)$  i.e.  $|M(R) - \mathcal{O}(S_{i,t}, R)|$ .

$R$	$\sum_j^N Pr(S_{i,t}) \in R$	$\sum_j^N  Pr(S_{i,t}) \in R \cap S_{i,t} $	$\mathcal{O}(S_{i,t}, R)$	$ M(R) - \mathcal{O}(S_{i,t}, R) $
0.05 - 0.10	1393	193	0.14	0.063
0.10 - 0.15	2268	401	0.18	0.051
0.15 - 0.20	2600	516	0.19	0.023
0.20 - 0.25	3016	643	0.21	0.012
0.25 - 0.30	3901	1094	0.28	0.005
0.30 - 0.35	3341	1005	0.3	0.025
0.35 - 0.40	1806	597	0.33	0.045
0.40 - 0.45	788	272	0.35	0.080
0.45 - 0.50	307	117	0.38	0.095
0.50 - 0.55	171	64	0.37	0.151
0.55 - 0.60	60	32	0.53	0.045
0.60 - 0.65	25	13	0.52	0.105

Table 1: Evaluating the Model's Prediction: This table illustrates the level of agreement between the model's predicted probability values  $Pr$  for each range  $R$  and the proportion of observations of the actual events  $\mathcal{O}(S_{i,t}, R)$ . The second column shows  $\sum_j^N Pr(S_{i,t}) \in R$ : the total number of occurrences that were predicted to occur in  $R$ . The third column shows  $\sum_j^N |Pr(S_{i,t}) \in R \cap S_{i,t}|$  the total number of times that the event did occur in  $R$ . The fourth column shows the overall proportion of observations that were predicted  $\mathcal{O}(S_{i,t}, R)$ . Finally, the last column is the accuracy measure for the model's prediction as the absolute difference value  $|M(R) - \mathcal{O}(S_{i,t}, R)|$ .

### 5.3 Results

We found that the level of agreement between the proportion of observations for predicted events  $\mathcal{O}(S_{i,t}, R)$  and the corresponding range  $R$  was high; the overall correlation between  $\mathcal{O}(S_{i,t}, R)$  and mid-point of the range  $M(R)$  was very high ( $r = 0.97, \rho_{(\mathcal{O}, M(R))} < 0.0001$ ). In general, the  $\mathcal{O}(S_{i,t}, R)$  fell within the corresponding range  $R$  or was close to  $R$ . For example, for  $Pr(S_{i,t})$  values between  $0.15 - 0.20$  i.e. ( $0.15 \leq Pr(S_{i,t}) \leq 0.20$ ), the value of the corresponding proportion  $\mathcal{O}$  i.e.  $\mathcal{O}(S_{i,t}, 0.15 - 0.20) = 0.19$  fell within the same  $R$  range as the  $Pr(S_{i,t})$ . In some cases, the value of  $\mathcal{O}$  fell just outside the corresponding range. For example, in the case where  $Pr(S_{i,t})$  value was between  $0.05 - 0.10$  (i.e. ( $0.05 \leq Pr(S_{i,t}) \leq 0.10$ )), the value of the corresponding  $\mathcal{O}$  i.e.  $\mathcal{O}(S_{i,t}, 0.05 - 0.10) = 0.14$  fell slightly outside of the  $R$  range of the  $Pr(S_{i,t})$ . However, even though this value does not exactly fall within the range, it is close to the  $R_{max}$  ( $R_{max} = 0.10$ ). Overall, our results suggest that in general, the prediction mechanism in our model is efficient i.e. the model is able to predict effectively.

We observe that there is a relationship between the total number of occurrences in the range ( $\sum_j^N Pr(S_{i,t}) \in R$ ) and the proportion of observations  $\mathcal{O}$  ( $\mathcal{O}(S_{i,t}, R)$ ); the bigger the number of cases (i.e. the bigger the value of  $\sum_j^N Pr(S_{i,t}) \in R$ ) or the sample size, the closer the proportion of observations values  $\mathcal{O}$  was to the corresponding range. The measure that we have for accuracy assessment of the model's prediction was  $|M(R) - \mathcal{O}(S_{i,t}, R)|$ . With this, we can estimate how the accuracy of the model's prediction changes as the sample size varies. For example, for range ( $0.50 \leq R \leq 0.55$ ), the total number of occurrences ( $\sum_j^N Pr(S_{i,t}) \in R$ ) was only 171. In this case, the absolute difference ( $|M(R) - \mathcal{O}(S_{i,t}, R)|$ ) was 0.151, indicating that the accuracy of the prediction is low. However, for range ( $0.25 \leq R \leq 0.30$ ), the total number of occurrences ( $\sum_j^N Pr(S_{i,t}) \in R$ ) was 3901. In this case, the absolute difference ( $\sum_j^N Pr(S_{i,t}) \in R$ ) was 0.005, indicating that the accuracy of the prediction is high. We can deduce from this that the difference between  $\mathcal{O}$  and  $R$  is smaller when the model was provided with a bigger sample size. This suggests that the model produces more accurate predictions when more information is available about the likelihood of occurrences, which is not surprising considering that this information can be used in the prediction calculation  $Pr(S_i, t)$ . We note that the low prediction accuracy observed in some ranges coincided with the range  $R$  with lower number of occurrences ( $\sum_j^N Pr(S_{i,t}) \in R$ ) i.e. the model has only perceived those events a small number of times. This indicates that the model is accurately predicting the probability in general and that the anomalous results are expected due to the small sample size.

In conclusion, our test demonstrated that the model has learned the relationships between events efficiently and is capable of calculating relatively accurate predictions based on these learned relationships.

## 6 Conclusion

In this paper, we propose an alternative approach to cognitive modeling. We design the learning mechanism so that it is able to identify reliable relationships between events, form features, and then categorise them in a way that simulates human learning as much as possible. We also incorporate in our model's prediction mechanism the frequentist probability theory, which has been shown to be applicable in human prediction processes [2], [3].

There are various alternative models in the literature that are successful in categorising and making predictions over time. Artificial Neural Networks (ANNs), using deep learning algorithms, are currently the most commonly used model in the field of prediction. While the model we present here is an account of human learning and is not intended to compete with such machine learning algorithms, a brief comparison is worthwhile. The main difference between our model and ANNs is that ANNs are associative models [15], [10], [27] whereas ours is representational. Unlike ANNs, our model includes explicit features that can be extracted and easily used. Moreover, as Gallistel and Gibbon highlighted [5], [4], [6], associative models create associations between two events that are frequently perceived together with a short time interval between them but these associations do not necessarily represent reliable relationships between the events, as these relationships are not evaluated. In contrast, representational models represent the learned data and store it in the "if-then rule" format (if an event  $A$  occurs, then after a specific time interval  $t$ , event  $B$  will occur) [16]. In representational models such as ours, the model can distinguish between causal and coincidental relationships through statistical calculation. In addition, representational models can learn temporal relationships even with a large time interval between events, to verify if there is a causal relationship between them.

In summary, we argue that the prediction model presented in this paper is a more realistic simulation of human learning than ANNs. The simulation is more realistic as its evaluation process is designed using frequentist probability theory that enables the model to form specific meaningful features that are easily accessible and can be used for prediction, unlike ANNs which lack this process.

In future work, we plan to conduct further tests to improve the learning model and incorporate additional features. The experiments and tests would focus on the model's ability to process more complex data to ensure that the prediction results are as close to human predictions as possible. We also plan to integrate the ability to predict actions as well as perceptions into our prediction mechanism with the overall aim of enabling our model to make decisions and perform goal driven actions.

## References

1. Balsam, P.D., Drew, M.R., Gallistel, C.: Time and associative learning. *Comparative Cognition & Behavior Reviews* 5, 1 (2010)
2. Costello, F., Watts, P.: Surprisingly rational: probability theory plus noise explains biases in judgment. *Psychological review* 121(3), 463 (2014)

3. Costello, F., Watts, P.: Peoples conditional probability judgments follow probability theory (plus noise). *Cognitive Psychology* 89, 106–133 (2016)
4. Gallistel, C.R., Gibbon, J.: Time, rate, and conditioning. *Psychological review* 107(2), 289 (2000)
5. Gallistel, C.R.: *The organization of learning*. The MIT Press (1990)
6. Gallistel, C., Gibbon, J.: Computational versus associative models of simple conditioning. *Current Directions in Psychological Science* 10(4), 146–150 (2001)
7. Geman, S., Johnson, M.: Probabilistic grammars and their applications. *International Encyclopedia of the Social & Behavioral Sciences* 2002, 12075–12082 (2002)
8. Griffiths, T., Yuille, A.: A primer on probabilistic inference. *The probabilistic mind: Prospects for Bayesian cognitive science* pp. 33–57 (2008)
9. Hampton, J.A.: Similarity-based categorization: the development of prototype theory. *Psychologica Belgica* 35(2-5), 104–125 (1995b)
10. Haykin, S.S., Haykin, S.S., Haykin, S.S., Haykin, S.S.: *Neural networks and learning machines*, vol. 3. Pearson Upper Saddle River, NJ, USA: (2009)
11. Hodges Jr, J.L., Lehmann, E.L.: *Basic concepts of probability and statistics*. SIAM (2005)
12. Holdershaw, J., Gendall, P.: Understanding and predicting human behaviour. In: ANZCA08 Conference, Power and Place. Wellington (2008)
13. Kruschke, J.K.: Alcov: An exemplar-based connectionist model of category learning. *Psychological Review* 99(1), 22–44 (1992)
14. Lafferty, J., Sleator, D., Temperley, D.: Grammatical trigrams: A probabilistic model of link grammar, vol. 56. School of Computer Science, Carnegie Mellon University (1992)
15. Michel, A.N., Farrell, J.A.: Associative memories via artificial neural networks. *IEEE Control Systems Magazine* 10(3), 6–17 (1990)
16. Nan, J.: A learning model that combines categorisation and conditioning (2013), [www.summon.com](http://www.summon.com), university Collage Dublin, School of Computer Science and Informatics
17. Nosofsky, R.M.: Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General* 115(1), 39 (1986)
18. Osberg, T.M., Shrauger, J.S.: Self-prediction: Exploring the parameters of accuracy. *Journal of Personality and Social Psychology* 51(5), 1044 (1986)
19. Pavlov, I.P.: *Lectures on conditioned reflexes*. vol. ii. conditioned reflexes and psychiatry. (1941)
20. Pavlov, I.P., Anrep, G.V.: *Conditioned reflexes*. Courier Corporation (2003)
21. Pavlov, I.P., Gantt, W.A.H.: *Lectures on conditioned reflexes*. Liveright, New York (1928)
22. Rosch, E.: Principles of categorization. *Concepts: core readings* pp. 189–206 (1999)
23. Savastano, H.I., Miller, R.R.: Time as content in pavlovian conditioning. *Behavioural Processes* 44(2), 147–162 (1998)
24. Skinner, B.F.: *The behavior of organisms: An experimental analysis*. (1938)
25. Skinner, B.F.: *The behavior of organisms: An experimental analysis*. BF Skinner Foundation (1990)
26. Zettlemoyer, L.S., Collins, M.: Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. arXiv preprint arXiv:1207.1420 (2012)
27. Zurada, J.M.: *Introduction to artificial neural systems*, vol. 8. West St. Paul (1992)