EAI Endorsed Transactions
on Context-aware Systems and Applications
Research Article ICST.ORG

Chronological states of viewer's intentions using hidden Markov models and features of eye movement

Minoru Nakayama*, Naoya Takahashi

Human System Science, Tokyo Institute of Technology
Ookayama, Meguro, Tokyo, Japan

Abstract

To determine the possibility of predicting viewer's internal states using the hidden Markov model, several features of eye movements were introduced to the model. Performance was measured using the data from a set of eye movement features recorded during recall tests which consisted of observations of three levels of task difficulty. The features were the temporal appearances of fixations and saccades, and combinations of 8 viewed directions during long and short eye movements. As a result, features of long eye movements, such as saccade information, contributed to prediction accuracy. Also, this prediction accuracy was regulated by the difficulty of the task.

Received on 19 May 2014; accepted on 27 June 2014; published on 05 September 2014

Keywords: User intention, hidden Markov model, features of eye movements

doi:10.4108/csa.1.1.e5

1. Introduction

As eye movements reflect user's mental activity [1], the prediction of user intentions can be studied using eye movements [2]. Here, the definition of intention is the user's decision. In this approach, key features of eye movements are extracted, and discriminant analysis is employed to estimate intention. Recently, when these predictions were made using a combined set of features of eye movements and machine learning techniques, performance improved significantly [3, 4].

Since eye movements are measured as sequential data, time-domain analysis such as conventional scan-path analysis is conducted frequently. The sequential data of eye movements can also be employed to predict user intention [5]. As the scan-paths depend on the features of visual images, mathematical analytical techniques have been introduced to extract significant information [6]. In particular, the hidden Markov model (HMM) [7] is sometimes employed to detect viewer's intentions and to create a cognitive process model using eye movement [8]. For example, in the analysis of the intentions of drivers of cars, a number of implicit cognitive states were created for HMM using the sequential data of driver's eye movements [9]. Also, sentence relevance has been evaluated using eye movement data from an information retrieval task [10]. The HMM is a famous tool for speech recognition, handwriting recognition, etc. The attributes of handwritten data are similar to sequential data of eye movements. Some feature extraction techniques such as "sub-stroke" description [11] may be useful for eye movement analysis. The selection of combinations of features of eye movements is also important to improve prediction performance, while simple combinations of saccades and fixations can be used in an HMM model to predict viewer's intentions [12].

In regards to the above mentioned issues, this paper will examine the feasibility of presenting the viewer's internal states to predict his or her intentions using HMM and selected features of eye movements. Also, the possibility of detecting the process used for internal information processing using the temporal changes recorded in the experimental data is explored. In particular, the performance improvement should be

*Corresponding author. Email: nakayama@cradle.titech.ac.jp

confirmed by comparing it with previous results [12] when the selected features were introduced. Again, the possibility of predicting the viewer's temporal changes in internal states using HMM and sequential data should be determined, in addition to estimating the intention using the overall data [3, 13].

2. Method

To determine the feasibility of detecting chronological changes of intention, previous experimental data [3, 13] was analyzed and applied to the hidden Markov model.

2.1. Experimental task

The task was a recall test for contextual understanding and memorization. First, participants were asked to read a number of definition statements which described locational relationships between two objects (Figure 1a) [3, 13]. A set of definition statements containing a number of statements (K= 3, 5, or 7) was presented for 5.0 seconds for each statement, as shown in Figure 1a, and one of the 10 question statements in Figure 1b was presented for 10.0 seconds following that. The subject was asked to create a knowledge base from the definition statements, and then asked to evaluate whether a question statement was "True" or "False" using inference. Therefore, all statements used simple objects, and were carefully created to possess the same levels of comprehension. The question was a two alternative forced choice (2AFC) task where each question statement was "Yes (True)" or "No (False)" (Figure 1b).

Five sets of statements were created for each of the three levels of tasks. In total, there were 150 data responses divided into 15 task sets. The subjects were 6 male university students ranging from 23 to 33 years of age. They had normal visual acuity for this experiment.

The accuracy rates and mean reaction times of responses to question statements are summarized in Figure 2 [3]. Deviation in the reaction time suggests that the decision is sometimes not stable when thorough understanding does not occur. Therefore, the certainness of decision varies with the level of viewer's understanding, and this may influence their responses. When the response time was long, the certainty was low because participants could not decide quickly. Therefore, all answers were processed as incorrect responses when the reaction time was longer than 4.0 seconds. Since the mean reaction time for incorrect responses was less than 4.0 seconds, it may mean that the longer responses are not certain.

2.2. Eye-movement measuring

The task was displayed on a 20 inch LCD monitor positioned 60 cm from the subject. During the

Figure 1. The screen on the left (a) shows a sample of a definition statement: "A theater is located on the east side of the police station." The screen on the right (b) shows a sample of a question statement: "There is a post office on the south side of the theater."

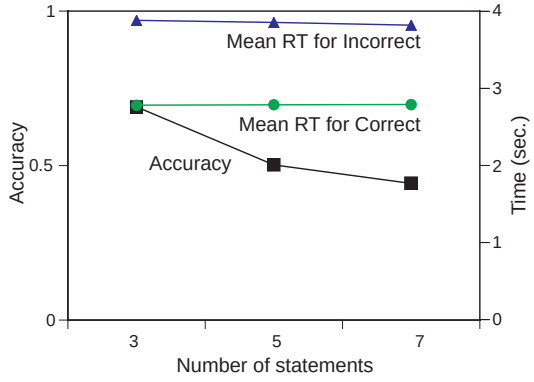

Figure 2. Accuracy rate and mean reaction times for responses

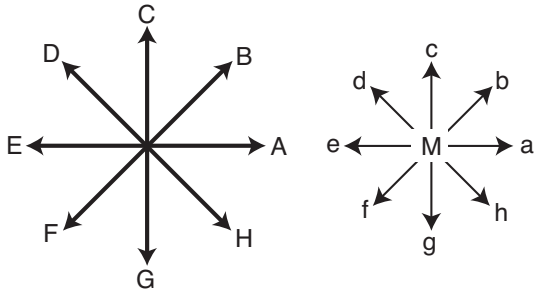

Figure 3. Eye movement categories: 'A'–'H' ('a'–'h') are directions of saccades (micro-saccades) and 'M' is a fixation.

experiment, participant's eye movements regarding definition and question statements were continually observed, using a video-based eye tracker (EMR-8NL). Eye-movement was tracked on a 640 by 480 pixel screen at 60 Hz.

The tracking data was converted into visual angles, and eye movements were divided into saccades and fixations using a threshold of 40 degrees per second [14].

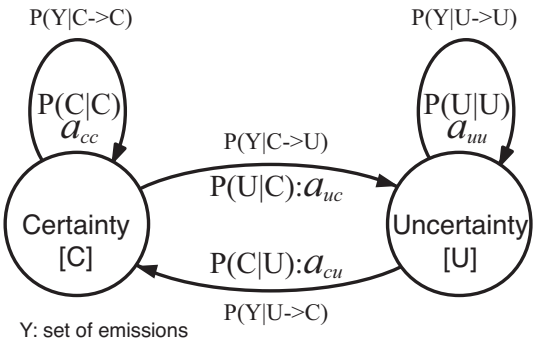

Figure 4. HMM diagram.

3. Hidden Markov modeling

According to the correct and incorrect responses in the task results, it is supposed that there are two levels of internal states used when making choices regarding the certainness mentioned above. These two internal states are defined as "Certainty" and "Uncertainty" for the prediction of response correctness. The probability of a correct response may be high when an internal state remains at "Certainty". The state of being certain starts from a state of "Uncertainty", and then moves between "Uncertainty" and "Certainty". During these state transitions, some emissions are observed as sets of eye movement features.

Using the series of observed eye movements as a set of time series data, the observation O and a set of eye movement features Y can be defined as follows:

$$\begin{aligned} O &= \{o_1, o_2, o_3, \cdots, o_t\} \\ o_t &\in Y \\ t &= T \times 60 - 1, T : sampling\ time(sec.) \\ & \quad 60 : sampling frequency(Hz) \end{aligned} \tag{1}$$

A study of HMM-based handwriting recognition introduced features such as "sub-stroke" writing motion [11]. This trace analysis of handwriting was employed to extract features of eye movements [8]. A similar method of eye movement analysis has been conducted using EOG [15]. As mentioned above, the eye movements were classified into saccades and fixation, but were all scalar. Using the categories in Figure 3, the saccades and micro-saccades during fixation were categorized into one of 8 directions, as were "sub-stroke" classifications from the handwriting study. The directional categorization of eye movements was based on an idea from a previous study [3, 13]. The set of emissions Y can be noted using a combination of features, as in the following formula:

$$\begin{aligned} Y &\ni \{Y_{f-s}, Y_{f9-s8}, Y_{f9-s}, Y_{f-s9}\} \\ Y_{f-s} &= \{fix, sac\} \\ Y_{f9-s8} &= \{a, \ldots, h, M, A, \ldots, H, \} \\ Y_{f9-s} &= \{a, \ldots, h, M, sac\} \\ Y_{f-s9} &= \{fix, A, \ldots, H\} \end{aligned} \tag{2}$$

Figure 4 illustrates a hidden Markov model of two states of certainty of responses and a set of emissions during transitions. The two states indicate levels of high or low certainty for the two responses, known as "Certainty" and "Uncertainty".

As a result, the model of HMM λ can be defined as follows:

$$\begin{aligned} \lambda &= \{S, Y, A, B, \pi\} \\ S &= \{s_i | "C" : certain \vee "U" : uncertain\} \\ A &= \{a_{ij}, i, j = C, U | a_{cc}, a_{cu}, a_{uu}, a_{uc}\}, \sum_j a_{ij} = 1 \\ B &= \{b_{ij}(k), i, j = C, U, k \in Y\}, \sum_k b_{ij}(k) = 1 \\ \pi &= \{\pi_C = 0, \pi_U = 1\} \end{aligned} \tag{3}$$

A set of parameters, $\theta \equiv (A, B)$ is optimized using experimental data. The Baum-Welch algorithm, which uses a likelihood function, is employed as shown in (4) [7].

$$\log P(S|O, \theta) \tag{4}$$

The Forward algorithm provides a series of state transitions S, which maximize the likelihood function [7].

Additionally, the probability (Pr) of remaining in one state can be defined as c for "certain" and u for "uncertain", while $Pr(c) + Pr(u) = 1$. According to the features of the Markov transition, the transitional probability can be calculated as shown in (5). Then, the probability of certainty can be stated as a time series.

$$\lim_{m \to \infty} \begin{bmatrix} a_{cc} & a_{uc} \\ a_{cu} & a_{uu} \end{bmatrix}^m \times \begin{bmatrix} c^{ini.} \\ u^{ini.} \end{bmatrix} \to \begin{bmatrix} c \\ u \end{bmatrix} \tag{5}$$

4. Results

The parameters of HMM models were optimized using data sets as sequential data between the observation time and stimulus onset, and performance evaluations were conducted every 0.05 seconds after stimulus onset. The results were then summarized using the same format as in the previous study [12].

4.1. Prediction accuracy

The performance of predicting the states and responses was evaluated using the leave-one-out technique. As previously indicated, correct and incorrect responses were predicted, and the estimation accuracy was also evaluated during the time series. An example of prediction results for 3 statements using a feature set of directional saccades and micro-saccades (Y_{f9-s8}) is summarized in Table 1.

The temporal change in prediction accuracy using the 3 statement condition is summarized in Figure 5. Performance in three conditions (Y_{f-s}, Y_{f9-s8}, Y_{f-s8}) is compared in the figure. Most levels of accuracy for the models with Y_{f9-s8} and Y_{f-s8} are higher than for the ones with Y_{f-s}, as has been previously reported

Table 1. Prediction accuracy of 3 statements using set (Y_{f9-s8}) at 450 msec. (Accuracy=70.7%)

	Prediction	
	Certainty	Uncertainty
Correct	188	19
Incorrect	69	24

$\chi^2 = 13.1, df = 1, p < 0.01$

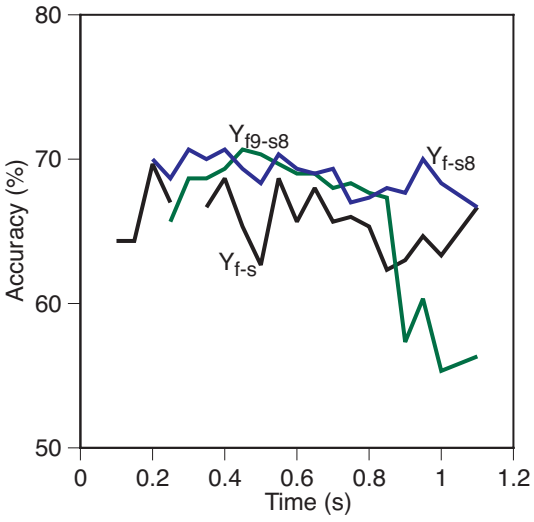

Figure 5. Prediction accuracy across the viewing time (K=3).

[12]. In particular, the levels of accuracy from 0.3 to 0.7 seconds are the highest. On the other hand, the level of accuracy for the models with Y_{f9-s8} suddenly decreases after 0.9 seconds, and the level of accuracy for the models with Y_{f9-s} remains at chance. These all suggest that the appropriate combination of features of eye movements provide accurate predictions of viewer's certainty during a chronological change of state, while the level of accuracy changes with the given data set of temporal eye movements. Since performance is high when detailed directional saccade information is employed, prediction performance is a significant source of information for the prediction of intention. This result coincides with previous analysis of the estimation of response correctness [13]. Additionally, the prediction can be made using shorter time series data, though this prediction performance is lower than the performance of previous analysis. This is a benefit of the proposed procedure.

However, performance remains at chance when the number of definition statements is 5 or 7 (K=5 and 7). Prediction performance using the original responses was significant if even the reaction time was longer than 4.0 seconds [12]. It may be possible that either the overall correct rates influence prediction accuracy, or there are other calculation-related issues. At least, chronological prediction is possible when the number of definition statements is 3 (K=3), since the mental workload in this condition is not high [13]. A detailed analysis of this will be a subject of our further study.

4.2. Probability of staying in "Certainty"

As mentioned in Equation (5), the probability can be calculated using the transition matrix, which is estimated using a set of features of eye movements in a time series. The probability of remaining in the "Certainty" state indicates the degree to which the content is understood, so that the viewer can make correct responses. The temporal changes in probability for the sets with Y_{f-s}, and Y_{f-s8} are calculated using the three levels of the number of definition statements (K=3,5,7), and summarized in Figure 6. Again, the number of definition statements determines the task difficulty. Figure 6 shows the probability relationship between the temporal time and the task difficulty. For example, a viewer has a strong inclination to make correct responses when question statements are presented, so the probability of remaining in the "Certainty" state exists. However the inclination sometimes decreases as the statement is understood, so the probability also decreases. The probabilities with Y_{f-s8} are higher than the ones with Y_{f-s} while the probabilities decrease with the number of definition statements. These results seem reasonable in view of the experimental conditions.

Most levels of probability are high during the early stage of reading, in the area of 0.1-0.3 seconds, and then decrease through the initial probability sets to 0 at stimulus onset. After 0.8 seconds, the levels of probability for the conditions (K=5, 7) are the lowest. As the results of temporal accuracy changes suggest, the features of eye movements during early stage of reading a statement affect the level of certainty of the response.

The level of probability of a set with Y_{f9-s8} is summarized in Figure 7 using the same procedure as with other sets. Using this method of calculation, the results present a different tendency. The levels of probability for a condition (K=3) remain at the same level, which is higher than 0.8 during the initial second following stimulus onset. The level of probability was high during a specific time when the number of definition statements increased (K=5, 7). As Figure 7 shows, the level of probability for condition (K=5) suddenly dropped after 0.6 seconds. Also, another condition (K=7) presents a high level of probability between 0.4 and 0.65 seconds. According to the results of transition simulations using this data set, viewers remain in the "Certainty" state for all conditions between 0.4 and 0.6 seconds. This result may show that there is a common period of certainness to be inclined to respond correctly when the knowledge

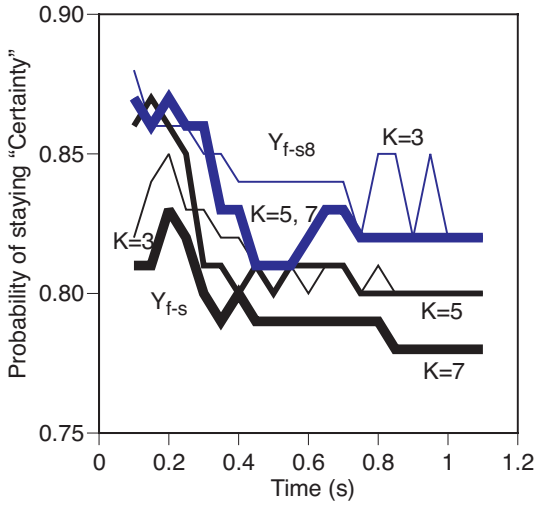

Figure 6. Probability changes of remaining in "Certainty" with Y_{f-s} and Y_{f-s8}.

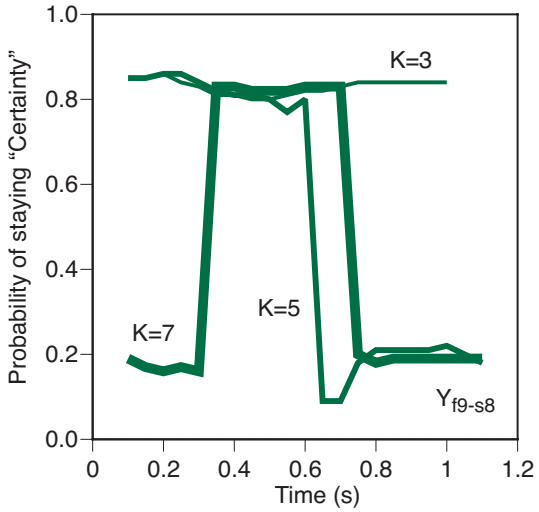

Figure 7. Probability changes of staying in "Certainty" with Y_{f9-s8}.

base is complicated (K=5,7) though it is almost always possible when the knowledge base is the simplest (K=3). Therefore, viewers may receive key information useful in a recall test during this short period of time.

This phenomenon is likely to be similar to previous studies regarding text comprehension and information processing, and should be discussed in more detail. However, it is not clear which exact type of processing is occurring during this period. A more detailed analysis should be conducted and will be a subject for our further study. Also, a model setting which includes model selection and tuning should be considered for use in future studies. Determination of the possibility that this procedure can be applied to other tasks will be a subject of our further study.

5. Conclusion

To obtain chronological transitions of internal states in order to determine viewer's intentions, the hidden Markov model (HMM) was applied to sequential data sets of eye movements. Regarding the features of handwriting motions for HMM, temporal eye movements such as saccades and fixation were categorized using eye movement directions, and then the parameters of HMMs were calculated using the data which was observed during a recall test. Levels of prediction accuracy regarding internal state transitions were measured using HMMs and sequential features of eye movements. Performance depended on details of features presented and sequential data in the early stage of viewing, and performance using those features improved rather than ones with simple combination of saccades and fixations. The probability of remaining in an internal state can be calculated using the sequential data, and the specific time period can be extracted as a stage of information processing using HMM models.

References

[1] Jacob, R.J.K. and Karn, K.S. (2003) Eye tracking in human–computer interaction and usability research: Ready to deliver the promises. In Hyona, Radach and Deubel [eds.] *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (Oxford, UK: Elsevier Science BV).

[2] Goldberg, J.H. and Schryver, J.C. (1995) Eye-gaze determination of user intent at the computer interface. In Findlay, J., Walker, R. and Kentridge, R. [eds.] *Eye Movement Research, Volume 6: Mechanisms, Processes and Applications (Studies in Visual Information Processing)* (Elsevier), 491–502.

[3] Nakayama, M. and Hayashi, Y. (2010) Estimation of viewer's response for contextual understanding of tasks using features of eye-movements. In Hyrskykari, A. and Ji, Q. [eds.] *Proceedings of ACM Symposium on Eye-Tracking Research & Applications (ETRA2010)* (New York, USA: ACM): 53–56.

[4] Bednarik, R., Vrzakova, H. and Hradis, M. (2012) What do you want to do next: A novel approach for intent prediction in gaze-based interaction. In Mulligan, J.B. and Qvarfordt, P. [eds.] *Proceedings of ETRA 2012: ACM Symposium on Eye-Tracking Research & Applications* (New York, USA: ACM): 83–90.

[5] Goldberg, J.H. and Kotval, X.P. (1999) Computer interface evaluation using eye movements: methods and constructs. *Industrial Ergonomics* **24**: 631–645.

[6] Choi, Y.S., Mosely, A.D. and Stark, L.W. (1995) String editing analysis of human visual search. *Optometry and Vision Science* **72**(7): 439–451.

[7] Bishop, C.M. (2006) *Pattern Recognition and Machine Learning* (New York, USA: Springer Science+Business Media).

[8] Salvucci, D.D. and Anderson, J.R. (1998) *Tracing Eye Movement Protocols with Cognitive Process Models*. Tech. Rep. Paper 48, Carnegie Mellon University, Department Psychology.

[9] Liu, A. (1998) What the driver's eye tells the car's brain. In Underwood, G. [ed.] *Eye Guidance in Reading and Scene Perception* (Elsevier), 431–452.

[10] Puolamäki, Y., Salojärvi, J., Savia, E., Simola, J. and Kaski, S. (2005) Combining eye movements and collaborative filtering for proactive information retrieval. In Heikkil, A., Pietik, A. and Silven, O. [eds.] *Proceedings of ACM-SIGIR 2005*, ACM (New York, USA: ACM Press): 145–153.

[11] Nakai, M., Akira, N., Shimada, H. and Sagayama, S. (2001) Substroke approach to HMM-based on-line Kanji handwriting recognition. In *Proceedings of the Sixth International Conference on Document Analysis and Recognition (ICDAR 2001)*: 491–495.

[12] Takahashi, N. and Nakayama, M. (2013) Chronological prediction of certainty in recall tests using Markov models of eye movements. In *Proceedings of BIOTECHNO 2013: The Fifth International Conference on Biometrics, Biocomputational Systems and Biotechnologies* (IARIA): 55–60.

[13] Nakayama, M. and Hayashi, Y. (in Press) Prediction of recall accuracy in contextual understanding tasks using features of oculo-motors. *Universal Access in Information Society* .

[14] Ebisawa, Y. and Sugiura, M. (1998) Influences of target and fixation point conditions on characteristics of visually guided voluntary saccade. *The Journal of the Institute of Image Information and Television Engineers* **52**(11): 1730–1737.

[15] Bulling, A., Ward, J.A., Gellersen, H. and Tröster, G. (2011) Eye movement analysis for activity recognition using electrooculography. *IEEE Transasction on pattern analysis and machine intelligence* **33**: 741–753.