Analysis of Gaze Trajectories in Natural Reading with Hidden Markov Models

Maksim Volkovich

Lomonosov Moscow State University

Abstract. The process of natural reading, registered with a modern eye-tracking system generates a signal of complicated structure that can be considered as a time series consisting of gaze point coordinates. Signal properties are supposed to depend on various properties of presented text as well as on current cognitive condition of a reader such as attention focus, level of fatigue, level of text understanding and other parameters. The task of cognitive state recognition can be approached with the modeling of gaze trajectories using probabilistic models, which parameters may contain information relevant to read text properties and reader's cognitive state. In this work a new approach of gaze trajectories modeling based on Hidden Markov Models is proposed. HMM's transition probability matrix corresponds to probabilities of saccades between words and emission probability functions correspond to words coordinates and overall measurement noise. Two variants of HMM are proposed: text-related HMM models multiple gaze trajectories collected on the same text from different readers, subject-related parametric HMM models gaze trajectories produced by a single reader on a set of consecutive pages from the same text. A series of experiments on simulated data were performed to estimate a required sample size and a required level of measurement accuracy for a forthcoming data collection procedure.

Keywords: Eye-tracking · Natural reading · Hidden Markov Models

1 Introduction

Natural reading is a complex task that includes eye movement processes, lexico-semantic processing dependent on reader attention and visual features of the text being read.

The process of reading consists of long relatively rare movements (”saccades”) between areas of high attention where eyes are fixated on a word for some time depending on the skill of reader.

Eye movements during reading are under the direct control of linguistic processing [1]. There are three properties of a word that influence its ease of processing: word's frequency, length and predictability in context, the so-called Big Three [2]. These words properties affect time of word's processing, length of saccade and probability of word's skipping. Thereby longer words require more processing time from reader because they consist of larger number of letters then shorter, so they include more information to process. Although the word's frequency is correlated with it's length in general case, it has been shown that more frequent words are processed faster then infrequent words with similar lengths [3]. Moreover, word's length affects not only the duration of current fixation but the length of next saccade and duration of next fixation [4, 5]. More predictable words are more likely to be skipped and require shorter fixations.

Medical conditions also influence eye-movements during reading. For example, schizophrenia patients read slower, make shorter saccades, have longer forward fixation durations and make a greater number of regressive saccades [6]. Several studies show that children with dyslexia read slower, make a larger number of long progressive saccades, make twice more fixations on long words and less often skip short words then children without this condition [7, 8].

Eye-tracking signals are successfully used in a large number of different tasks. Subject's attention during reading can be determined with respect to type of reading: reading, skimming and scanning based on features extracted from eye-tracking data: amplitude, angularity, velocity of saccades and duration of fixations [9]. Eye-movements data may be used in tasks of language proficiency determining [11]. Students' eye movements taken while IELTS reading test completion have been analyzed and significant difference between samples taken from successful and unsuccessful attempts to pass the exam has been revealed [12]. In a part-of-speech tagging task data extracted from eye-tracker allows to improve the performance of an approach based on Second-order Hidden Markov Models. The extracted data was transformed into 22 features encoding information about fixation durations, probabilities, number of fixations, refixations and regressions related to current word and its neighbours [13]. The same features are found to be useful in a domain of Named Entity Recognition in approach based on a usage of a bidirectional LSTM. Using embeddings based on these features it became possible to improve performance of the previous state-of-the art model [14].

An approach based on applying Slip-Kalman filtering is used to track the progression of reading. This approach works particularly well for determining the event of changing a read line but also shows good results with respect to noise reduction [15].

At the works discussed above different variations of Hidden Markov Models are used in order to determine the part of speech, the correct coordinate at which the eye is directed etc. In this work we propose to model eye-tracking trajectories using Hidden Markov Models: associate a set of hidden states with individual read words and fit the matrix of transition probabilities between them. We assume that this matrix can be used as a set of features for a models solving various cognitive state recognition tasks.

2 Proposed Approach

2.1 Hidden Markov Models

A first-order Hidden Markov Model is a probabilistic model which is based on Markov Chain model. An HMM is defined by the following components:

- $X = \{x_1, \dots, x_N\}, x_n \in \mathbb{R}^d$ – a sequence of observed values
- $T = \{t_1, \dots, t_N\}, t_n \in \{0,1\}^K, \sum\limits_{j=1}^{K} t_{nj} = 1$ – a sequence of hidden states that correspond to observed values.
  $t_i = \begin{cases} 1, & \text{if a model in a state } i \\ 0, & \text{otherwise} \end{cases}$
- A transition probability matrix $A^{K \times K}$, where $a_{ij} = p(t_{n,j}|t_{n-1,i})$ is a probability of transition from the hidden state $t_{n-1,i}$ to the state t_{nj}, $\sum\limits_{j=1}^{k} a_{ij} = 1 \; \forall i$
- $p\left(x_n|t_n\right)$ – observation likelihoods or emission probabilities, expressing probabilities that observed value x_n would be generated from a hidden state t_n. It is assumed that conditional distribution $p\left(x_n|t_n\right)$ is known up to parameters $\phi_k, k \in \{1, \dots, K\}$, so if $t_{ni} = 1$ then x_n is from $p(x_n|\phi_i)$
- $\pi = \{\pi_1, \dots, \pi_K\}$ – an initial probability distribution. π_i is the probability of HMM being started in state i. $\sum\limits_{j=1}^{N} \pi_i = 1$

A first-order HMM instantiates 2 assumptions. The first one is Markov assumption: the value of the hidden state t_n depends only on the state at the previous moment t_{n-1}.

$$p(t_n|t_1, \dots, t_{n-1}) = p(t_n|t_{n-1}) \tag{1}$$

Second, the value of the observed value x_n depends only on the current hidden state. It is known as Output Independence assumption.

$$p(x_n|X, T) = p(x_n|t_n) \tag{2}$$

Having a sequences of observed values (in our case, coordinates taken from an eye-tracker) we need to determine from which hidden states (read words) these coordinates were led from. It is needed to solve HMM learning problem: learn the transition probability matrix A and observation likelihoods $p\left(x_n|t_n\right)$ from given observation sequences X and set of possible hidden states where each hidden state would represent read word. Let's denote the set of HMM parameters as $\Theta = \{\pi, A, \phi\}$. For an HMM model parameters estimation Expectation-Maximization or EM algorithm can be applied.

$$\Theta_* = \arg\max_{\Theta} p(X|\Theta) = \arg\max_{\Theta} \sum_T p(X, T|\Theta)$$

$$p(X|\Theta) = \sum_T p(X, T|\Theta) \to \max_{\Theta} \Leftrightarrow \log\left(\sum_T p(X, T|\Theta)\right) \to \max_{\Theta}$$

1. Initialization step. At the beginning it is needed to set $\Theta = \{\pi, A, \phi\}$ parameters.
   - π and A are usually set randomly but corresponding to restrictions $\sum\limits_{j=1}^{K} \pi_i = 1$ and $\sum\limits_{j=1}^{k} a_{ij} = 1 \; \forall i$

  - ϕ initialization depends on $p(x|\phi)$ distributions
2. Expectation step. Θ_{old} are fixed.

$$\mathbb{E}_{T|X,\Theta_{old}} \log p(X,T|\Theta) = \sum_{T} \log p(X,T|\Theta) p(T|X,\Theta_{old})$$

3. Maximization step. $p(T|X,\Theta_{old})$ are fixed.

$$\Theta_{new} = \arg\max_{\Theta} \mathbb{E}_{T|X,\Theta_{old}} \log p(X,T|\Theta)$$

4. Expectation and Maximization steps are repeated until convergence

Baum-Welch algorithm is the special case of the standard HMM-training algorithm which allows to perform E and M steps more efficiently due to optimization related to HMM assumptions 1 and 2.

Our primary goal is to propose an approach for extracting subject-dependent features from eye-tracking data that would correspond to cognitive states and a group of text-dependent features. Examples of cognitive states that can be extracted are a level of fatigue, stress, focus, a level of text understanding, emotional state. One of the features of the text can be ”difficult” words that take a longer reading time, require higher number of fixations and longer fixations that can be explained by the requirements for the level of proficiency in the language or subject area. Or we can try to discern the influence of words which are unpredictable in context on eyes movements patterns. Eye-tracking data can be represented as a sequence of coordinates taken from eye-tracker with resolution of several hundred coordinates per second. Each coordinate from a sequence can be assigned to an area of high attention such as a read word. It is proposed to learn an HMM using coordinates as observed values and determine the set of hidden states based on knowledge about the number of words. From practical point of view, one of two following situations is considered for further analysis: either a same text is read by a certain number of different subjects, or a single subject reads a set of texts of sufficient size. Thus, two different models are proposed for these scenarios, text-dependent and subject-dependent.

2.2 Task 1: Text Analysis

In a ”same text, different readers” scenario, text-dependent features remain the same for every session, but reader-dependent features should be different for different readers.

The main objective in the analysis of this scenario is to determine a vector of text-dependent parameters given a set of observation sequences related to one text fragment read by a certain number of subjects. It is assumed that a set of observations related to a large enough set of subjects can be used to estimate a subject-independent HMM parameters which can be analysed for the purpose of extracting text-related features while subject-dependent features would be suppressed. Thus the text-dependent HMM consists of the following components:

- $X_j = \{x_{1j}, \ldots, x_{nj}\}$ – is a set of observed values taken from subject j.
- $T = \{t_1, \ldots, t_K\}$ – is a set of hidden states determine by a number of words in the text.
- $p(x_n|t_n)$ – emission probabilities – set of two-dimensional Gaussian distributions
  $N = \{N_1(\mu_1, \sigma_1), \ldots, N_n(\mu_n, \sigma_n)\}$ with means defined by geometric centers of words $C = \{c_1, \ldots, c_K\}$, $\mu_i = c_i$ and standard deviations σ represents overall noise of measurements of the eye-tracking device and is estimated from the data.
- $\pi = \{\pi_1, \ldots, \pi_K\}$ – initial probability distribution. It also can be estimated from samples.
- $A^{K \times K}$ – a transition probability matrix that can be fitted from set of eye-tracking measurements taken from different subjects.

Parameters of the model trained from samples taken from different subjects reading the same text may potentially represent such text characteristics as sentence structure, word frequency, general text complexity, etc.

2.3 Task 2: Subject Analysis

Let's consider a scenario when a single subjects reads a certain number of text fragments. It may be assumed that during one session cognitive states of a subject should not change significantly over time and therefore model parameters should be similar for every single page. The objective of the analysis in this scenario is to estimate these subject-dependent parameters. Since the data is presented as a set of eye-tracker trajectories collected from different text fragment, each single trajectory is assumed to be sampled from a corresponding text-dependent HMM. It is assumed that parameters of each HMM can be presented as a function of higher-level parameters that refer to current cognitive state of a reader. For example, a level of fatigue can be modeled as the average duration of a fixation on a word. In order to train a subject-dependent HMM on these data, a parametric family of HMM is proposed.

Suppose that θ is a vector of parameters that represents a cognitive state of a subject. Such parameters as average number of fixations, average fixation duration, average number of regressions could be represented as functions dependent on the vector θ and words frequencies. In a simplest case, a parametric family of HMM can be proposed in a following form:

- $M = \{m_1, \ldots, m_P\}$ is a set of HMMs, m_i is a HMM corresponding to text i.
- For each HMM emission probabilities are defined by two-dimensional Gaussian distributions with parameters considering geometry of text and measurements noise.
- For each HMM a set of hidden states is determined by number of words $W = \{w_1, \ldots, w_{K_p}\}$ presented in a text fragment.

- Reading process is modeled using probabilities $\alpha(\theta, W)$, $\beta(\theta, W)$, $\delta(\theta, W)$, $\epsilon(\theta, W)$, where $\alpha(\theta, W)$ is a probability of continuing a current fixation, $\beta(\theta, W)$ is a probability of saccade to a next word, $\delta(\theta, W)$ is a probability of saccade to a previous word, $\epsilon(\theta, W)$ is a probability of long forward or backward saccade.

Thus a transition probability matrices have the following form:

$$A^{K \times K} = \begin{pmatrix} \alpha(\theta, w_1) & \beta(\theta, w_1) & \epsilon(\theta, w_1) & \dots & \dots & \epsilon(\theta, w_1) \\ \delta(\theta, w_2) & \alpha(\theta, w_2) & \beta(\theta, w_2) & \dots & \dots & \epsilon(\theta, w_2) \\ \epsilon(\theta, w_3) & \delta(\theta, w_3) & \alpha(\theta, w_3) & \dots & \dots & \epsilon(\theta, w_3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \epsilon(\theta, w_{K-1}) & \dots & \dots & \dots & \alpha(\theta, w_{K-1}) & \beta(\theta, w_{K-1}) \\ \epsilon(\theta, w_K) & \dots & \dots & \dots & \delta(\theta, w_K) & \alpha(\theta, w_K) \end{pmatrix}$$

Thus, a parametric form of subject-dependent HMM can be defined by setting a vector of parameters θ and choosing a set of functions $\alpha, \beta, \delta, \epsilon$. The model can be more detailed if other text-related parameters would be taken into account.

3 Eye-tracking Corpora

For future studies and experiments two existing eye-tracking datasets were chosen: Zurich Cognitive Language Processing Corpus [16] and Ghent Eye-Tracking Corpus [17].

ZuCo is a publicly available dataset containing eye-tracking and EEG data recorded from 12 native English speakers reading natural English text. Subjects were presented with three tasks: normal reading task in which participants had to give an assessment of movie described in a read text fragment, normal reading task with multiple choice questions about read content and task-specific reading task in which subjects had to focus on a certain semantic relation type. The corpus includes high density eye-tracking data recorded with a calibrated infrared eye-tracker. Fixations, saccades and blinks are identified by the tracker software. The dataset also includes such statistics of gaze trajectories as time after first fixation on sentence for every single fixation, number of fixations for each word and sentence and mean pupil size, gaze duration (GD) during first word reading, sum of reading time of word (total reading time or TRT), first fixation duration (FFD), the duration of first and single fixation on a word (single fixation duration or SFD) and go-past time (GPD) which is the sum of all fixations preceding saccade to the right.

GeCo is a corpus of eye-tracking data taken from 14 monolingual and 19 bilingual participants reading a novel. Bilinguals were classified as English speakers with proficiency level from lower-intermediate to advanced. Bilingual participants read half of the novel in their first language and the other half in their second language. The size of the text read by each subject was about 5,000 sentences. As in ZuCo dataset in addition to raw data extracted by the tracker there are presented word-level reading measurements: GD, FFD, SFD, TRT and GPT.

4 Experiments

A set of experiments was executed to prove a concept that eye-tracking trajectories can be generated using HMM transition probability matrix and then HMM can be fitted the model parameters can be fitted close to the parameters of the original model. For a given set of sentences we can obtain coordinates of exact positions of words on a page. The distance was measured in relation to the size of a printed letter. A Hidden Markov Model was initialized in the following way according to our vision of what they should look like. The number of hidden states was set according to the number of displayed words. A transition probability matrix was generated as diagonally dominant since the number of saccades and therefore hidden states transitions has to be much smaller then number of eye movements inside areas of gaze fixations on a single word when the hidden state does not change. Initial probability distribution was chosen to be geometric distribution in order to simulate a task in which the subject must read the text from the beginning. A set of gaussian 2-d distributions with means located in word centers was chosen as a set of emission probabilities.

A simulated training dataset of sequences of observed values was generated using the defined HMM. For a texts with several dozen words each sequence consists of 300 observed values. We perform an experiment to find out how many observation is needed for a precise fitting of HMM parameters. The process of model training was run for datasets with different number of observations. For each size of training sample a new model was trained. Each model was trained using its own dataset. Then mean squared error between original transition probability matrix and trained transition probability matrix and mean squared error between main diagonals of original and trained matrices for each trained model were measured.

$$\text{MSE} = \frac{1}{K * K} \sum_{i=1}^{K} \sum_{j=1}^{K} (A_{ij} - A^*_{ij})^2 \tag{3}$$

$$\text{MSE diagonal} = \frac{1}{K} \sum_{i=1}^{K} (A_{ii} - A^*_{ii})^2 \tag{4}$$

MSE mean values and their 95% confidence intervals were also calculated. Confidence interval were defined as $\overline{x} \pm Q(1 - \frac{\alpha}{2}) * s$ where $\overline{x}$ is the sample mean, $s = \sqrt{\frac{(\sum\limits_{i=1}^{N} x_i - \overline{x})^2}{N-1}}$ is the sample standard deviation, α is the confidence level, N

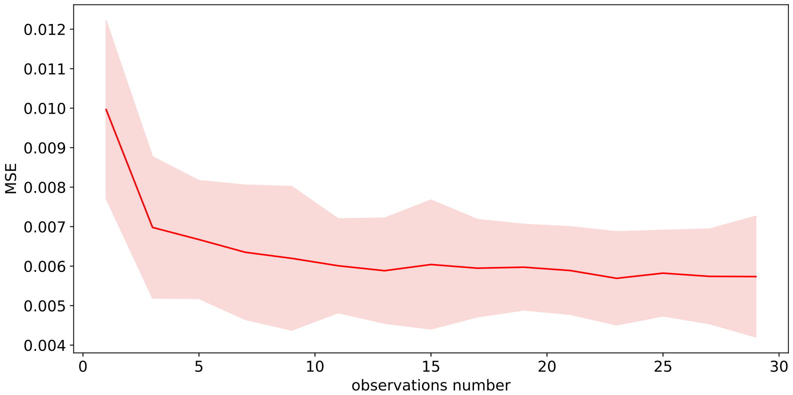

Fig. 1. Mean Squared Error over the difference of original and a trained transition probability matrices. The red line indicates mean MSE. The light red area includes the 95% MSE confidence interval.

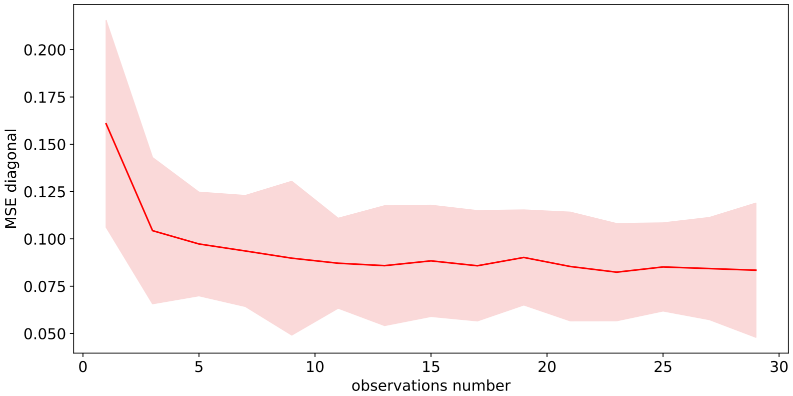

Fig. 2. Mean Squared Error over the diagonals difference of original and a trained transition probability matrices. The red line indicates mean MSE. The light red area includes the 95% MSE confidence interval.

is the sample size, Q is quantile finction: $Q(p) = \inf\{x \in \mathbb{R} : p \leq F(x)\}$, $F(x)$ is the cumulative distribution function for the Student's t-distribution with $N - 1$ degrees of freedom. A confidence interval gives a more descriptive estimate of a parameter than a point estimation. Results of our experiments are presented on Fig. 1 and Fig. 2.

The experiments were executed using python package named hmmlearn [18]. As is shown in the figures our metrics reach values close to optimal at observations number between 10 and 15. With a further increase in observations numbers MSE over full transition probability matrices and MSE over diagonals do not decrease significantly. So for a suboptimal quality it could be enough to collect data from dozen of subjects.

5 Conclusions and Future Work

In this work various approaches to the analysis of eye movement in different text-dependent and subject-dependent scenarios have been considered. A new approach of gaze trajectories modeling based on Hidden Markov Models is proposed for both scenarios. An experiment conducted as part of a preliminary study helped us determine the approximate size of the sample we would need for a further work. It is planned to apply a proposed approach on real data taken from ZuCo and GeCo datasets in one of the following tasks.

1. Binary classification task: was the answer given by subject in a sentiment task from ZuCo dataset was correct or not.
2. If information about the answer time is available it may be possible to evaluate this parameter using a model, because it is supposed that time required for answer is dependent on how thoroughly the text was read.
3. The fatigue classification task. It is assumed that at the end of the long reading session a fatigue level is higher then at the beginning of the session. Data from GeCo corpus would be useful.

It is also planned to collect a new dataset containing samples of eye movement recordings taken while reading text fragments in Russian. Several dozen subjects will read several texts at different levels of fatigue. Fatigue level will be measured in two ways: by interviewing subjects and using a binary classifier trained to recognize the beginning and end of the session. Using the second method, we will assume that fatigue level rises at the end of the session.

References

1. Dambacher, M., Slattery, T. J., Yang, J., Kliegl, R., Rayner, K. Evidence for direct control of eye movements during reading. Journal of Experimental Psychology: Human Perception and Performance, 39(5):1468 (2013)
2. Clifton Jr, C., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., Schotter, E. R. Eye movements in reading and information processing: Keith Rayner's 40 year legacy. Journal of Memory and Language, 86:1-19 (2016)
3. Rayner, K., Duffy, S. A. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & cognition 14.3:191-201 (1986)
4. White, S. J., Rayner, K., Liversedge, S. P. The influence of parafoveal word length and contextual constraint on fixation durations and word skipping in reading. Psychonomic bulletin & review 12.3:466-471 (2005)

5. Juhasz, B. J., et al. Eye movements and the use of parafoveal word length information in reading. Journal of Experimental Psychology: Human Perception and Performance 34.6:1560 (2008)
6. Whitford, V., et. al. Reading impairments in schizophrenia relate to individual differences in phonological processing and oculomotor control: Evidence from a gaze-contingent moving window paradigm. Journal of Experimental Psychology: General 142.1:57 (2013)
7. Olson, R. K., Kliegl, R., Davidson, B. J. Dyslexic and normal readers' eye movements. Journal of Experimental Psychology: Human Perception and Performance 9.5:816 (1983)
8. De Luca, M., et. al. Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia. Neuropsychologia 37.12:1407-1420 (1999)
9. Mozaffari, S. S. et al. Reading Type Classification based on Generative Models and Bidirectional Long Short-Term Memory. Joint Proceedings of the ACM IUI 2018 Workshops. CEUR Workshop Proceedings 2068 (2018)
10. Kunze, K., et. al. I know what you are reading: recognition of document types using mobile eye tracking. Proceedings of the 2013 International Symposium on Wearable Computers (2013)
11. Kunze, K. et. al. Towards inferring language expertise using eye tracking. CHI'13 Extended Abstracts on Human Factors in Computing Systems, 217-222 (2013)
12. Bax, S. Readers' cognitive processes during IELTS reading tests: Evidence from eye tracking. British Council, ELT Research Papers 13-06 (2013)
13. Barrett, M., et. al. Weakly supervised part-of-speech tagging using eye-tracking data. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Volume 2: Short Papers (2016)
14. Hollenstein, N., Zhang, C. Entity Recognition at First Sight: Improving NER with Eye Movement Information. arXiv preprint arXiv:1902.10068 (2019)
15. Bottos, S., Balasingam, B. A Novel Slip-Kalman Filter to Track the Progression of Reading Through Eye-Gaze Measurements. arXiv preprint arXiv:1907.07232 (2019)
16. Hollenstein, N., et al. ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading. Scientific data 5.1:1-13 (2018)
17. Cop, U., et al. Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. Behavior research methods 49.2:602-615 (2017)
18. hmmlearn library. https://hmmlearn.readthedocs.io
19. Ulutas, B. H., Özkan, N. F., Michalski, R. Application of hidden Markov models to eye tracking data analysis of visual quality inspection operations. Central European Journal of Operations Research, 28:761–777 (2020)