

Scanpath Complexity

Modeling Reading Effort Using Gaze Information

ABHIJIT MISHRA, DIPTESH KANOJIA, SEEMA NAGAR, KUNTAL DEY, AND

PUSHPAK BHATTACHARYYA

Overview

Measuring reading effort is useful for practical purposes such as designing learning material and personalizing text comprehension environment.

We propose a quantification of reading effort and call the measure Scanpath Complexity.

It is modeled as a function of various properties of gaze fixations and saccades, collected from the eye movement patterns of readers.

Our method personalizes reading effort for a person unlike the readability scores which generalize the effort.

Introduction

The effort spent by the reader on a task often determines the reward associated with the task.

- In education, reading effort controls the motivation and learning experience.
- For text annotation which involves reading, reading effort controls the financial incentives.

From a NLP perspective, quantifying reading effort for text-annotation tasks may give rise to better annotation-cost-models vis-à-vis ones that rely on word and sentence counts.

Methods like MRI, and EEG are limited to laboratories, and are prohibitively expensive.

Our work relies on eye-movement data, and is based on the eye-mind hypothesis i.e. when a subject views a word/object, he or she processes it cognitively for approx. the same time they fixate on it.

Crux: Gaze patterns in the form of Scanpath as input indicate conceptual difficulty the reader experiences.

Related Work

A number of successful models of eye-movement control for reading include the one from Reichle and Laurent (2006), the E-Z Reader (Reichle, Rayner, and Pollatsek 2003; Reichle, Pollatsek, and Rayner 2006), SWIFT (Engbert et al. 2005) and Bayesian inference based models (Bicknell and Levy 2010; Engbert and Krugel 2010).

Rayner and Duffy (1986) show how fixation time is associated with different forms of lexical complexity in the form of word frequency, verb complexity, and lexical ambiguity.

Demberg and Keller (2008) relate complex eyemovement patterns to the syntactic complexity present in the text. von der Malsburg and Vasishth (2011) show that complex saccadic patterns (with higher degree of regression) are related to syntactic re-analysis arising from various forms of syntactically complex structures (*e.g.*, garden-path sentence).

Malsburg, Kliegl, and Vasishth (2015) also propose a method to determine scanpath regularity, and observe that sentences with short words and syntactically more difficult sentences elicited more irregular scanpaths.

Scanpath Complexity

Mathematically, Scanpath complexity denoted as *ScaComp* is,

$$ScaComp = f(X, \theta)$$

where X correspond to a set of N attributes $x_1 ; x_2 ; x_3 ; \dots ; x_n$

and θ corresponds to model parameters.

Heuristic ScaComp (*ScaComp_H*) *i.e.* scanpath complexity to be linearly proportional to each scanpath attribute.

$$ScaComp = \theta \times \prod_{i=1}^N x_i + C$$

Supervised ScaComp (*ScaComp_L*) *i.e.* designed as a weighted sum of constituents.

$$ScaComp = \sum_{i=1}^N w_i x_i + C$$

Baseline Absence?

In absence of prior baselines that address how the model attributes can be combined to get the most effective model possible, we considered two rudimentary functions (linear sum and product) *prima facie*.

We draw inspirations from general science where it is very standard in case of modeling of physical phenomena to take product of all influencing factors or their inverses- as the case may be (e.g., in laws relating pressure, temperature and volume), and in case of statistical phenomena to use linear regression like expressions.

We thought it is quite important to gain a first level insight, and most importantly, creating a baseline for future research.

Scanpath Attributes

Various attributes corresponding to fixations and saccades combine to form scanpath complexity.

We divide these attributes into two categories - Fixational attributes and Saccadic attributes, as shown in the table. We do not normalize the attributes by text length assuming that reading effort is often associated with the length of the text, hence, normalization would rule out its effect.

Attributes	Intent
Basic Fixational Attributes	
Total Fixation Duration (<i>FD</i>)	Sum of all fixation duration
Total First-Fixation Duration (<i>FFD</i>)	Sum of duration of fixations during the first pass reading of words
Total Regression-Fixation Duration (<i>RFD</i>)	Sum of duration of fixation on a regressed word
Total Fixation Count (<i>FC</i>)	Count of all fixations
Skipped Word Percentage (<i>SKIP</i>)	Fraction of words which have no fixation on them (or skipped)
Basic Saccadic Attributes	
Total regression count (<i>RC</i>)	Count of regressions
Total saccade distance (<i>SD</i>)	Sum of saccadic distance in terms of character count.
Total regression distance (<i>RD</i>)	Sum of regression distance in terms of character count
Complex Saccadic Attributes (Introduced by us)	
Negative Saccade log-likelihood (<i>NLL</i>)	Negative of the log-likelihood of saccade transitions with respect to an ideal saccade transition model (refer to section 4.1)

Negative Saccade log-likelihood (NLL)

We first propose a saccade transition model that is based on an ideal reading behavior.

Malsburg, Kliegl, and Vasishth (2015) find that in the Potsdam Sentence Corpus (Kliegl et al. 2004), 50% of the saccades target the next word in a sentence; in 19% of the saccades, the next word is skipped; 17% of the saccades result in refixations of the current word; and 8% are regressive saccades landing on the word directly receding the current word.

We propose a bi-modal ideal saccade transition distribution which comprises two asymmetric Gaussian distribution (denoted by \mathcal{N}_{assym}); one for progressions and the other for regressions.

$$P(s) = \psi * \mathcal{N}_{assym}(\mu_p, \sigma_{p1}, \sigma_{p2}) \\ + (1 - \psi) * \mathcal{N}_{assym}(\mu_r, \sigma_{r1}, \sigma_{r2})$$

where ψ is the probability of performing a progressive saccade, $1 - \psi$ is the probability of performing a regressive saccade. μ_p , σ_{p1} and σ_{p2} are mean and standard deviation associated with the left part and the right part of the asymmetric Gaussian distribution for the progressive saccades. μ_r , σ_{r1} and σ_{r2} are mean and standard deviation associated with the left part and the right part of the asymmetric Gaussian distribution for the regressive saccades.

NLL Distribution

The distribution N_{assym} with parameters μ , σ_1 and σ_2 can be described as,

$$\mathcal{N}_{assym}(\mu, \sigma_1, \sigma_2) = \frac{1}{Z} \exp\left(-\frac{(s - \mu)^2}{2\sigma^2}\right)$$

and,

$$\sigma = \begin{cases} \sigma_1, & s < \mu \\ \sigma_2, & s \geq \mu \end{cases}$$

And the normalization constant Z is given by,

$$Z = \sqrt[2]{\frac{\pi}{2}} (\sigma_1 + \sigma_2)$$

Experiment Setup

We compute scanpath complexity in two ways, by following equations 2 and 3. Our technique requires scanpath data to be available. To combine scanpath attributes using supervised statistical techniques (equation 3), we need data annotated with scores representing reading/annotation effort.

We collected 32 paragraphs of 50 - 200 words on 16 different topics belonging to the domains of history, geography, science and literature. For each topic, two comparable paragraphs were extracted from **Wikipedia** and **simple Wikipedia**.

The documents are annotated by 16 participants. 13 of them are graduate/post-graduate students with science and engineering background in the age group of 20 - 30 years, with English as the primary language of academic instruction.

- The other 3 are expert linguists and they belong to the age group of 47 - 50.

The task given to the participants is to read one document at a time and assign the paragraph with a “reading difficulty” score of 1 to 10.

- Higher scores indicate higher degree of difficulty.

Choice of NLL model parameters

For experimental purposes the parafoveal range is often considered to be 7 characters to the left and 12 characters to the right of the current fixation (Bicknell and Levy 2010).

We fix the value of μ_r and μ_p to be -8 and 13 respectively. The shape parameters σ_{p1} , σ_{p2} , σ_{r1} and σ_{r2} are empirically set to 22; 18; 3; 13 respectively by trial and error, plotting the distribution.

Probability of regression ($1 - \psi$) is kept as 0.08 considering that around 8% of the total saccade transitions are regressions.

Computing Scanpath Complexity

Annotation scores (which are to be taken as measures of scanpath complexity) obtained from participants are highly subjective and vary from person to person.

We normalize these scores across all the documents for each individual by scaling them down to a range of [0,1]. Scanpath attributes are also normalized for computational suitability.

Total reading time has been considered as a measure of effort.

In eye-tracking setup, *total annotation time* often amounts to total fixation duration or total gaze duration.

We perform a series of univariate linear regression tests where the cross correlation between each attribute and the dependent variable are measured and are converted to ANOVA F-scores and p-values.

Evaluation

Reading difficulties can broadly be related to two factors

- Linguistic complexity, textual attributes, readability of the given text *etc.*
- Individual factors (age, domain knowledge and language skills).

We evaluate scanpath complexity using the various measures presented in table on the next slide, pertaining to linguistic complexity, textual attributes and readability.

We evaluate our techniques using Spearman's rank correlation coefficients between scanpath complexity and the linguistic complexity, basic textual and readability measures.

Complexity Measures

Attributes	Intent
	Basic Properties
Word Count (<i>W</i>) Sentence Count (<i>S</i>) Characters per Word (<i>C/W</i>) Syllables per Word (<i>S/W</i>) Words per Sentence (<i>W/S</i>)	
	Readability Scores
Flesch-Kincaid (<i>FK</i>) (Kincaid et al. 1975) Gunning-Fog (<i>GF</i>) (Gunning 1969) SMOG (<i>SMOG</i>) (Mc Laughlin 1969) LEXILE (<i>LEX</i>) (Stenner et al. 1988)	
	Lexical Complexity
Total Degree of Polysemy (<i>DP</i>) Lexical Sophistication (<i>LS</i>) Lexical Density (<i>LD</i>) Out-of-vocabulary Words (<i>OOV</i>)	Sum of number of Wordnet senses of all content words. Lexical Sophistication Index proposed by Lu (2012) Ratio of content words to total number of words Ratio of words not present in GSL (jbauman.com/gsl.html) and AWL (victoria.ac.nz/lals/resources/academicwordlist) to total words.
	Syntactic Complexity
Dependency Distance (<i>DD</i>) Non-terminal to Terminal ratio (<i>NN</i>) Clause per Sentence (<i>CLS</i>) Complex Nominal per Clause (<i>CN/C</i>)	Average distance of all pairs of dependent words in sentence (Lin 1996) Ratio of the number of non-terminals to the number of terminals in the constituency parse of a sentence
	Semantic Properties
Discourse Connectors (<i>DC</i>) Co-reference Distance (<i>CD</i>) Perplexity (<i>PP</i>)	Number of Discourse Connectors Sum of token distance b/w co-referring entities of anaphora in sentence Trigram perplexity using language models trained on a mixture of sentences from the Brown corpus

Insights

This evaluation criterion is chosen to gain insights into whether any variation in such textual properties is related to the way scanpath is formed on the text.

Since scanpath complexity is considered as a personalized measure, we compute the correlation coefficients for each participant to demonstrate the effectiveness of our technique.

Table on the next slide shows the averaged correlation coefficients. For measures pertaining to lexical complexity; the baseline method correlates well with the complexity measures.

- *ScaComp_L*, on the other hand is better correlated with syntactic, semantic complexity measures and readability.
- *ScaComp_H* does not perform better than the baseline for our dataset.

Results

	W	S	C/W	S/W	W/S	FK	GF	SMOG	LEX	DP	LS
Baseline	0.84	0.41	0.56	0.46	0.55	0.60	0.56	0.57	0.58	0.61	0.41
ScaComp_H	0.92	0.50	0.51	0.42	0.51	0.56	0.54	0.56	0.58	0.70	0.35
ScaComp_L	0.94	0.51	0.52	0.44	0.52	0.58	0.58	0.59	0.59	0.72	0.33
p	0.0001	0.0007	0.032	0.32	0.03	0.02	0.04	0.03	0.008	0.0001	0.008

	LD	OOV	DD	NN	CL/S	CN/C	DC	CD	PP
Baseline	0.30	0.08	0.56	-0.05	0.30	0.69	0.46	0.30	-0.02
ScaComp_H	0.23	0.03	0.55	-0.04	0.29	0.65	0.53	0.30	-0.11
ScaComp_L	0.22	-0.01	0.57	-0.08	0.32	0.63	0.53	0.33	-0.17
p	0.0004	0.003	0.008	0.1	0.2669	0.002	0.005	0.13	0.0001

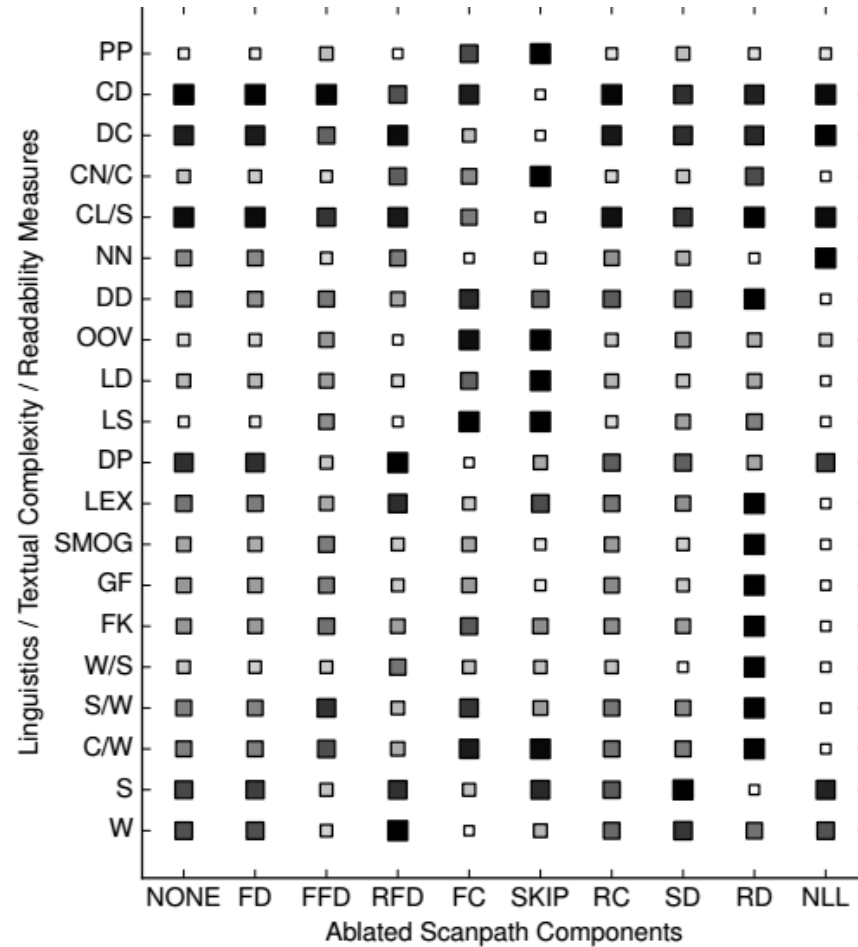
Ablation Test

We perform a series of ablation tests to see how each scanpath component described in Table 1 affect our scanpath complexity measures. Ablation of one scanpath component at a time largely results in a reduction of correlation coefficients observed in both *ScaComp_H* and *ScaComp_L* settings.

It is worth noting that ablation of components like *FD* and *RC*, which are often used in psycholinguistic literature, results in a slight degradation of correlation values, whereas our proposed *NLL* measure proves to be very important, as its ablation results in a significant degradation.

We also tried ablating *FD*, *RC* and *NLL* together and observed a great reduction of correlation values. On the other hand, considering only these three components makes the model as good as the one with all components. Yet, in some cases, the “all-component” combination beats the “*FD - RC - NLL*” combination by a good margin.

Ablation Test Results



Conclusion and Future Work

Our work tries to model readers eye-movement behavior to quantify the cognitive effort associated with reading processes.

We showed that the **measurement of complexity of scanpaths leads to better cognitive models that explain nuances in the reading better than total annotation time**, a popular measure of cognitive effort.

We have **validated Scanpath complexity by obtaining correlation between the measure and various levels of linguistic complexities** associated with the text.

Our work **does not yet address effects individual factors** (*viz.* age, domain expertise and language skills) on scanpath complexity, studying which is on our future agenda.

In future, we would also like to jointly model fixations and saccades for scanpath complexity measurement, instead of treating these attributes separately.

Thank you!

Questions?

We thank members of Center for Indian Language Technology (CFILT) Lab for their immense help and support during the course of this work.

References

Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124(3):372.

Reichle, E. D., and Laurent, P. A. 2006. Using reinforcement learning to understand the emergence of "intelligent" eye-movement behavior during reading. *Psychological review* 113(2):390.

Reichle, E. D.; Pollatsek, A.; and Rayner, K. 2006. E-z reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research* 7(1):4–22.

Reichle, E. D.; Rayner, K.; and Pollatsek, A. 2003. The ez reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences* 26(04):445–476.

Engbert, R.; Nuthmann, A.; Richter, E. M.; and Kliegl, R. 2005. Swift: a dynamical model of saccade generation during reading. *Psychological review* 112(4):777.

References

- Bicknell, K., and Levy, R. 2010. A rational model of eye movement control in reading. In *Proceedings of the 48th annual meeting of the ACL*, 1168–1178. ACL.
- Engbert, R., and Krugel, A. 2010. Readers use bayesian estimation for eye movement control. *Psychological Science* 21(3):366–371.
- Rayner, K., and Duffy, S. A. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 14(3):191–201.
- Demberg, V., and Keller, F. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2):193–210.
- von der Malsburg, T., and Vasishth, S. 2011. What is the Scanpath signature of syntactic reanalysis? *Journal of Memory and Language* 65(2):109–127.
- Malsburg, T.; Kliegl, R.; and Vasishth, S. 2015. Determinants of scanpath regularity in reading. *Cognitive science* 39(7):1675–1703.
- Kliegl, R.; Grabner, E.; Rolfs, M.; and Engbert, R. 2004. Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology* 16(1-2):262–284.