Predicting Term-Relevance from Brain Signals

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Manuel J. A. Eugster[2*], Tuukka Ruotsalo[1*], Michiel M. Spapé[1*], Ilkka Kosunen[3], Oswald Barral[3], Niklas Ravaja[14], Giulio Jacucci[13], Samuel Kaski[23]

Helsinki Institute for Information Technology HIIT & [1] Aalto University | [2] Aalto University, Department of Information and Computer Science | [3] University of Helsinki, Department of Computer Science | [4] University of Helsinki, Department of Social Research | [*] equal contribution

In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '14). ACM, New York, NY, USA, 425-434. DOI 10.1145/2600428.2609594.

I. Motivation

Relevance prediction is a central challenge of Information Retrieval (IR) research as it determines the information presented to the user. Term-Relevance Prediction from Brain Signals (TRPB) is proposed to automatically detect relevance of information directly from brain signals.

Research questions:

1. How well can we predict relevance judgments on terms from the brain signals of unseen users?
2. Which parts of the EEG signals are important for the prediction?

II. Neural-Activity recording experiment

Scenario:

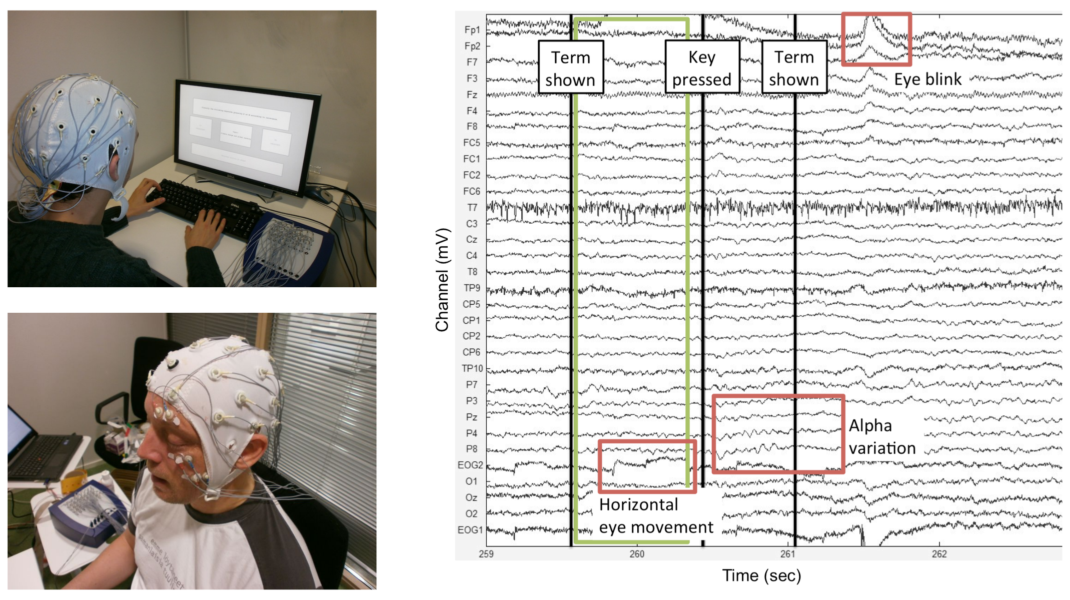
- Each participant read and judged hand-picked terms in six topics
- One term at a time; no repetitions
- balanced ground-truth

Examples:

Entrepreneurship: business risk, startup company, …
Iraq war: US army, Saddam Hussein, …
Irrelevant words: shopping, video-games, …

Data:

38 participants, ca. 1368 relevance judgments



(a) Participants of the experiment with the full EEG sensor setup to record the raw EEG signals reading the instructions for the next task. (b) Excerpt of an participant's captured raw EEG signals with annotations.

III. Feature engineering

EEG feature representations in the frequency domain (frequency-band based) and in the time domain (event-related-potential based).

Views	v_k	Features
Relevance judgement view:		
Relevance		A binary relevance judgement provided by a participant for a term for a given topic
Frequency-band-based views:		
Theta	1	40 features for each frequency band: 20 features of average power over 1 second epochs before the relevance judgement; 20 features of average power over entire period, minus power of the second before term onset
Alpha	2	
Beta	3	
Gamma1	4	
Gamma2	5	
Engage	6	
Event-related-potential-based view:		
ERPs	7	80 features of average amplitude: 20 features for 80–150 ms, P1; 20 features for 150–250 ms, N1/P2; 20 features for 250–450 ms, N2 or P3a; 20 features for 450–800 ms: N4 or P3b

IV. Classification setup

- Bayesian Efficient Multiple Kernel Learning,

$$y(\mathbf{x}_*) = \mathbf{a}^T \left(\sum_{k=1}^{K} e_k \mathbf{v}_{k,*} \right) + b$$

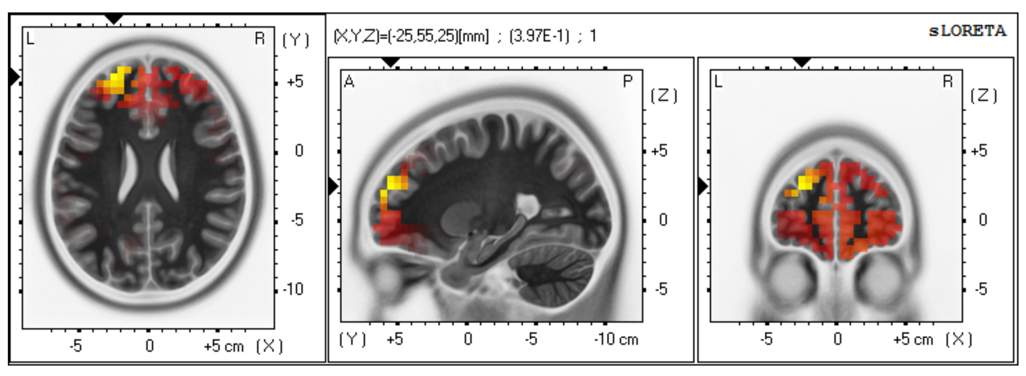
- with $\boldsymbol{y}$ the binary judgments, $\boldsymbol{v}_k$ the views, e_k the kernel weights [1] (RQ2).
- Leave-one-participant-out strategy to estimate the classification accuracy (RQ1).
- Only observations that conformed to the ground truth, balance between relevant and irrelevant observations, five repetitions.
- Simple automatic feature selection procedure based on the *t*-statistic.

V. Predictive power

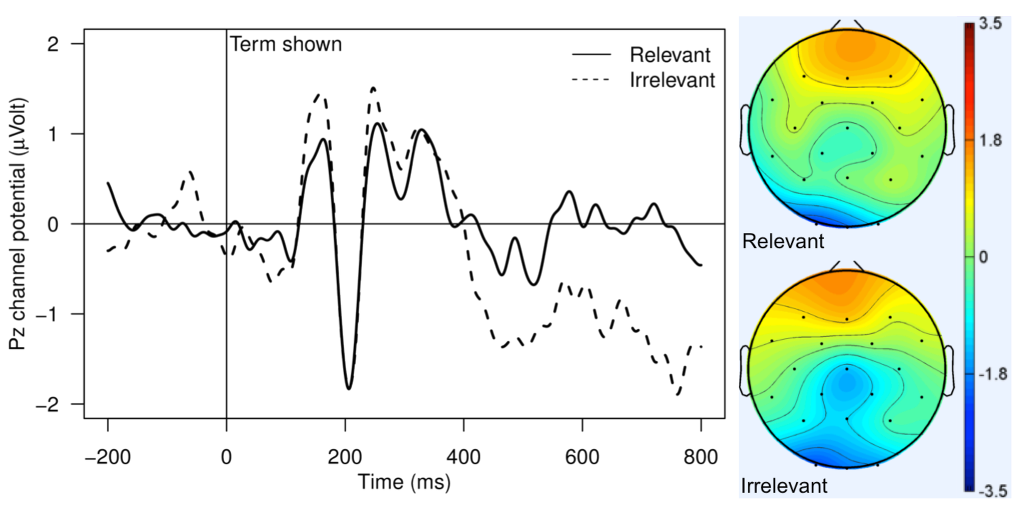
Views	Mean accuracy	*p*-value	Mean improvement
All	0.5415	**0.0003**	8.30%
Selected combined views:			
Al+Ga1	0.5429	0.0014	8.59%
Al+E	0.5475	**0.0007**	9.50%
Ga1+E	0.5528	**0.0002**	10.55%
Al+Ga1+Be	0.5369	0.0022	7.37%
Al+Ga1+E	0.5586	**<0.0001**	11.72%
Individual views:			
Alpha (Al)	0.5242	0.0265	4.83%
Gamma1 (Ga1)	0.5143	0.1445	2.86%
Beta (Be)	0.5005	0.4838	0.10%
Gamma2	0.5101	0.2003	2.02%
Theta	0.5000	0.4984	0.01%
ERPs (E)	0.5312	0.0092	6.24%
Engage	0.4773	0.9673	−4.55%

Bold entries denote that improvements are statistically significant at a level alpha = 0.01, p-value < alpha with correction for multiple testing.

VI. Physiological findings



Localization of Alpha change assoziated with relevance mapped to a normalized brain: Brodmann Area 10 is associated with a range of cognitive functions that are important for relevance judgments, such as recognition, semantic processing, and memory recall [4].



Grand average of the ERP in the Pz channel: ERP after irrelevant and relevant term onset with significance difference after 450ms, maximizing at 747ms. The latency and topography of the potential suggest the involvement of a P3-like potential [3].

References:
[1] Gönen, 2012, ICML, Bayesian Efficient Multiple Kernel Learning.
[2] Ruotsalo, et al., 2013, CIKM, Directing exploratory search with interactive intent modeling.
[3] Polich, 2007, Clinical Neurophysiology, Updating p300: An integrative theory of P3a and P3b.
[4] Moshfeghi, et al., 2013, Advances in Information Retrieval, Understanding Relevance: An fMRI Study.

VII. Application: Topic representation

In certain IR applications the target is to detect true positive terms (i.e., relevant with very high probability) that represent a user's search intent [2]. Topic-wise prediction using a **high-precision classifier** with p > 0.99 as threshold for a term being classified as relevant:

Topic	Count all	Count relevant	Precision	Recall	Top 5 relevant terms
Climate change and global warming	209	111	0.5238	0.0991	Snowmelt, Elevated CO2, Climate change, *hardware synchronization*, *sightseeing*
Entrepreneurship	199	110	0.6897	0.1818	business risk, startup company, business creation, *shopping*, *virtual relationships*
Immigration integration	204	105	0.5238	0.1048	citizenship, ethnic diversity, xenophobia, *arsonist*, *morse code*
Intelligent Vehicles	185	109	0.8000	0.1101	pedestrian tracking, collision sensing, remote driving, radar vision, *arsonist*
Iraq war	208	111	0.6296	0.1532	Saddam Hussein, US army, Tony Blair, *morse code*, *rock n roll*
Precarious employment	204	106	0.5714	0.1132	minimum wage, employment regulation, job instability, *virtual relationships*, *video-games*
Mean	202	109	0.6231	0.1270	

Normal font indicates a relevant term according to the ground truth, italics indicates an irrelevant term.

VIII. Summary

- Relevance judgments happen in the brain and therefore the most intriguing way to predict relevance is to directly use the brain signals.
- We showed that term-relevance prediction using only brain signals captured via EEG is possible.
- We demonstrated its usage to construct meaningful sets of terms for unknown topics.
- For future developments and all our other research related to IR, visit http://augmentedresearch.hiit.fi/.

Helsinki Institute for Information Technology HIIT is a joint research institute of Aalto University and the University of Helsinki for basic and applied research on information technology.

Aalto University   UNIVERSITY OF HELSINKI

This work has been partly supported by the Academy of Finland (Multivire, 255725; and the Finnish Centre of Excellence in Computational Inference Research COIN, 251170), Re:Know funded by TEKES, and MindSee (FP7 -- ICT; Grant Agreement # 611570).