Data in Cognitive Science

Frank Keller

Institute for Language, Cognition and Computation School of Informatics, University of Edinburgh

October 30, 2014

(日) (同) (三) (三)

Introduction

2 Eye-tracking Data

- Visual Attention
- Language Processing

3 Applications

- Human Parsing
- Object Detection

< ∃ >

Cognitive Science

The aim of cognitive science is to figure out how the mind works.



(日) (同) (三) (三)

Cognitive Science

Cognitive scientists do this by studying a range of cognitive processes:

• vision;



A B > A B >

Cognitive scientists do this by studying a range of cognitive processes:

- vision;
- auditory processing;



Image: A math a math

Cognitive scientists do this by studying a range of cognitive processes:

- vision;
- auditory processing;
- language;



▲ 同 ▶ → 三 ▶

Cognitive scientists do this by studying a range of cognitive processes:

- vision;
- auditory processing;
- language;
- motor control.



(日) (同) (三) (

Cognitive scientists do this by studying a range of cognitive processes:

- vision;
- auditory processing;
- language;
- motor control.



▲ @ ▶ < ∃ ▶</p>

Here, we will focus on vision and language.

Visual Attention Language Processing

Eye-tracking

By recording their eye-movements, we can study how humans process linguistic or visual information.



Visual Attention Language Processing

Eye-tracking

By recording their eye-movements, we can study how humans process linguistic or visual information.



Visual Attention Language Processing

Eye-tracking

By recording their eye-movements, we can study how humans process linguistic or visual information.



Visual Attention Language Processing

Visual Attention

When people view an image, their eyes remains static for short periods of time (fixation), and then jump to a new location (saccade).



(日) (同) (三) (

.∋...>

Visual Attention Language Processing

Visual Attention

When people view an image, their eyes remains static for short periods of time (fixation), and then jump to a new location (saccade).



Visual Attention Language Processing

Visual Attention

When people view an image, their eyes remains static for short periods of time (fixation), and then jump to a new location (saccade).



.∋...>

Visual Attention Language Processing

Visual Attention

When people view an image, their eyes remains static for short periods of time (fixation), and then jump to a new location (saccade).



(日) (同) (三) (

Visual Attention Language Processing

Visual Attention

When people view an image, their eyes remains static for short periods of time (fixation), and then jump to a new location (saccade).



More fixations on salient regions, out-of-context objects, animate objects, task-relevant areas.

Buck did not read the newspapers, or he would have known that trouble was brewing, not alone for himself, but for every tide-water dog, strong of muscle and with warm, long hair, from Puget Sound to San Diego. Because men, groping in the Arctic darkness, had found a yellow metal, and because steamship and transportation companies were booming the find, thousands of men were rushing into the Northland. These men wanted dogs, and the dogs they wanted were heavy dogs, with strong muscles by which to toil, and furry coats to protect them from the frost.

Buck lived at a big house in the sun-kissed Santa Clara Valley. Judge Miller's place, it was called. It stood back from the road, half hidden among the trees, through which glimpses could be caught of the wide cool veranda that ran around its four sides.

¹⁶⁸ **Buck** did not read the newspapers, or he would have know that trouble was brewing, rotatione torthimself, but or every tide-water epge 3trong of muscle and with warm on 52 hair, from Puget Sound to ຮັສກ²⁴⁵ເອີດວ. ອີຣັດລມຮູ້ຂຶ້າຫຼາຍ grouping in the Arctic dankness, had found ayellow field, and because steam ship and transportation companies were boond the thousands of men were rushing into the 188 Northiand. These ine wanted bogs, and the bogs the wanted were freavy dogs, with strong muscles by which to toil, and turry coats to protect them from the frost. 164_64 148_ 168 Bock lived at a big house in the sun kissed Santa Clara Valley. ²³udge Miller Splace, it was called ₂₂It stood back from the road, half ²hidden among the trees, through which glimpses could be caught of the wide cool veranda that ran around its four sides.

Britk did hot read the new spapers, or he way to have know that 168148 14644 the dale was browing, not stone for himself, but for every the water toge strong of multiple and with Warm, long hair, from Puget Sound to San Bago, Besause new groping in the Arctic dallmess, had found $\frac{212}{3}$ yell $\frac{256}{5}$ and $\frac{258}{5}$ and $\frac{258}{5}$ and $\frac{2127}{5}$ and $\frac{2129}{5}$ and $\frac{2129}{5}$ and $\frac{2129}{5}$ and $\frac{178}{5}$ and $\frac{178}{$ Not Hand. These mented togs, and the dogs they wanted were 8813223 188₂₅ heavy togs, with strong muscles by which to toil, and turry coats to protect them trops the flost.

Buck fried at a big house in the sun kissed Santa Clata Valley. Audo Millers place, it was called It stood back from the road, half hidden arteng the trees, through which glimps could be caught of the wide opol veranda that ran around its four sides

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

< ⊡ > < ⊡ > <</p>

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

(日) (同) (三) (

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

Human parsing is word-by-word incremental. Example:

A (1) > A (2) > A

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

Human parsing is word-by-word incremental. Example:

(1) Banks

(日) (同) (三) (

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

Human parsing is word-by-word incremental. Example:

(1) Banks will

▲□ ► ▲ □ ► ▲

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

Human parsing is word-by-word incremental. Example:

(1) Banks will take

▲□ ► ▲ □ ► ▲

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

Human parsing is word-by-word incremental. Example:

(1) Banks will take measures.

A (1) > A (1) > A

Human Parsing Object Detection

Human Parsing

Incremental TAG model proposed by Demberg et al. (2013):



1. NP $\rightarrow \langle \{A0,A1,A2,A3\},Banks,nil \rangle$

2. VP \rightarrow (AM-MOD,will,nil)

(日) (同) (三) (三)

Human Parsing Object Detection

Human Parsing

Incremental TAG model proposed by Demberg et al. (2013):



Human Parsing Object Detection

Predicting Reading Times



Human Parsing Object Detection

Object Detection

We can use eye-tracking data for a classic computer vision task: object class detection.



Object detectors are trained on images that are manually annotated with bounding boxed around the objects.

Human Parsing Object Detection

Object Detection

Alternative: infer bounding boxes from eye-tracking data (Papadopoulos et al. 2014):



- much faster: 1 s per image vs. 22 s for BB drawing;
- no need for trained annotators, guidelines, etc.

Human Parsing Object Detection

From Fixations to Bounding Boxes



イロン イロン イヨン イヨン

Human Parsing Object Detection

From Fixations to Bounding Boxes

Evaluation using CorLoc (intersection over union > 0.5):



Other Topics in Cognitive Data Science

- Build an incremental semantic role labeler which can be evaluated against eye-tracking data (Konstas et al. 2014);
- use synchronous grammars to align image structure with linguistic structure, e.g., for image description (Elliott & Keller 2013);
- build sequence models (e.g., based on HMMs) that human predict fixation behavior, either for images or for text;
- use eye-tracking data to improve unsupervised PoS tagging or parsing models.

(日) (同) (三) (三)

References

- Demberg, V., Keller, F., & Koller, A. (2013). Incremental, predictive parsing with psycholinguistically motivated tree-adjoining grammar. *Computational Linguistics*, 39, 1025–1066.
- Elliott, D., & Keller, F. (2013). Image description generation from visual dependency representations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (pp. 1292–1302), Seattle, WA.
- Konstas, I., Keller, F., Demberg, V., & Lapata, M. (2014). Incremental semantic role labeling with Tree Adjoining Grammar. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (pp. 301–312), Doha.
- Papadopoulos, D. P., Clarke, A. D. F., Keller, F., & Ferrari, V. (2014). Training object class detectors from eye tracking data. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (eds.), *Proceedings of the 13th European Conference on Computer Vision*, vol. V, (pp. 361–376), Zurich.

イロト イポト イヨト イヨト