A Bayesian Model of Memory for Text

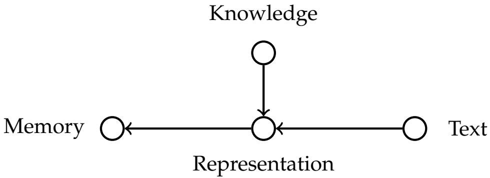
Mark Andrews
Psychology Department, Nottingham Trent University

mark.andrews@ntu.ac.uk
@xmjandrews
https://github.com/lawsofthought/bps-cog-2018

August 30, 2018

Memory for text

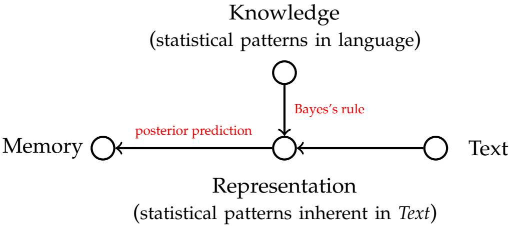
- The seminal study on memory for text is usually attributed to Bartlett (1932).
- From this, and schema based accounts of text memory (see, e.g., Bower, Black, & Turner, 1979), there has been something close to a consensus on the broad or general characteristics of human text memory.
- According to this general account — which we can summarize by the following schematic:



— the recognition or recall of items in a text is based on querying a representation of the text that is built up on the basis of background knowledge and experience.

Probabilistic account of memory for text

- We begin with the assumption that our background knowledge that is relevant for our memory of text is knowledge of the distributional statistics.
- Given a probabilistic language model, we may use Bayes's rule to infer the statistical patterns inherent in any given text.
- We may then predict, via posterior predictive inference, the words are and are not typical of this inferred statistical representation.
- As such, this provides a computational description of the previous schematic, i.e.,



Probabilistic topic models

- *Probabilistic topic models* (see, e.g., Griffiths, Steyvers, & Tenenbaum, 2007) have proved effective in capturing the statistical patterns that characterize coarse-grained "discourse topics".
- Topic models model each word in a text as a sample from one of many underlying probability distributions over a vocabulary.
- For example, here is a sample of 6 topics from an inferred model:

theatre	music	league	prison	rate	pub
stage	band	cup	years	cent	guinness
arts	rock	season	sentence	inflation	beer
play	song	team	jail	recession	drink
dance	record	game	home	recovery	bar
opera	pop	match	prisoner	economy	drinking
cast	dance	division	serving	cut	alcohol

- In this study, we used a *Hierarchical Dirichlet process topic model* (HDPTM), which is a non-parametric probabilistic topic model.

Inference and predictions in topic models

- Given this model, we obtain a formal model of background knowledge, text representation, and recall or recognition memory:
  - *Background knowledge*: From a corpus of example texts, we can infer the posterior distribution over the topics.
  - *Text representation*: For any given text, we can infer how well it is characterized by each inferred topic.
  - *Memory*: Given the text representation, we can then predict which words are typical or expected on the basis of that inferred representation.
- More formally, having inferred topics, given a new text w_j, we infer the posterior probability over π_j, which is the probability distribution over the discourse topics in w_j. We then use the posterior predictive distribution to infer the words that are typical of the topics inherent in w_j:

$$P(w_{ji}|w_j, \mathcal{D}) = \int P(w_{ji}|\pi_j, \mathcal{D})P(\pi_j|w_j, \mathcal{D})d\pi_j$$

Example text 1: Posterior predictions

Improve your mood and counteract stress: Ask anyone who exercises regularly and they will tell you that they always feel exhilarated at the end of a session — even if they had begun by feeling that they were not in the mood for exercise and had almost forced themselves to continue. Physical fitness also provides considerable protection against stress and the illnesses it can cause…maintain an already high level of stress.

relaxation feel *mind* exercise *people* *exercising* *stretching*
walking *stamina* *build* *energy* *routine* *walk* *swimming* fit
training *weight* *aerobics* *health* *yoga* *anxiety* *programme* *rest* session
fitness *increase* life *running* *week* *jogging* *rate* level *aerobic* *tension*
exercises regular stress *start* *begin* *muscles* *gym* *minutes*
mood *heart* *strength* *body* *muscle* physical day time

Example text 2: Posterior predictions

Developmental norms are an attempt to provide an indication of the ages at which one might expect ordinary children to show evidence of certain skills or abilities. Since children vary with respect to the ages at which they demonstrate any particular behaviour, norms represent an 'average' obtained from an examination of the developmental changes occurring in a large number of children. Data from a large sample . . . normally developing children.

data time carried play individual children items scores
cent found measured information average school samples sample extent adults family
reliability set population behaviour test parent ability testing aged
assessment adult score low childhood increase level result provide scale performance
tested parents measure results mother age compared
child home validity tests

Training and test corpus

- As our training corpus, we used the British National Corpus (BNC).
- From the entire BNC, we extracted approximately 200,000 texts, each with between 250 and 500 words.
- This gave a corpus of approximately 80m word tokens, and 50K word types.
- We randomly selected exactly 50 texts from the training corpus, and removed them.
- Having inferred the topics on the basis of the training corpus, we calculated the posterior predictive distribution for each of the 50 test texts.

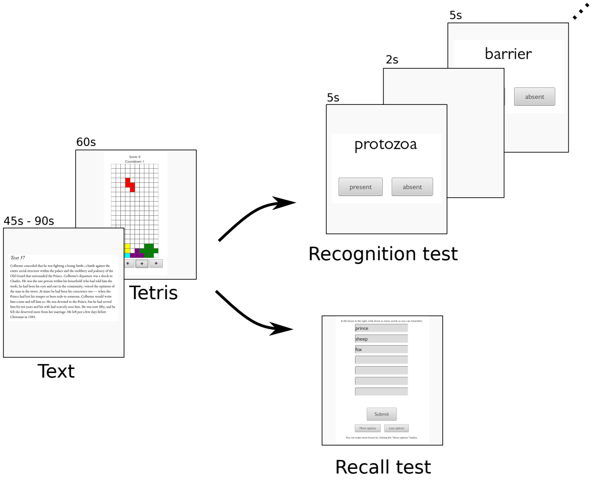
Comparison models

- We compare to predictions made by two *associative* models:
  1. From the BNC, we calculate the conditional probability of w_k and w_l as follows:
  $$\mathrm{P}_{\mathrm{C}}(w_k|w_l) = \frac{\mathrm{P}_{\mathrm{C}}(w_k, w_l)}{\mathrm{P}_{\mathrm{C}}(w_l)},$$
  where $\mathrm{P}_{\mathrm{C}}(w_k, w_l)$ is the co-occurrence probability of w_k and w_l.
  2. Using the *small world of word* association norms, we calculate conditional probability of word w_k given w_l with
  $$\mathrm{P}_{\mathrm{A}}(w_k|w_l) = \frac{A_{kl}}{\sum_{i=1}^{V} A_{il}},$$
  where A_{kl} indicates frequency that word w_k is stated as associated with word w_l.
- Given a text $w_j = w_{j1}, w_{j2} \ldots w_{jn_j}$, these models' predictions are:

$$\mathrm{P}_{\mathrm{C}}(w_k|w_j) = \frac{1}{n_j}\sum_{i=1}^{n_j} \mathrm{P}_{\mathrm{C}}(w_k|w_{ji}), \quad \mathrm{P}_{\mathrm{A}}(w_k|w_j) = \frac{1}{n_j}\sum_{i=1}^{n_j} \mathrm{P}_{\mathrm{A}}(w_k|w_{ji}),$$

Experiment

- 216 people (113 female, 103 male) participated in the experiment.
- Each participant read three randomly chosen texts (from the set of 50).
- Their memory was tested with either a recall or a recognition test.



Recognition memory analysis

- We analysed the predictions of each of the three models using multilevel logistic regression.
- For each model, we modelled the log-odds of recognizing as a linear function of the log of the model's predictions.
- In each case, this linear function varied randomly by participant, and by text, and had random intercept for each word type.
- Each model was evaluated using its *deviance information criterion* (DIC), which is a measure of out-of-sample generalization:

Topic model	Cooccurrence model	Association model	Null model
5232	5259	5320	5352

Recall memory analysis

- We analysed the predictions of each of the three models using multilevel multinomial logistic regression.
- For each model, we modelled the probability of recalling a word as probability mass function over the vocabulary that is (normalized exponential) linear function of the log of the model's predictions.
- In each case, this linear function varied randomly by participant, and by text.
- As in the recognition analysis, each model was evaluated using its DIC:

Topic model	Cooccurrence model	Association model
23798	26324	26825

Conclusion

- We have proposed a Bayesian account of how we form memories for spoken and written language.
- This account models how we use our background knowledge to form memories as a process of Bayesian inference of the statistical patterns that are inherent in each text, followed by posterior predictive inference of the words that are typical of those inferred patterns.
- We tested these predictions in a behavioural experiment with 216 participants.
- The results of the analysis from both the recognition and recall data provided strong evidence in favour of the Bayesian model relative to non-trivial alternative models.

References

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Bower, G., Black, J., & Turner, T. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*(2), 177–220.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244.