

Learning to Learn Implicit Queries from Gaze Patterns

Kai Puolamäki, Antti Ajanki, Samuel Kaski

Helsinki Institute for Information Technology HIIT
Department of Information and Computer Science
Helsinki University of Technology (TKK)

Collaborations

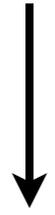
- Builds on previous work with David Hardoon and John Shawe-Taylor (published in AISTATS'07)
- Implicit feedback from eye movements - PinView (<http://www.pinview.eu/>). EU FP7 collaborative project with TKK and...
 - University of Southampton
 - University College London
 - University of Leoben
 - Xerox Research Centre Europe
 - celum gmbh
- PASCAL2

Contents

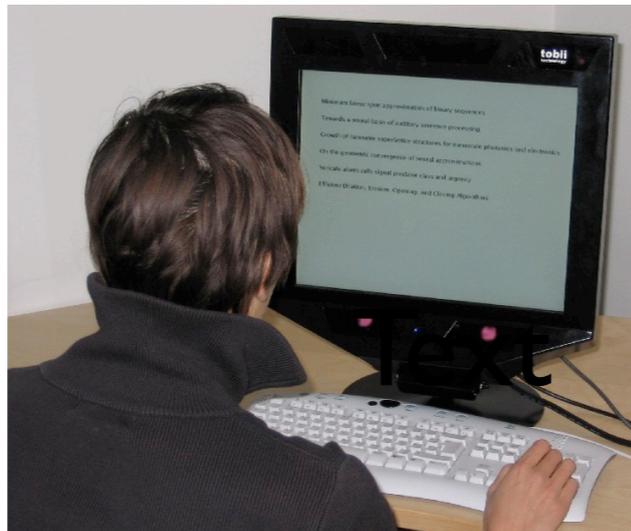
- **Application task**
- Machine learning task
- Experiment
- Conclusions

Previously unseen textual query

Task



Non-league
Football (soccer)
American football.
The Scottish
Front Office
Football is a text
sim which places
you at the helm of
an NFL football
team. Designed



Non-league football refers to football in England played at a level below that of the FA Premier League and The Football League. The term non-league came into common use before 1992 when the top football clubs in England all belonged to The Football League; all clubs who were not a part of The Football League were therefore non-League clubs.

Despite the non-league name, there are many leagues below the level of The Football League, and some, such as the Northern League, are almost as old as the League itself. Most of these are regional leagues organised by The Football Association, the sport's governing body in England, into a National League System (NLS). The NLS has seven levels or steps, and includes over 50 separate leagues, many with more than one division.

Revise ranking of yet unseen documents

Non-league football refers to football in England played at a level below that of the FA Premier League and The Football League. The term non-league came into common use before 1992 when the top football clubs in England all belonged to The Football League; all clubs who were not a part of The Football League were therefore non-League clubs.

Despite the non-league name, there are many leagues below the level of The Football League, and some, such as the Northern League, are almost as old as the League itself. Most of these are regional leagues organised by The Football Association, the sport's governing body in England, into a National League System (NLS). The NLS has seven levels or steps, and includes over 50 separate leagues, many with more than one division.

Application task

- We are looking for articles about ancient Rome
- We make a textual query “Rome”
- The search engine outputs articles also about the modern Rome
- Use the gaze pattern to infer a new query (that would be more specific for *ancient* Rome)
- Yet unseen documents can be re-ranked to reflect the new query

Application task: problems

- In the original training phase we show user several queries, gaze patterns are recorded and the user labels the relevant documents
- In the actual application...
 - The specific query has not been seen before
 - There is no explicit feedback
 - We must be able to infer the new query using implicit feedback (gaze patterns) only

Contents

- Application task
- **Machine learning task**
- Experiment
- Conclusions

Machine learning task

- Learn to infer relevance r of a document d for a previously unseen query w ,

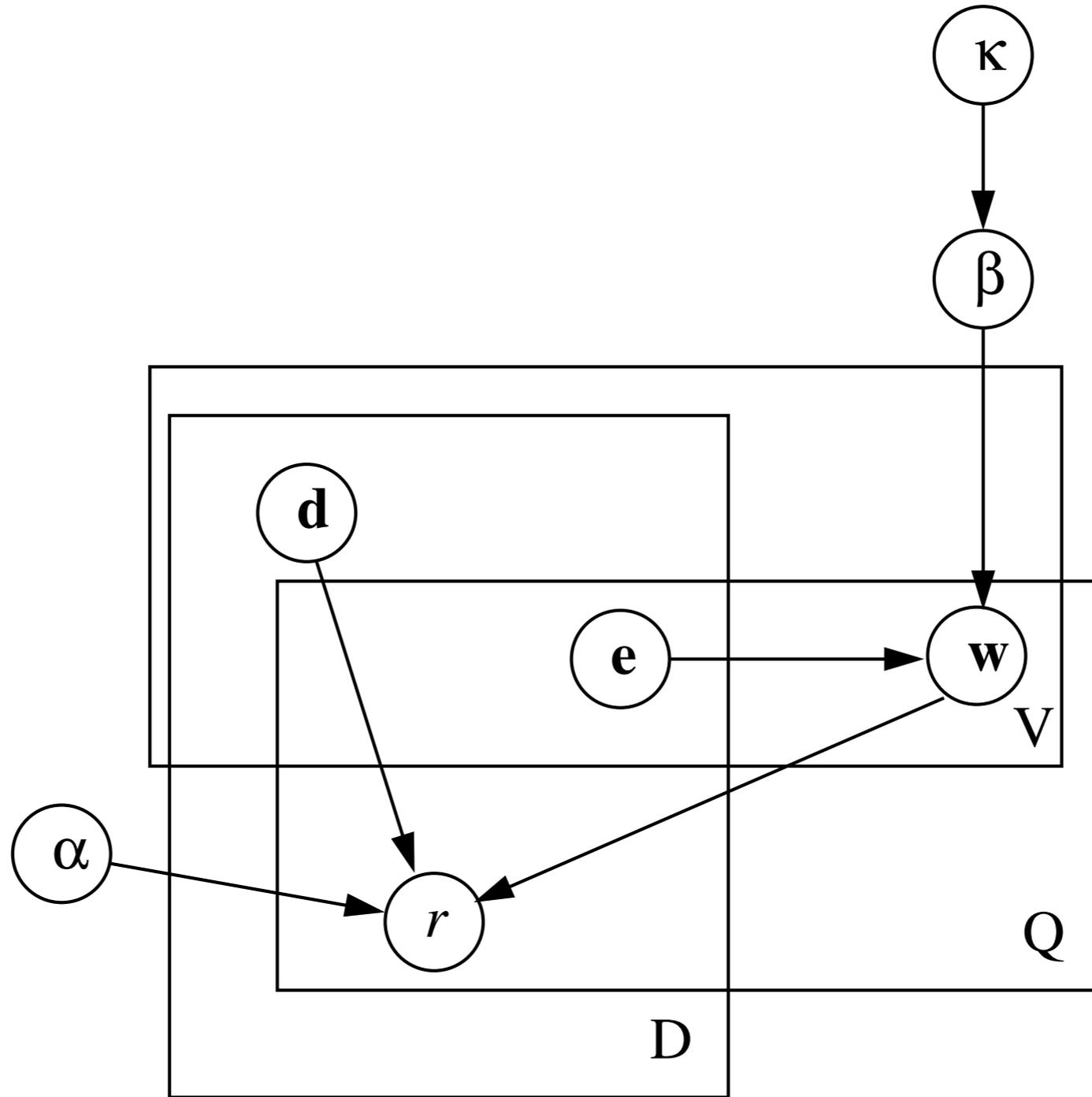
$$Pr(r|d,w;\alpha).$$

- Infer query w from gaze pattern e , $Pr(w|e;\beta)$
- Use inductive bias (parameters α, β) learned on original training queries
- “Learn to learn”

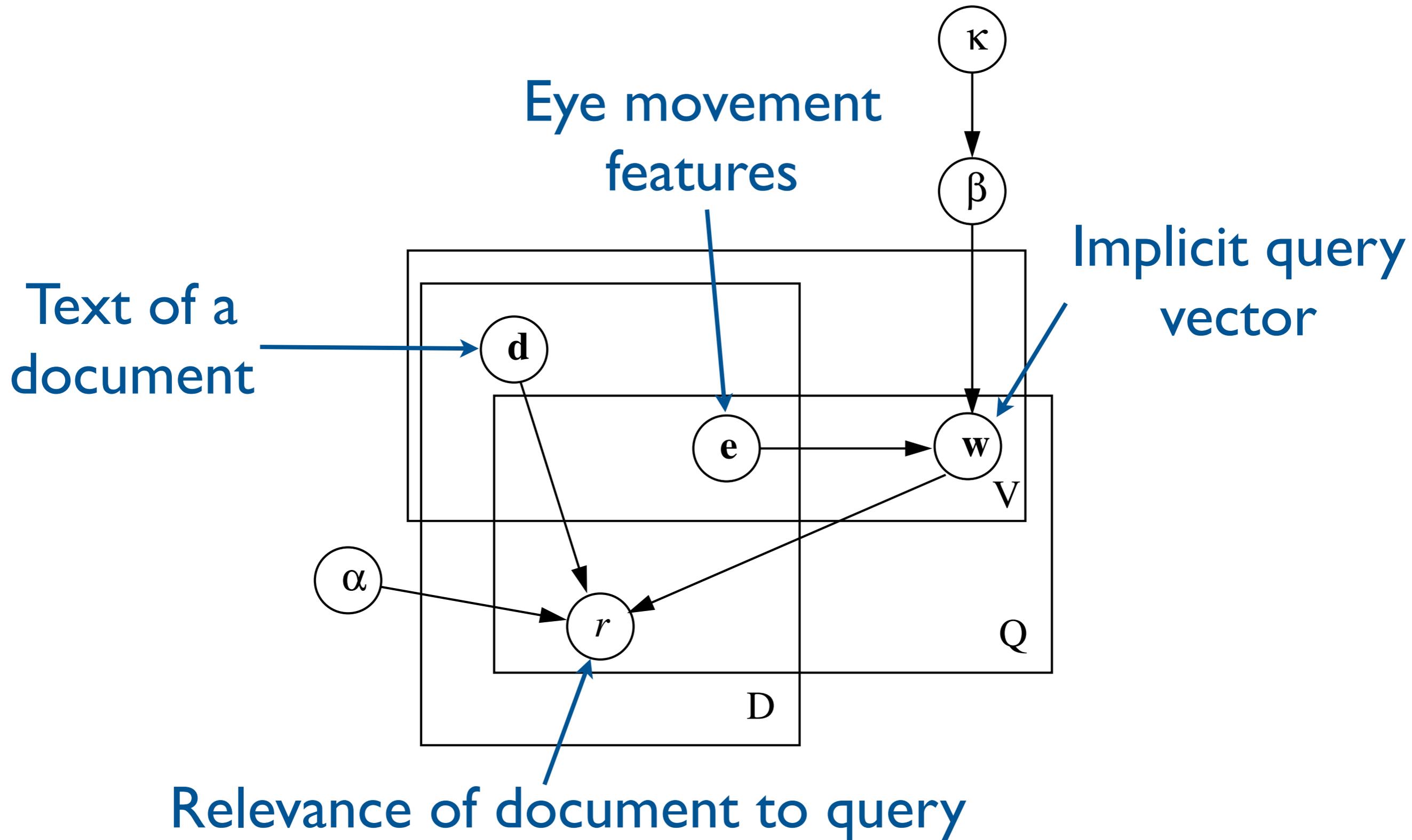
Model

- Assumption: there is a link between how a term is viewed and the term's importance in the query
- Bag-of-words representation
- Each term has...
 - Term features (IDF, # of characters, relative position)
 - Eye movement features (# of fixations, fixation durations etc.)
- Assumption: the semantic meaning of a term can be ignored, only the above features matter

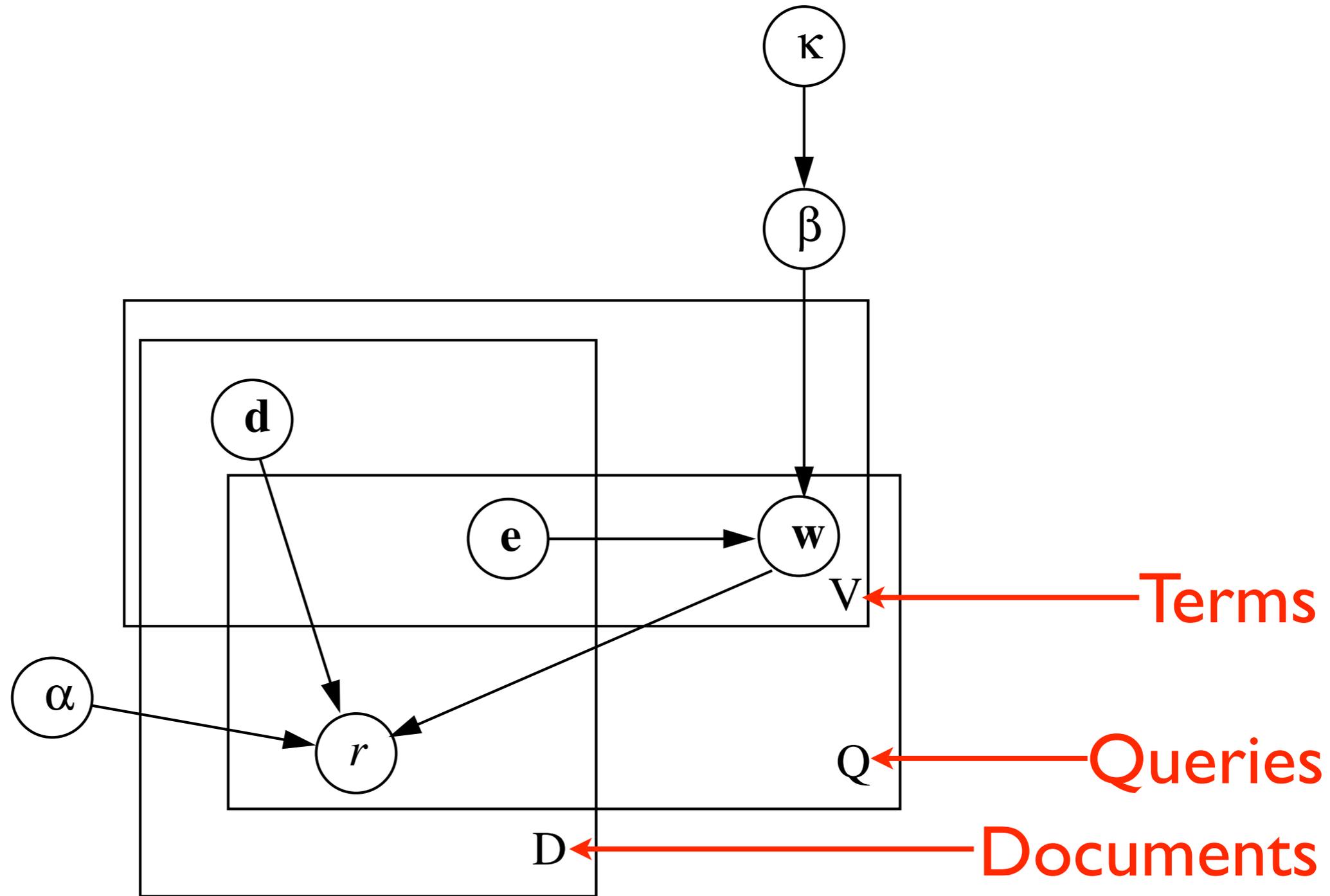
Model



Model

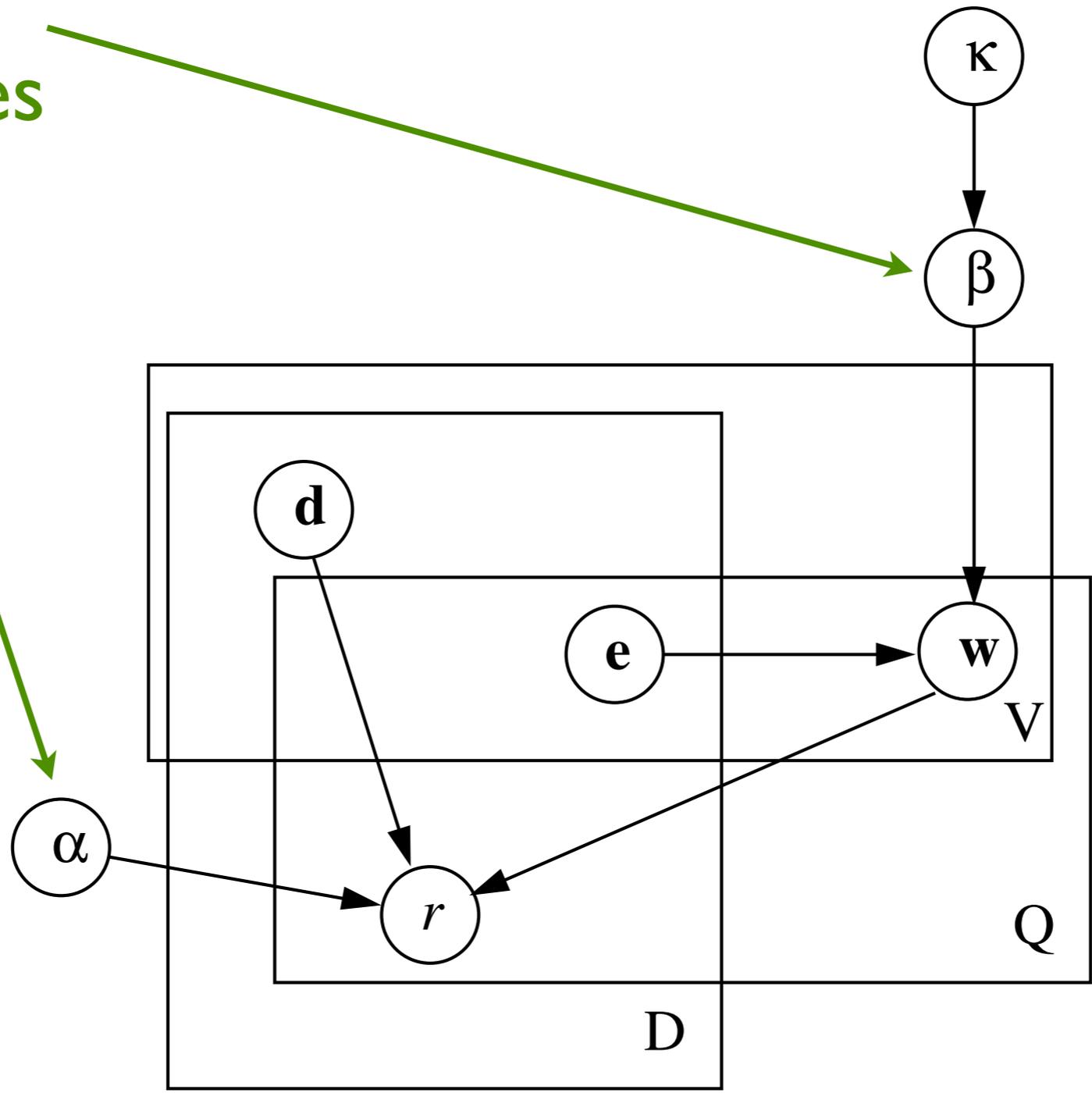


Model



Model

Parameters
common
for all queries



Training

- Learn the model:
 - Gaze patterns from several query tasks are shown, with explicit relevance feedback
 - Hyper-parameters α and β common to all tasks are learned by marginalizing all other parameters out (with appropriate approximations)

Training in two phases

- Learn the query:
 - Previously unseen search task
 - No explicit feedback
 - Using hyper-parameters common for all tasks we infer new query vector w

Connection to multitask, transfer learning

- Difference to multitask learning: we must learn a new classification task unlike the one seen before
- Potential application for multitask learning: learn different types of classification tasks (relevance from browsing, focused searching etc.)
- Transfer learning: inductive bias is coded into hyper-parameters that are made

Contents

- Application task
- Machine learning task
- **Experiment**
- Conclusions

Revising the ranking

- Documents are originally ranked using a textual query (“Rome”) and BM25
- User is shown top-5 documents, gaze pattern is recorded
- The ranking of top 6-10 documents are revised using the gaze pattern
- Simple and robust meta-search scheme (we used $\gamma=0.2$):

$$\text{score}(d) = \gamma \text{ ranking}(\text{BM25}) + (1 - \gamma) \text{ ranking}(\text{GAZE})$$

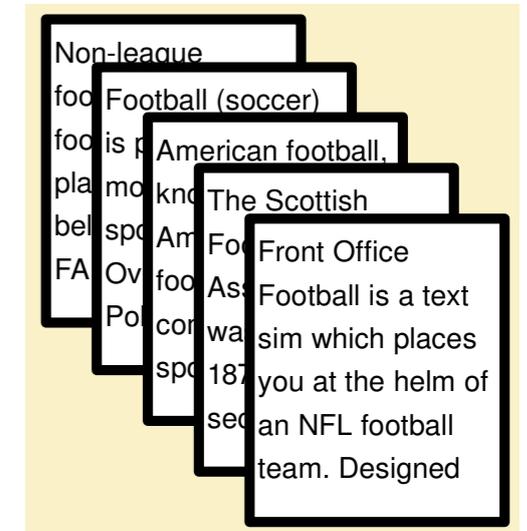
Experiment

- Corpus of truncated Wikipedia documents
- 3 naive users, 13 given search tasks

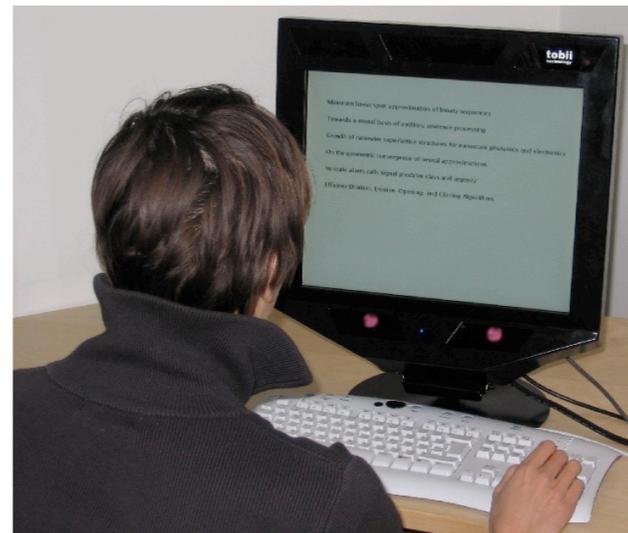
Task number	Desired topic	Search term
1	American football	football
2	Alternative medicine	medicine
3	Ancient Rome	Rome
4	Adhesive tape	tape
5	Environmental conservation	conservation
6	Seal (marine mammal)	seal
7	Extra-solar planets	planets
8	Visual nervous system	vision
9	Internet forums	forum
10	Marketing strategies	strategies
11	National libraries	libraries
12	British Royal Navy	navy
13	Space shuttles	shuttle

For each user and query:

1. User was shown top-5 BM25 documents



2. Gaze pattern was recorded



Non-league football refers to football in England played at a level below that of the FA Premier League and The Football League. The term non-league came into common use before 1992 when the top football clubs in England all belonged to The Football League; all clubs who were not a part of The Football League were therefore non-League clubs.

Despite the non-league name, there are many leagues below the level of The Football League, and some, such as the Northern League, are almost as old as the League itself. Most of these are regional leagues organised by The Football Association, the sport's governing body in England, into a National League System (NLS). The NLS has seven levels or steps, and includes over 50 separate leagues, many with more than one division.

3. A new query was inferred

4. Top 6-10 documents were re-ranked

Non-league football refers to football in England played at a level below that of the FA Premier League and The Football League. The term non-league came into common use before 1992 when the top football clubs in England all belonged to The Football League; all clubs who were not a part of The Football League were therefore non-League clubs.

Despite the non-league name, there are many leagues below the level of The Football League, and some, such as the Northern League, are almost as old as the League itself. Most of these are regional leagues organised by The Football Association, the sport's governing body in England, into a National League System (NLS). The NLS has seven levels or steps, and includes over 50 separate leagues, many with more than one division.

Experiment

- Combination of textual query and proposed method is better than textual ranking alone
- Measure: mean improvement in average precision (larger is better)

	Search task												
	1	2	3	4	5	6	7	8	9	10	11	12	13
$\Delta AVGPREC_{Prob}$	0.17	0.28	0.53	0.18	0.22	-0.71	-0.03	0.16	0.04	0.14	-0.05	-0.08	0.34
$AVGPREC_{BM25}$	0.50	0.61	0.13	0.50	0.50	1.00	0.91	0.50	0.78	0.70	0.71	0.50	0.42

Conclusion

- We can learn to learn relevance information from a new search task where only implicit feedback is available
- Implicit feedback improves ranking
- The full two-stage learning process is presented within one probabilistic model (no ad hoc assumptions)