Query Intention Acquisition

A Case Study on
Inferring Query Term Intent

Leif Azzopardi and
Maarten de Rijke

Outline

- A Brief Overview of QIA
- The focus of the Case Study
  - Building the Dataset
  - Analyzing the data
  - Inferring Intentions
  - Examining the affect on retrieval models
- Discussion
- Conclusions and Further Work

Query Intent Acquisition

- What is QIA?
  - It is the process of acquiring a query and analyzing the query to extract its the meaning, semantics and nature.
  - The goal is to ascertain what intentions the user has with respect to their *information need*.
  - Key for any Contextual IRS

Query Intent Acquisition

- QIA is performed in two main ways -
  - **Implicit** (inferring from query, user actions)
  - **Explicit** (questioning, structured input)
- Capturing Intentions
  - **Language** (Xpath, Xquery, SQL, etc)
  - **Interface** (tab, radio buttons, fielded input, etc)
- Here, we consider implicitly trying to capture the intentions of the user from their query

Inferring Query Term Intents

- What is it?
  - Automatic (unsupervised) extraction of the meaning of query terms
  - Given the context of the search (such as document features, user profile, etc)
- Benefits:
  - Simplifies query interface
  - Doesn't preclude usage
  - Ascertains a better understanding of the query
  - Attempts to bridge the semantic gap

Related Work

- Inferring structure in Boolean Queries
- Finding Dependencies between Terms
  - With Language Models
  - Using Statistical Co-occurrence
- Inferring fields from queries submitted to digital library IRS
  - Used Inference Networks to infer where a term in a query came from in the structured document
- Past work has tended to integrate the structure into the models.

Evaluation

- Evaluation criteria:
  - Reliability
    - how accurately can we infer intentions?
  - Robustness
    - How tolerate is the retrieval method to incorrect inferences?
  - Retrieval Effectiveness
    - Do they improve the retrieval performance of the IRS?

Case Study

- Aims of Case Study
  - To provide an analysis of inferring query intentions over these criteria
  - In particular, considering how:
    - The difficulty of inferring intentions, and
    - The ambiguity in the expressed intentions
  - Impact on the utility of the performing QIA
    - Over different (un)structured retrieval models

Case Study

- Comprised of Four Main Tasks
  - Building a dataset of query intentions
  - Analysis of Intentions
  - Automatically Inferring Intentions
  - A study of ambiguity and difficultly on retrieval performance.

Building Intentions Dataset

- Data
  - TREC Enterprise 2005 Email Forum
    - Aprox. 170,000 emails.
  - Using the Known Item Task
    - Topics KI1-150
      - 150 topics
      - Used Title Only queries (with length of 3-7)
    - Each topic has only one relevant document

Building Intentions Dataset

- The query was then tagged by assigning the query term to the field that it appears in the known item
  - Assumption: If the term appeared in two or more fields it was assigned to the field that is most salient. (date, from, subject, body)
  - If a term didn't appear then it was tagged with type="about"
  - Stopwords were ignored except those that seemed to indicator some form of intention

Building Intentions Dataset

- Example Mark Up
  - <QUERY><QUERYNO> KI8 </QUERYNO>
    - <body>noun</body>
    - <body>phrases</body>
    - <body type="about">exact</body>
    - <body type="about">match</body>
  - </QUERY>
  - <QUERY><QUERYNO> KI11 </QUERYNO>
    - <subject>tag</subject>
    - <subject>minutes</subject>
    - <date>9</date>
    - <date>june</date>
    - <date>2003</date>
  - </QUERY>

Query Characteristics

Field	Num. Occurrences	Num. Queries.
Date	19	13
From	41 (2)	24 (1)
Subject	323 (28)	111 (22)
Body	160 (61)	62 (34)
Indicators	105	76
Non-word Ind.	28	25
	Out of:	150 Queries

Ambiguity in Queries

- Here, we quantify how "ambiguous" a query according to how much of the query terms are in the known item.
  - The more query terms that are not in the document, the more ambiguous the query.

Level	Conditions	Total
(0) not	0-1 about terms	97
(1) somewhat	1+ about terms, < 50%	33
(2) very	> 50%	20

Inferring Intentions

- Classification of query terms
  - Unigram Model
  - Bi-gram Model
- Using the Odds Ratio to decide
  - Whether the query term belongs to a field or not.
  - The term was assigned the field with the highest odds.

Classification Accuracy

	Date	From	Subject	Body	Out of:
Date	84.2%	5.3%	0.0%	10.5%	19
From	11.9	78.6%	0.0%	9.5%	42
Subject	2.8%	2.5%	72.4%	22.3%	323
Body	0.7%	2.1%	66.2%	31.0%	145

- Unigram Only (similar performance with Bi-gram)
- Overall accuracy: 62.0% and 62.4%

Difficulty to Predict Intent

- Here, we quantify how “difficult” it is to infer the intent of query terms (given our classifier)
  - Easy (47)
  - Medium (37)
  - Hard (65)
- Relationship between Ambiguity and Difficulty
  - A Chi-Squared test indicates that the two factors are dependent ($p < 0.001$).
  - I.e more ambiguous, the more difficult to predict.

Retrieval Models

- How does ambiguity and difficulty affect different retrieval models?
  - Hypothesis: As ambiguity/difficultly increases, retrieval effectiveness will decrease.
  - We examined three Language Models
    - Standard,
    - Fielded, and,
    - Combination

Retrieval Models

- Standard Language Model

$$p(q \mid d) = \prod_{t \in q} p(t \mid d)^{n(t,q)}$$

- where t is a term, q is a query, n(t,q) is the number of times t occurs in q, and d is a document.
  - Bag of Words Terms (no structure)

Retrieval Models

- Combination Language Model

$$p(q \mid d) = \prod_{t \in q} \left(\sum_{x \in d} p(t \mid x, d) p(x \mid d) \right)^{n(t,q)}$$

  - Where x is a field in a document d
  - Settings:
    - Uniform - Each field combined equally
    - Automatic - Each field combined proportional to the number of times a query was classified as into each field
  - Structure is encoded on the document modeling side

Retrieval Models

- Fielded Language Model

$$p(q \mid d) = \prod_{x \in d} p(q_x \mid d_x)$$

  - Where q_x and d_x are the fields of the query and document
  - Settings:
    - Explicit - manually tagged structure
    - Fuzzy - assigned to a field if the Odds Ratio was greater than a small threshold
    - Strict - assigned to the field with the max. Odds Ratio.
  - Structure in document and query

Retrieval Effectiveness

	LM	Setting			
			Overall	**Ambiguity**	
				None	**Some/Very**
a	**Standard**		0.466d	0.537	0.337
b	**Combination**	**uniform**	**0.631**ade	**0.719**	**0.469**
c		**auto**	0.627ade	**0.719**	0.458
d	**Fielded**	**strict**	0.355	0.436	0.208
e		**fuzzy**	0.546ad	0.667	0.325
f		**explicit**	0.581ad	0.679	0.400

Retrieval Effectiveness

	LM	Setting				
			Overall	**Difficulty**		
				Easy	**Med.**	**Hard**
a	**Standard**		0.466d	0.501	0.409	0.470
b	**Combination**	**uniform**	**0.631**ade	0.701	0.636	**0.578**
c		**auto**	0.627ade	**0.705**	0.638	0.565
d	**Fielded**	**strict**	0.355	0.516	0.455	0.186
e		**fuzzy**	0.546ad	0.693	0.574	0.425
f		**explicit**	0.581ad	0.687	0.425	0.479

Findings

- Combination LM performs the best
  - Significantly better than standard, strict fielded and fuzzy fielded LMs.
- Strict Fielded LM performed the worse
- The fuzzy and explicit fielded LMs
  - Significantly better than standard and strict fielded LMs.

Discussion

- These results confirm our hypotheses:
  - As the ambiguity/difficulty increases then retrieval effectiveness decreases.
  - But, this was not so for the Standard LM
    - It is robust to the difficulty!
    - It doesn't use such information, but its performance is poor comparatively.

Discussion

- Our results confirms previous research that structured queries can out perform non structured queries, but
  - Against the combination model this is not the case!
  - The combination model uses structure on the document side only. This leads to a robust estimate which clearly is superior.

Conclusions

- Predicting the intentions of users is a difficult problem.
- The uncertainty within a query affects the quality of retrieval.
- Models need to be robust enough to tolerate ambiguity and difficulty.

Further Work

- Attaining better inference accuracy
  - Using other classifiers (for example, SVM)
  - Using other features
    - Such as using indicators of intent
- Formulating different retrieval models
  - Such as a combination of the combination and fielded approaches,
    - I.e.(Subject+Body) * from * date
- Building a simple natural querying language
  - for instance, by attaching meaning to stop word like terms