Intent Triage
Quantifying the Severity of Poor Performance on Intent Classes

Fernando Diaz
Yahoo! Labs

July 23, 2010
**query intent**: a detailed, unambiguous representation of the user’s information need.

**query intent class**: a group of related query intents (e.g. shopping, travel research, checking email).
Outline

Prior Work

Retrieval Performance on Different Intent Classes

Summary
State of the Art

Query Intent Class Detection

- $I^3R^1$
- query classification$^2$
- task intent classification$^3$
- vertical intent classification$^4$
- ...

---

1 Croft and Thompson, “The use of adaptive mechanisms for selection of search strategies in document retrieval systems”.


4 Arguello et al., “Sources of Evidence for Vertical Selection”. 
State of the Art
Responding to Query Intent Class

- specialized ranking models\(^5\)
- vertical presentation\(^6\)
- query suggestion\(^7\)
- personalization\(^8\)
- ...

---


\(^6\) Diaz, “Integration of News Content Into Web Results”.

\(^7\) Cao et al., “Context-aware query suggestion by mining click-through and session data”.

\(^8\) Teevan, Dumais, and Liebling, “To personalize or not to personalize: modeling queries with variation in user intent”.
Managing a Taxonomy of Intent Classes

• Sources of evidence
  • corpus: defines the possible means of satisfying intents
  • query, browsing logs: defines the types of intents users have
  • ...

• Defining intent classes
  • data mining: analyze the available evidence
  • feedback driven: explicit/implicit user feedback (e.g. ‘report a problem’ link)
  • ...

DCG\((q, \theta) = f(\text{Intent-Class}(q))\)

The performance of a retrieval system may be influenced by the query’s intent class.\(^9\)

\(^9\)Bian et al., “Ranking with Query-Dependent Loss for Web Search”. 
How do we compare performance across different intent classes?
Approach 1: Uniform Impression Weighting

\[
\frac{1}{|Q'|} \sum_{q \in Q'} \text{DCG}(q, \theta) = \frac{1}{|Q'|} \sum_{q \in Q'} f(\text{Intent-Class}(q)) \\
= \sum_{i \in I} \frac{|Q'_i|}{|Q'|} f(i)
\]

- **Assumption**: Each impression is equally important.
- **Implication**: Each intent class is weighted by its volume in traffic.
Approach 2: Uniform Query Weighting

\[
\frac{1}{|Q|} \sum_{q \in Q} \text{DCG}(q, \theta) = \frac{1}{|Q|} \sum_{q \in Q} f(\text{Intent-Class}(q)) = \sum_{i \in \mathcal{I}} \frac{|Q_i|}{|Q|} f(i)
\]

- **Assumption**: Each query is equally important.
- **Implication**: Each intent class is weighted by its representation in the query set.
Approach 3: Uniform Intent Weighting

\[
\sum_{i \in \mathcal{I}} \frac{1}{|\mathcal{I}|} f(i) = \sum_{q \in \mathcal{Q}} \frac{1}{|\mathcal{I}|} \left( \frac{1}{|\mathcal{Q}_{\text{Intent-Class}(q)}|} f(\text{Intent-Class}(q)) \right)
\]

\[
= \sum_{q \in \mathcal{Q}} \frac{1}{|\mathcal{I}| \cdot |\mathcal{Q}_{\text{Intent-Class}(q)}|} \text{DCG}(q, \theta)
\]

- **Assumption**: Each intent class is equally important.
- **Implication**: Each query is weighted by the number of other queries with its same intent class.
Approach 4: Revenue-Based Intent Class Weighting

\[
\sum_{i \in I} \frac{\$i}{\sum_{i' \in I} \$i} f(i)
\]

- **Assumption**: Each intent class should be weighted according to its impact on revenue.
- **Implication**: Must carefully select horizon for \$i.
Approach 5: User-Based Intent Class Weighting

\[ \sum_{i \in I} u_i f(i) \]

- **Assumption**: Each intent class should be weighted according to its impact on the user.
- **Implication**: Must carefully define \( u_i \).
Measuring User Impact

\( u_i \propto \)

- delight: user delight from system success for \( i \)
- frustration:\(^{10}\) user frustration from system failure for \( i \)
- recovery effort:\(^{11}\) user effort to recover from system failure for \( i \)
- quality of life impact: quality of life impact to the user as a result of system failure for \( i \)
- ...

---

\(^{10}\) Feild, Allan, and Jones, “Predicting searcher frustration”; White and Dumais, “Characterizing and Predicting Search Engine Switching Behavior”.

\(^{11}\) Feild, Allan, and Jones, “Predicting searcher frustration”.
Opportunity

• The most catastrophic impact is for users unable to get information during an emergency crisis (e.g. earthquake, hurricane).
• Information is the most disorganized at this point; difficult to recover with reformulation, deep inspection, etc.
• Information retrieval techniques are designed to support this kind of task.
Summary

• Can we be more formal about defining, detecting, organizing intent classes?
• Should we reason about intent class importance when computing system performance?
• What are the most important queries for users which systems are not currently supporting?


Li, Xiao, Ye-Yi Wang, and Alex Acero. “Learning query intent from regularized click graphs”. In: SIGIR ’08:


Teevan, Jaime, Susan T. Dumais, and Daniel J. Liebling. “To personalize or not to personalize: modeling queries with variation in user intent”. In: SIGIR ’08: Proceedings of the 31st annual international ACM SIGIR conference on