Visual Interactive Failure Analysis: Supporting Users in Information Retrieval Evaluation

Nicola Ferro, and Gianmaria Silvello
University of Padua, Italy

Marco Angelini, Giuseppe Santucci
“La Sapienza” University of Rome, Italy
Outline

- Motivations
- Models for Interaction
- The Prototype and Application Examples
- The Domino Effect
- Final Remarks and On-Going Works
Motivation

Reduce the time needed to understand and analyze the behavior of an IR system providing a **visual interactive tool for evaluation**
Motivation

Reduce the time needed to understand and analyze the behavior of an IR system providing a visual interactive tool for evaluation

Failure Analysis

or “Understand What is Wrong”.

Is it better to re-rank or to re-query?
Motivation

Reduce the time needed to understand and analyze the behavior of an IR system providing a visual interactive tool for evaluation

Failure Analysis

or “Understand What is Wrong”.

Is it better to re-rank or to re-query?

What-if Analysis

“Try to anticipate the effect of a modification of the system”.

Visual Analytics is not only a mean to improve the presentation of results, but also a mean allowing the users to analyze and interact with data.
Models for Interaction

- Rank Gain/Loss Model
- Clustering via Learning to Rank
- What-If Analysis Model
Models for Interaction

Rank Gain/Loss Model

DCG is good if we want to compare performances between different systems but what if we want to look at misplaced documents?
Models for Interaction

Rank Gain/Loss Model

DCG is good if we want to compare performances between different systems but what if we want to look at misplaced documents?

Relative Position = RPos let us determine how much a document is misplaced with respect to its ideal rank.
Models for Interaction

Rank Gain/Loss Model: R_Pos

The graph illustrates the Rank Gain/Loss Model (R_Pos) with various lines representing different positions or categories. The x-axis represents the number of steps or trials, while the y-axis shows the gain or loss in rank. The graph shows how the rank changes as the number of steps increases.
Models for Interaction

**Rank Gain/Loss Model: R_Pos**

By means of R_Pos we can say if a document is ranked too soon or too late. But, can we quantify “too soon” and “too late”?

**How much these misplaced documents impact on the dcg?**

---

lunedì 20 agosto 12
Rank Gain/Loss Model: Delta Gain

Models for Interaction

Information Interaction in Context (IIiX 2012)
Nijmegen, the Netherlands, August 21-24, 2012

Gianmaria Silvello
Models for Interaction

Rank Gain/Loss Model: Delta Gain

- Rank Position (R Pos)
- Delta Gain

- DCG: 10.592209799285
- DeltaGain: -0.39647972634112
- RPos: -18

Graph showing the DCG and Delta Gain over rank.
Failure Analysis
Models for Interaction

Clustering via Learning to Rank

Learn the ranking model of the IR system under investigation in order to simulate the way in which it ranks the documents.
Clustering via Learning to Rank

- **Training Data**: \( \{ q_1, x_1^{(1)}, x_2^{(1)}, \ldots, x_m^{(1)}, y^{(1)} \} \) for each query \( q_1 \), \( q_2 \), \( q_n \).

- **Learning System**: Produces ranking scores \( y^{(2)} \) for each query.

- **Ranking System**: Uses the scores from the learning system to produce a ranked list.

- **Test Data**: \( \{ q, x_1, x_2, \ldots, x_m \} \) for a new query \( q \).

- **Model \( h \)**: Takes the test data and produces a prediction \( h(x) \).

The diagram illustrates the process of clustering via learning to rank, where the learning system takes the training data and produces ranked results, which are then used by the ranking system to produce a final prediction for the test data.
Clustering via Learning to Rank

- Lever on the clustering hypothesis
- Group together the documents which are similar from the considered ranking model point-of-view

Generating the clusters

1. Submit each doc in Dj as a query and retrieve a set of docs Di;
2. determine Ci = Dj \ Di;
3. ranking the documents in Ci by employing the learned ranking model
Models for Interaction

What-if Analysis

Visual comparison of Ranked Result Cumulated Gains

Topic ID: 351 Experiment: bbn1
What-if Analysis

Visual comparison of Ranked Result Cumulated Gains

Topic ID: 351 Experiment: bbn1

Gianmaria Silvello
To Summarize: Data Pipeline

1. Experiment
   - The ideal ranked document list
2. Visualization
   - Clusters of documents for the selected $t \in T$
   - New ranked document list for a new topic $t_k \in T$, $t_k \neq t$
3. Machine learning algorithm
   - The new ranked document list for the selected $t \in T$
4. Set of ranked document lists for each $t \in T$
5. Rank Gain/Loss Model
   - Learned Rank Model
   - What-if Analysis Model

$T \xrightarrow{D} \text{Experiment} \xrightarrow{\text{Visualization}} \text{Visualization} \xrightarrow{\text{Machine learning algorithm}} \text{Set of ranked document lists for each } t \in T$
Application Example: Free Movement

Visual comparison of Ranked Result Cumulated Gains

Topic ID:351 Experiment: bbn1

R_Pos | ΔGain
---|---
OK BELOW ABOVE | OK GAIN LOSS

Graph showing the comparison of ranked results cumulated gains with different lines representing different experiments and gains.
Application Example: Free Movement

Visual comparison of Ranked Result Cumulated Gains

Topic ID:351 Experiment: bbn1

R_Pos

ΔGain

0 50 100 150 200

0 5 10 15 20 25

Ideal
New Optimal
Old Optimal
New Experiment
Old Experiment
Application Example: Capped Movement

Visual comparison of Ranked Result Cumulated Gains

Topic ID:351 Experiment: bbn1

R_Pos vs ΔGain

- Ideal
- New Optimal
- Old Optimal
- New Experiment
- Old Experiment

Gianmaria Silvello
Application Example: Capped Movement

Visual comparison of Ranked Result Cumulated Gains

Topic ID: 351 Experiment: bbn1

Visualization of ranked result cumulative gains for different conditions and experiments.
Application Example: New Entry Evolution

Visual comparison of Ranked Result Cumulated Gains

Topic ID: 351 Experiment: bbn1

Gianmaria Silvello
Application Example: New Entry Evolution

Visual comparison of Ranked Result Cumulated Gains

Topic ID: 351 Experiment: bbn1

R_Pos

△Gain

Information Interaction in Context (IIiX 2012)
Nijmegen, the Netherlands, August 21-24, 2012
Gianmaria Silvello
The Domino Effect

How does the change for topic 351 affect the other topics?

Visual comparison of Ranked Result Cumulated Gains
Topic ID: 351 Experiment: bbn1

[Graph showing the comparison]
The Domino Effect

The change for 351 worsens the DCG curve of topic 355
The Domino Effect

The change for 351 improves the DCG curve of topic 400
The Domino Effect

Aggregate view for the whole set of documents before and after the movement

**Boxplot**
- Upper limit
- Upper quartile
- Median
- Lower quartile
- Lower limit

**DCG (Aggregate)**

<table>
<thead>
<tr>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>150</td>
</tr>
<tr>
<td>200</td>
</tr>
</tbody>
</table>

Legend:
- Ideal
- Optimal
- Experiment
- Median
- Upper/Lower quartile
- Upper/Lower limit
- New mean DCG
- Old mean DCG
Final Remarks and On-Going Work

- We presented the visual interactive tool allowing analysts to perform failure and what-if analyses.
- We described the prototype implementing an actual fusion between IR evaluation and visual analytics tested on the TREC7 collection.
- We are performing additional tests employing different learning to rank algorithms to construct the clusters.
- We are investigating whether (and how) custom features extraction and selection may allow us to understand on which component of the IR system a change of ranking has an impact on.