Metrics on Collaborative Information Retrieval

Wei Xia

January 10, 2024

Basic Concepts of Collaborative Information Retrieval and Information Retrieval Metrics

- **goals of collaborative information retrieval:** A systematic approach to improve information finding by leveraging group interactions and shared strategies.
- **Addressing information overload:** Tackling the challenge of too much information.
- **Purpose and importance of information retrieval metrics:** Evaluating the effectiveness of retrieval systems.

Collaborative Information Retrieval

- **Definition:** Collaborative process of explicit information seeking that benefits all participants involved
- **Activities:** Social searching, concurrent search, collaborative exploratory search, co-browsing, and collaborative information synthesis
- **Main processes:** Information seeking and collaborative grounding
- **Research areas:** Human-computer interaction, information retrieval, library and information science, and more

Relationship with Computational Linguistics and Intersection with Collaborative Information Retrieval

- **Definition and scope of computational linguistics:** The scientific study of language from a computational perspective.
- **Role of computational linguistics in information retrieval:** Enhancing search through linguistic algorithms.
- **How computational linguistics techniques can enhance collaborative information retrieval:** Advanced processing for better search results.
- **Examples of text analysis, natural language processing, and topic modeling:** Practical applications in improving search accuracy.

Metrics for effectiveness in collaborative search process

User Engagement

- Focused on user engagement as a key indicator of search performance in collaborative environments.
- Engagement measured through a Likert scale, assessing user responses to various aspects of their search experiences.
- This approach emphasizes the subjective experience of users as a vital component of information retrieval success.

Evaluation of Search and Learning Resources

- Analyzed time spent on Lookup and Exploratory searches and the use of learning resources.
- Employed statistical analysis on log data to relate resource use to search success and learning outcomes.
- Used structured diaries and a five-point Likert scale to assess the frequency of learning resource usage.

Leeder and Shah's Quality of Search Analysis

- Focused on evaluating the quality of searches through parameters such as usage time, variety of resource types, and the overall quality of the search resources identified.
- Developed a comprehensive scoring rubric to quantitatively rate the quality of sources discovered during the search.

Table: Scoring rubric

The recent history, current condition, and future prospects for the sector as a whole
The recent history, current condition, and future prospects for companies within the sector
Which companies might be winners and/or losers in the sector
Information which would be important to a potential investor in the sector

Based on Na and Lee, 2016

- Investigated the query behaviors of CIS (Collaborative Information Searching) participants with a focus on effectiveness and diversity.
- Identified 4 key indicators: the number of queries, query vocabulary richness, and similarity of query results, and Successful Query Rate.
- Analyzed five distinct query behaviors: query terms used, dwell time on pages, bookmarking behavior, total web pages visited, and the extent of content coverage.
- Highlighted the significance of diverse query formulation and high query success rates in collaborative search scenarios.

Query Vocabulary Richness based on Na and Lee

Equation QVR

$$\text{QVR} = \frac{\Gamma(\text{Unique Query Terms})}{\Gamma(\text{Queries})}$$

- Measures the diversity of terms used across queries.

Query Results Similarity based on Na and Lee

Equation QRS

$$\text{QRS}(p_1, p_2) = \frac{|\psi(p_1) \cap \psi(p_2)|}{|\psi(p_1) \cup \psi(p_2)|}$$

- Calculates the overlap between document results from different queries.

Successful Query Rate based on Na and Lee

Equation sq

$$SQ = \frac{\#(\text{Queries with items collected})}{\#(\text{Queries})}$$

- Assesses the effectiveness of queries based on the number of items saved.

Based on Shah et al. 2017

- Focused on mathematical concepts used in measuring group search behavior.
- It included metrics like effectiveness, efficiency, and query diversity.
- These metrics are crucial for analyzing the dynamics of group search and collaboration.
- The analysis aids in understanding how groups interact with information search and retrieval processes.

Effectiveness of Group's Web Search based on Shah et al. 2017

Effectiveness

$$\text{Effectiveness}(i) = \frac{|\cup_i \{p_n(\text{dwell time}_{p_n} \geqslant 30\text{secs})\}|}{|C(i)|}$$

- Measures the proportion of web pages visited for at least 30 seconds to total distinct web pages visited.
- $|C(i)|$ represents the total number of distinct web pages visited by the group.
- It assesses the extent to which the group's web search

Efficiency of Group's Search based on Shah et al. 2017

Efficiency

$$\text{Efficiency}(i) = \frac{\text{Effectiveness}(i)}{|Q(i)|}$$

- Normalizes effectiveness by the number of distinct queries used.
- $|Q(i)|$ refers to the number of distinct queries used by the group.
- evaluates if the group achieves high-quality outcomes,by Using a minimal number of queries.

Relevant Coverage based on Shah et al. 2017

Relevant Coverage

$$RC(i)$$

- Focuses on distinct web pages found and marked as relevant based on collected snippets.

Based on Shah et al. 2015

- This section covered key measures for evaluating exploratory search tasks.
- It highlighted the importance of understanding both the quantity and quality of information exposure and relevancy.
- Metrics such as unique coverage, LD metric, and information search entropy were introduced.
- These measures provide comprehensive insights into user behavior during exploratory searches.

Query Diversity based on Shah et al. 2015

Query Diversity

$$\text{Query diversity} = \text{Mean}\{\text{lev}(Q_a, Q_b)\}, \forall Q_a, Q_b \in Q(i), Q_a \neq Q_b$$

- Calculates the average Levenshtein Distance between pairs of distinct queries.
- Reflects the diversity of search queries.
- quantifies the variation between different queries within a group

Information Exposure based on Shah et al. 2015

- Pertains to the quantity of information encountered during an exploratory search.
- Coverage and unique coverage are key metrics.

Unique Coverage

$$\text{UniqueCoverage}_t = \text{Coverage}_t \setminus \bigcup_{t_i \in T \setminus \{t\}} \text{Coverage}_{t_i}$$

- U represents the set of unique web pages accessed.
- highlights the unique contributions of individual users or teams to the overall search process.

LD Metric based on Shah et al. 2015

Difficulty Level (LD_{wp_i})

$$LD_{wp_i} = \frac{-1 \cdot n\{wp_i\}}{|U|}$$

- LD_{wp_i}: Represents the difficulty level associated with discovering the web page i.
- $n\{wp_i\}$: The number of users or teams that accessed web page i.
- $|U|$: The total number of unique web pages accessed by all users or teams.
- A higher LD_{wp_i} score suggests that the web page was less frequently visited, indicating a higher difficulty level in discovering that page.

LD Metric for User/Team Score based on Shah et al. 2015

User/Team Score (LD_t)

$$LD_t = \frac{\sum_{i=1}^{|\text{Coverage}_t|} LD_{wp_i}}{|\text{Coverage}_t|}$$

- LD_t: The overall difficulty score for a user or team t based on their information discovery experience.
- The numerator, $\sum_{i=1}^{|\text{Coverage}_t|} LD_{wp_i}$, adds up the difficulty levels for all web pages accessed by the user or team.
- $|\text{Coverage}_t|$: The total number of web pages visited by the user or team.
- This score aggregates the difficulty levels of all accessed pages, reflecting the challenge faced by the user or team in their search.

Based on Soulier et al.,2016

- This section emphasized the evaluation of collaborative group performance in search tasks.
- It introduced various metrics such as precision, recall, and F-measure at the group level.
- The metrics provide a framework to assess the effectiveness of group collaboration in information retrieval.
- Such evaluations are key to optimizing search strategies and tools for collaborative environments.

Information Relevancy based on Soulier et al.

Relevant Coverage (Relevant Coverage$_t$)

- **Formula:** Relevant Coverage$_t$ = Coverage$_t$ $\cap$ U_r
- Represents the intersection of the documents retrieved (Coverage$_t$) and the set of relevant documents (U_r).
- Measures the quantity of retrieved documents that are relevant to the search query

Unique Relevant Coverage (UniqueRelevant Coverage$_t$)

- **Formula:** UniqueRelevant Coverage$_t$ = UniqueCoverage$_t$ $\cap$ U_r
- Focuses on the unique documents retrieved by a user or team (UniqueCoverage$_t$) that are also relevant (U_r).
- Highlights the unique and relevant contributions of the user or team to the search task.

Entropy of a Query based on Soulier et al.

Entropy

The entropy of a query is a measure of the randomness or diversity of the words within that query. It is calculated using the formula:

$$\text{Entropy}_{Q_a} = \sum_{u=1}^{|\text{unigrams}_{Q_a}|} -p_u \log_2 p_u$$

Where:

- $|\text{unigrams}_{Q_a}|$ is the count of unique words in the query Q_a.
- p_u is the probability of occurrence of the u^{th} unigram in the query.
- The sum of $-p_u \log_2 p_u$ for all unigrams gives the total entropy.

A higher entropy indicates a query with greater diversity or complexity.

Average Information Content of Queries based on Soulier et al.

Average Information Content

The Average Information Content is a measure of the average diversity or complexity of queries made by a user or team. It is calculated using the formula:

$$\text{AvgInfoContent}_t = \frac{\sum_{a=1}^{|Q_t|} \text{Entropy}_{Q_a}}{|Q_t|}$$

Where:

- The sum $\sum_{a=1}^{|Q_t|} \text{Entropy}_{Q_a}$ represents the total entropy across all queries made by the user or team t.
- $|Q_t|$ is the total number of queries made by the user or team t.

This metric reflects the richness of the queries on average, where a higher value suggests more diverse or complex queries.

Performance Evaluation based on Soulier et al.

- Focuses on the efficiency and effectiveness in finding useful information.
- The F-score metric can be used for performance evaluation.

Extension of Evaluation Metrics

- Soulier et al. extended metrics to include group-level precision and recall.
- These metrics assess collaborative group performance based on collective document selection.

Group Coverage and Relevant Coverage

- Group Coverage ($Cov(g)$): Total distinct documents retrieved by the group.
- Relevant Coverage ($RelCov(g)$): Distinct relevant documents retrieved.
- Universe (U): Entire set of retrievable documents.

Precision and Recall

Precision ($Prec(g)$)

$$\text{Prec}(g) = \frac{\text{RelCov}(g)}{\text{Cov}(g)}$$

Recall ($Recall(g)$)

$$\text{Recall}(g) = \frac{\text{RelCov}(g)}{U_r}$$

F-measure (F-Score)

$$F(g) = \frac{2 * \text{Prec}(g) * \text{Recall}(g)}{\text{Prec}(g) + \text{Recall}(g)} \tag{1}$$

Unique Coverage Metrics by Htun et al.

- Different approach to coverage calculation, focusing on discovered documents.

Traditional IR Evaluation Metrics

- Number of Queries (Q), Average Query Length (QL), Query Success (QS)
- Number of Viewed Documents (V), Viewed Documents per Query (VDQ)

references

- Andolina, Salvatore, et al. "Querytogether: Enabling entity-centric exploration in multi-device collaborative search." Information Processing Management 54.6 (2018): 1182-1202.
- Wu, Dan, Shaobo Liang, and Wenting Yu. "Collaborative information searching as learning in academic group work." Aslib Journal of Information Management 70.1 (2018): 2-27.
- Leeder, Chris, and Chirag Shah. "Collaborative information seeking in student group projects." Aslib Journal of Information Management 68.5 (2016): 526-544.
- Na, Kyoungsik, and Jisu Lee. "When two heads are better than one: Query behavior, cognitive load, search time, and task type in pairs versus individuals." Aslib Journal of Information Management 68.5 (2016): 545-565.
- Shah, Chirag, Chathra Hendahewa, and Roberto González-Ibáñez. "Two's company, but three's no crowd: Evaluating exploratory web search for individuals and teams." Aslib Journal of Information Management 67.6 (2015): 636-662.

references

- Shah, Chirag, Chathra Hendahewa, and Roberto González-Ibáñez. "Two's not always company: collaborative information seeking across task types." Aslib Journal of Information Management 69.1 (2017): 22-35.
- Soulier, Laure, Lynda Tamine, and Chirag Shah. "MineRank: Leveraging users' latent roles for unsupervised collaborative information retrieval." Information Processing Management 52.6 (2016): 1122-1141.