

# 08\_06\_2023

## Writing progress - Gaps

### Gap 1 - Lack of forward applications of labeling techniques

#### 1. Introduction:

1. How often are the proposed approaches used by future researchers?
2. Which **role** do they play in these future applications?

#### 2. Verification (applied only to novel fully automated approaches)

1. **Forward search** of citing documents using SemanticScholar.
2. **Analysis of citations statements** in order to understand the role of the original work.
3. Additionally, we record:
  1. The presence of **code repositories**.
  2. The presence of **author overlap**.

#### 3. Encountered **forward implementations**

1. 11/27 have forward implementations
2. In most cases forward applications are a small percentage of total ones,
  1. 9/11 only 1 or 2 forward implementations.
  2. The two exceptions primarily propose Topic modeling approaches.
  3. Most of the time these approaches are used as **baselines** (rather than active tools in the research).

#### 4. Recurring authors

1. Overlapping authors in 40% of cases.
2. 4/7 cases this is the only type of forward implementation
3. Having a labeling approach re-used by (some of) the same authors can suggest that the technique has not found a wider audience among researchers operating in the field.

#### 5. Availability of code repositories

1. 8/27 documents have code repositories (6 of which have forward implementations)
2. Of the 19 without one:
  1. 14 have no forward representation
  2. 3 have author overlap

#### 6. Addressing the gap

1. If the studies aims to propose a new technique → Add an accessible code repository and proper documentation.

2. If the study requires a labeling technique → Use Table 7.1 (and the chapter “Insights from individual studies” as a guide.

## Gap 2 - Shortcomings in the quality evaluation procedures

### 1. Introduction:

1. Contextualise the lack of quality evaluation.
2. Explore how the few encountered evaluations are shaped (and their limitations).

### 2. Verification:

1. Trivial verification from Data Item 16 (22/108 documents including quality evaluation steps.
2. Further characterisation - Quality evaluation vs **focus** on labeling (are labels mostly evaluated when labeling is of primary focus?)
  - ◆ This is the case for 17/22 papers.
  - ◆ 3/5 without a primary focus provide an informal evaluation

### 3. Evaluation criteria and dimensions

1. How many **(different) evaluation criteria** in each study.
  1. 7/22 show more than one criteria (in values ranging from 2 to 5).
  2. 4/7 show a **mix** of human and automated approaches.
2. To which degree are **multiple dimensions (or factors)** considered in the evaluation.
  1. I.e. how many papers go beyond the good fit / bad fit evaluation?
  2. 6/22 consider more than one factor.
    1. Relevance, coverage and discrimination (3 documents)
    2. Relevance and adherence to gold-standard labels (2 documents)
    3. Finally, one document "*One Rating to Rule Them All? Evidence of Multidimensionality in Human Assessment of Topic Labeling Quality*" provides a deeper outlook on the topic
      - Here, evaluation is made based on 14 items → grouped into two factors by means of EFA (Suitable and Objectionable).
      - These findings are used again at the end of the section.

### 4. Information considered by the evaluation criteria

1. Four distinct kinds of information are found: terms, documents, gold-standard labels, other generated labels.
2. Most only consider one kind of information
  1. 9/22 only consider terms
  2. 5/22 only documents
  3. 3/22 use other labels (Coverage metric)

4. 6/22 use gold standard labels ← Existence of gold standard labels implied that we are not in a “practical” scenario
3. Only 2/22 document use **both terms and documents**
5. **Considerations on human ratings**
  1. Human rating 16/22 documents ← Most prevalent evaluation type
  2. Consideration on **subjectivity** and **bias** should be made ← Supported by Hämäläinen and Alnajjar (2021)
6. **Addressing the gap**
  1. Strive to include at least a basic form of (preferably non-informal) quality evaluation activity
  2. Include more than one criteria (technique) for evaluation.
    - ♦ Try to propose a mix of manual and automated approaches
  3. Employ a multi-dimensional approach to evaluation
    - ♦ Use the findings in Hosseiny Marani et al. (2022), namely
      - ♦ Quality evaluation of topic labels has multiple dimensions (two where found: Suitable & Objectionable)
      - ♦ Exploiting these additional dimension can ultimately provide more robust assessment
      - ♦ Develop means of assessing performance that go beyond asking whether a given data point is assigned the correct label
  4. Use multiple types of information in the evaluation procedure
  5. Involve domain experts to combat bias and subjectivity

### Gap 3 - Lack of sufficient research on alternative representations

1. **Introduction:**
  1. Q1: To which degree is research in this direction lackluster?
  2. Q2: Are these alternative representations more effective?
2. **Verification**
  1. Q1 is easily verifiable, only 13/108 use alternative representations
  2. Q2 requires an **individual analysis** to verify whether a **formal comparison to n-gram** labels exists
    1. Only two papers provide a formal comparison
      1. Aletras et al. (2017) → Information retrieval task + human-based evaluation
      2. Smith et al. (2017) → Best / Worst separate evaluation procedure
    2. Two studies provide n-gram labels but lack formal comparisons
3. **Backward search of alternative representations**

1. Stemming from Intro. and Related Work of the 13 papers
2. Three documents are gathered
  1. Wan and Wang (2016a) highlights substantially higher metrics for Relevance, Coverage and Discrimination for summaries
  2. Barawi et al. (2017) finds similar improvements (using a 4-level scale) for sentence labels
  3. The last paper does not have a formal comparison step
3. Here not enough data is found to reject the gap
4. **Addressing the gap**
  1. Difficult given the little existing research
  2. Use domain knowledge to choose the most appropriate representation
  3. Produce (and potentially share) more than one representation for the same topic
  4. Try to include formal comparisons among the different representations

## Upcoming activities

- Insights from selected studies (9-15)
- Contributions / Conclusion / Related work (16-22)

## Thesis title

In addition to the traditional informational output provided as the final result of any SLR, our work has a particular focus on the methodological rigor followed throughout the entirety of the process, where each step has been clearly structured and justified.

Should the title reflect the strongly methodological focus of this review?

Some proposals:

- "Enhancing the methodological approach to Systematic Literature Reviews: A practical application on Topic Labeling research"

#Thesis/Weekly notes#