Specialist English: Assignment 6

Rebecca J. Stones
rebecca.stones82@nbjl.nankai.edu.cn

Date due: 13 November 2017

This sixth assignment (worth 5% of the final mark) looks at abstracts.

I'll scale the marks on this assignment according to $m \mapsto \min(m, 10)$ for Master's students and $m \mapsto \lceil m/1.3 \rceil$ for Ph.D. students.

My marking will be affected by (a) your English writing, (b) your LaTeX typesetting, (c) your mathematical presentation, and (d) your understanding of the underlying computer science. Basically, I will "peer review" your assignments.

The example abstracts (and snippets) are from published papers as quoted.

Problem 1 Consider the following abstract:

> Given a large volume of Web documents, we consider problem of finding the shortest keyword sequences for each of the documents such that a keyword sequence can be rendered to a given search engine, then the corresponding Web document can be identified and is ranked at the first place within the results. We call this system as an Inverse Search Engine (ISE). Whenever a shortest keyword sequence is found for a given Web document, the corresponding document can be returned as the first document by the given search engine. The resulting keyword sequence is search-engine dependent. The ISE therefore can be used as a tool to manage Web content in terms of the extracted shortest keyword sequences. In this way, a traditional keyword extraction process is constrained by the document ranking method adopted by a search engine. The significance is that the whole Web-searchable documents on the World Wide Web can then be partitioned according to their keyword phrases. This paper discusses the design and implementation of the proposed ISE. Four evaluation measures are proposed and are used to show the effectiveness and efficiency of our approach. The experiment results set up a test benchmark for further researches.
>
> — Chen et al., ADC, 2009.

1. Describe the most important problem with the overall structure of this abstract. [1 mark]
2. What's wrong with the word "researches" in the last sentence? [1 mark]
   (Misusing "researches" is a common error among Chinese authors.)

Problem 2 The following is an example of the demonstration part of the abstract.

> The proposed algorithm is evaluated on Ren-CECps, a Chinese blog emotion corpus. Experimental results show that the coarse-to-fine emotion classification algorithm improves the sentence-level emotion classification by 19.11% on the average precision metric, which outperforms the baseline methods.
>
> — Xu et al., CIKM (2012).

This snippet is both succinct and specific, but it does not directly say that the 19.11% improvement is vs. the baseline methods (otherwise it's an excellent example). Putting that aside, identify two (or more) things that the authors are specific about in this snippet. [2 marks]

(Continued on next page.)

Problem 3 Identify the major structural components of the following abstract (if present): introduction, solution, demonstration, and the implications. [2 marks]

> Classical collaborative filtering, and content-based filtering methods try to learn a static recommendation model given training data. These approaches are far from ideal in highly dynamic recommendation domains such as news recommendation and computational advertisement, where the set of items and users is very fluid. In this work, we investigate an adaptive clustering technique for content recommendation based on exploration-exploitation strategies in contextual multi-armed bandit settings. Our algorithm takes into account the collaborative effects that arise due to the interaction of the users with the items, by dynamically grouping users based on the items under consideration and, at the same time, grouping items based on the similarity of the clusterings induced over the users. The resulting algorithm thus takes advantage of preference patterns in the data in a way akin to collaborative filtering methods. We provide an empirical analysis on medium-size real-world datasets, showing scalability and increased prediction performance (as measured by click-through rate) over state-of-the-art methods for clustering bandits. We also provide a regret analysis within a standard linear stochastic noise setting.
>
> — Li, Karatzoglou, and Gentile, SIGIR (2016).

Problem 4 Consider this sentence from an abstract:

> A *supergraph query*, q, on a graph database D is to retrieve all graphs in D such that q is a supergraph of them.
>
> — Zhang et al., EDBT (2009).

The mathematical notation q and D are not used elsewhere in the abstract, and thus it is beneficial for the reader if we avoid introducing this notation. Rewrite this sentence without using mathematical notation. [2 marks]

Problem 5 Rephrase the following sentence to use the active voice (introducing the agent "we"). [2 marks]

> Adversarial learning is implemented as an interplay between two processes.
>
> — Wang et al., MM (2017).

Problem 6 The noun "work" in the following snippet is uncountable (i.e., we don't have "one work", "two works", etc.), but by writing "previous works", it's written as if it's countable. Explain how to correct this common error. [1 mark]

> Previous works on annotating domain entities from biomedical references suffer from several issues, such as a data flexibility problem, language dependency, and limitations with respect to word sense disambiguation.
>
> — Tian et al., WI-IAT (2013).

(Continued on next page.)

Problem 7 For the following four snippets from abstracts indicate, if necessary, how to change their tenses to match my recommendation in lectures: simple past tense when referring to other papers, simple present tense ordinarily, and (possibly) continuous present tense if it reads better. If no changes are necessary indicate how you come to that conclusion. [2 marks]

> In this paper, we have proposed an approach to enhancing the semantic interoperability of reuse repositories ...
>
> — Pan et al., SIGSOFT (2004).

> We conducted experiments on a real data set showing the superiority of our EDS distance measure.
>
> — Xie, SIGMOD/PODS (2014).

> We empirically evaluated our system using a large GPS dataset collected by 162 users over a period of 2.5 years in the real-world.
>
> — Zheng et al., WWW (2010).

> In 2008 Hölbl et al. proposed a password-based protocol for remote user authentication and password changing.
>
> — Yang et al., iiWAS (2009).

(*Note*: My recommendation is because: 1. it's simple, so it's difficult to get wrong, and easiest for the reader, 2. it's the most common, and 3. it's uncontroversial.)

P.S.: If you feel like using it, my color scheme is typeset as follows:

- Poor examples: `\begin{tcolorbox} ... \end{tcolorbox}`,
- Good examples: `\begin{tcolorbox}[colback=white,colframe=green!60] ...`,
- Other examples: `\begin{tcolorbox}[colback=blue!15,colframe=blue!50] ...`,
- introduction `\textcolor{red}{introduction}`,
- solution `\textcolor{blue}{solution}`,
- demonstration `\textcolor{green!50!black}{demonstration}`, and
- implications `\textcolor{purple!50!black}{implications}`.