Specialist English: Assignment 6 (solutions)

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

November 17, 2017

Here's my solutions; your solutions needn't be identical.

Problem 1 This is the 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.* The introduction is too long, and it overwhelms the other important material. (It wouldn't be as big of a problem if the other components were proportionally longer, but then the abstract as a whole might be too long.)

2. *What's wrong with the word "researches" in the last sentence?*

   I've discussed this with other native English speakers, and most (like me) think that using "researches" as the plural of "research" is simply wrong. The general attitude is that it shows lack of familiarity with English, and lack of familiarity with research.

   Most native English speakers are not even aware of the old-fashioned usage of "researches" as a noun (as I mentioned in the solutions to Assignment 5). One of my friends (who's a paralegal) mentioned that it's still infrequently used in law, which may explain why it's still in some dictionaries.

   On top of what I wrote in the solutions to Assignment 5:

   - The MacMillian dictionary[1] says "research" is uncountable; they write:

     Research is an uncountable noun, and so:

     - it is hardly ever used in the plural
     - it never comes after **a** or a number

[1] https://www.macmillandictionary.com/dictionary/british/research_1

✗ Her latest work confirms the findings of earlier ~~researches~~.

✓ Her latest work confirms the findings of earlier research.

✗ According to ~~one recent research~~, women's earnings are still 27% lower than men's.

✓ According to recent research, women's earnings are still 27% lower than men's.

- Here's three examples of IELTS instructors considering it wrong:
  - *IELTS Simon* writes[2]:

    Below are some common uncountable nouns. Do you know any others?
    - give advice, some advice, a piece of advice (~~an advice~~, ~~advices~~)
    - do research, carry out research (~~a research~~, ~~researches~~)
    - gain knowledge, share knowledge, increase knowledge (~~knowledges~~)
  - *5 Common Grammar Mistakes in IELTS Writing* writes[3]:

    If a noun is uncountable, you cannot use:
    - a plural verb: "~~Recent researches have shown~~." It should instead be: "Recent research has shown."
  - *Sample Feedback on IELTS Writing Task 2* writes[4]:

    "Research" is uncountable, so "researches" is wrong.

Problem 2 Regarding this demonstration component:

> 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).

... identify two (or more) things that the authors are specific about in this snippet.
These are the major things:

- The dataset is specified: "Ren-CECps, a Chinese blog emotion corpus".
- What kind of improvement is specified: "the sentence-level emotion classification".
- How improvement is measured is specified: "average precision metric".
- The amount of improvement is specified: "19.11%". (However, writing "19%" would have been just as informative; what difference does 0.1% make?)

On top of this:

- What they test is specified: "the coarse-to-fine emotion classification algorithm". However, this is what the whole paper is about, so we needn't be that specific, e.g., "the proposed algorithm" is fine.
- They also write "baseline methods", which is specific by stating that they compare to a baseline method, but unspecific about what the baseline is (and this is fairly obvious in any case).

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

[2] http://ielts-simon.com/ielts-help-and-english-pr/2017/08/ielts-grammar-uncountable-nouns.html
[3] https://magoosh.com/ielts/5-common-grammar-mistakes-ielts-writing/
[4] https://ielts.allearsenglish.com/sample-feedback-ielts-writing-task-2/

I was expecting this:

> 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).

The sentence beginning "The resulting algorithm ..." could alternatively be considered part of the demonstration (the boundary is blurry in this case, as it often is).

There is no implications component, i.e., we don't talk about the impact of the research outside of the paper (e.g., in what systems the proposed method might be used); this is not a significant loss in this case.

Problem 4 *Rewrite this sentence without using mathematical notation.*

> 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).

One thing that confused a lot of students was the use of "supergraph query" to refer to both (a) the process of querying the database, and (b) the query itself (i.e., the input query graph). Indeed, Zhang et al. conflated these two distinct concepts.

I was expecting something like:

> On a graph database, a supergraph query retrieves all graphs for which the input query graph is a supergraph.

Problem 5 *Rephrase the following sentence to use the active voice (introducing the agent "we").*

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

I was expecting something like:

> We implement adversarial learning as an interplay between two processes.

Here it is "adversarial learning" and not "the adversarial learning".

Problem 6 *The noun "work" in the following snippet is uncountable ... Explain how to correct this common error.*

> 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).

One way to fix this is to change "works" to "work", which means we need to also change "suffer" to "suffers".

> Previous **work** on annotating domain entities from biomedical references **suffers** from several issues, such as a data flexibility problem, language dependency, and limitations with respect to word sense disambiguation.

Another way is to change "works" to e.g. "papers" or "studies".

> Previous **studies** 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.

It's arguably technically correct to write "some previous works", "all previous works", or "many previous works", but (a) this changes the meaning (the original sentence leaves this ambiguous), and (b) we should opt for uncomplicated grammar (we don't write so the reader has to think about grammar—we keep it simple, and write so the reader thinks about the content of the paper).

Problem 7 *For the following four snippets from abstracts indicate, if necessary, how to change their tenses to match my recommendation in lectures ...* I highlight the verbs below:

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

Here, "propose" (simple present tense) is my recommendation over "have proposed", just because it's simpler. The verb "enhancing" is in continuous present tense; I would say "enhance" reads slightly better (but both are okay). So we change it to:

> In this paper, we **propose** an approach to **enhance** the semantic interoperability of reuse repositories ...

The next example is:

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

Here, "conduct" (simple present tense) is my recommendation over "conducted". In this case, it makes it clear that the experiments are in the paper. The verb "showing" is in continuous present tense; I feel "which shows" (with a comma added) breaks apart the sentence better, but it's not a big deal:

> We **conduct** experiments on a real data set, **which shows** the superiority of ~~our~~ the proposed EDS distance measure.

I also recommend "the proposed" over "our" (think: it's ours, not yours).
The next example is:

> 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).

Here, "evaluate" (simple present tense) is my recommendation over "evaluated" (simple past tense). Also, "collected" (simple past tense) is appropriate since it refers to prior work.

> We empirically **evaluate** our system using a large GPS dataset **collected** by 162 users over a period of 2.5 years in the real world.

Also we should write "real world" (noun), and not "real-world" (compound adjective).
The last example is:

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

The tense is fine, but it could benefit from a comma after "2008".

> In 2008, Hölbl et al. proposed ...

In fact, 2008 is the publication date, and it's not necessarily the year they made their proposal. So it's better to write:

> Hölbl et al. (2008) proposed ...