Specialist English: Assignment 2 (solutions)

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

October 28, 2018

Here's my solutions; your solutions needn't be identical. There may be problems with the sample snippets I haven't listed.

Problem 1 The snippet is

> feature selection process. This weighting scheme follows the vector space model in text retrieval. Specifically the *tf* counts for the concept popularity and the *idf* counts for the concept specificity.
>
> $$\textit{c-tf-idf(c,d)} = \textit{freq(c,d)} \log \frac{|D|}{\textit{frec(c)}}$$
>
> Based on the *c-tf-idf* measurement, we can select concepts from relevance feedback with the intuition that concepts with higher
>
> – Wang et al. *Learning Structured Concept-Segments for Interactive Video Retrieval*, 2008.

Aside from "freq" being misspelled "frec", the main problem is that the equation is just floating, and is not part of a sentence. The equation is not connected to anything, so it's hard to tell how it's related to the surrounding material.

It is also noteworthy that the font used in the equation does not match the the font in the main text (compare f with f). Also the vertical lines surrounding $|D|$ are not vertical. Moreover, writing "freq" in italics is inconsistent with not writing "log" in italics: the equation should be typeset in LaTeX

$$\text{c-tf-idf}(c,d) = \text{freq}(c,d) \log_2 \frac{|D|}{\text{freq}(c)}$$

which is generated from

```
\[\text{c-tf-idf}(c,d) = \text{freq}(c,d) \log_2 \frac{|D|}{\text{freq}(c)}\]
```

In this case, `\mathrm` is not a good alternative to `\text`, e.g. `$\mathrm{c-tf-idf}(c,d)$` compiles to $\mathrm{c-tf-idf}(c,d)$ which is not appropriate.

Here, I'm guessing the base of the logarithm is base 2, i.e., $\log_2$ and not $\log_{10}$ or $\log_e$, or something else. It requires 1 extra character to clear up confusion—it's worth it! (The reader should not have to guess!)

Problem 2 The snippet is

> It can be further formulated into the following problem:
>
> $$\begin{aligned} &\min_{\mathbf{v}} \mathbf{v}^T\mathbf{v}+\mathbf{v}^T\mathbf{h} \\ &\text{s.t.} v_1 + v_2 = 1, 0 \le v_1, v_2 \le 1 \end{aligned} \tag{8}$$
>
> – Zhang et al. *Direction based Graphical Model for Face Sketch Synthesis*, 2016.

The second line in equation (8) is difficult to parse for the following reasons:

- The abbreviation "s.t." can be avoided by replacing it with "such that" or "subject to".
- There are mathematical expressions separated by commas: it's hard to determine where the boundaries are.
- It's ambiguous whether $0 \leq v_1, v_2 \leq 1$ means "$0 \leq v_1$ and $v_2 \leq 1$" or "$0 \leq v_1 \leq 1$ and $0 \leq v_2 \leq 1$", or something else.

As a result, I'm not sure what this actually means—it should be simple to work out what it means, but it's not. Elsewhere in the paper, we see $\mathbf{v}$, $\mathbf{u}$, and $\mathbf{h}$ only have two elements, thus we needn't use vector notation. Clearly, this is not written with the reader in mind.

Some students preferred writing

$$\min_{\mathbf{v}}(\mathbf{v}^T\mathbf{v} + \mathbf{v}^T\mathbf{h})$$

to clear up the ambiguity with $\min_{\mathbf{v}}(\mathbf{v}^T\mathbf{v}) + \mathbf{v}^T\mathbf{h}$. While it was not the anticipated answer, this seems reasonable to me, and was awarded marks.

Problem 3 The snippet is

> the core power capacity reference of the chip P_{0_c}. The power capacity of core j therefore is
>
> $$P_{core\_j} = r_j P_{0\_c} = r_j \min(P_{0\_i}), 1 <= i <= N. \quad (14)$$
>
> – Chen et al. *TSocket: Thermal Sustainable Power Budgeting*, 2016.

There are several problems with this snippet:

- There are mathematical expressions separated by a comma; these should be separated so the boundaries are made clear.
- The notation <= is incorrect, and should be replaced with $\leq$ typeset `\leq` (or `\le`), or alternatively $\leqslant$, typeset `\leslant`.
- It is not explicitly stated what the minimum is over; I guess it's over all $i \in \{1, \ldots, N\}$, and this should be incorporated underneath the "min".
- The equation number (14) is not used in the paper, and is not needed.

After changing to more suitable notation, we might write:

> The power capacity of core j is therefore
>
> $$\begin{aligned} \mathrm{P}_j^{(\mathrm{core})} &= r_j \mathrm{P}_c^{(0)} \\ &= r_j \min_{i \in \{1,\ldots,N\}} P_i^{(0)}. \end{aligned}$$

I prefer "P" to "P", but in this case, it's a matter of personal preference. I also prefer two lines (using `\begin{align*} ... \end{align*}`) to emphasize how there are two equations. I feel "is therefore" sounds better than "therefore is" (although I believe they are both correct, understandable, and identical in meaning—this is a quibble).

Problem 4 The snippet is

> Specifically, we perform a linear interpolation between two probabilities of the NMT generations and SMT word recommendations, to update the prediction probability:
>
> $$p(y_t|\mathbf{y}_{<\mathbf{t}}, \mathbf{x}) = (1-\alpha_t)p_{nmt}(y_t|\mathbf{y}_{<\mathbf{t}}, \mathbf{x}) + \alpha_t p_{smt}(y_t|\mathbf{y}_{<\mathbf{t}}, \mathbf{x}) \quad (15)$$
>
> Note that $p_{smt}(y_t|\mathbf{y}_{<\mathbf{t}}, \mathbf{x}) = 0$ for $y_t \notin V_t^{smt}$, as the SMT word recommendation vocabulary at decodint step t V^{smt} is N , , ,
>
> – Wang et al. *Incorporating Statistical Machine Translation Word Knowledge Into Neural Machine Translation*, 2018.

Here NMT and SMT are acronyms, so it's not appropriate to switch them to lowercase. It's also not appropriate to write them in math mode, e.g. $nmt = n \times m \times t$. In this particular case, the spacing between n, m, and t in math mode is particularly large.

The notation Pr(...) is typically used to denote a probability function. Thus, it's better to choose notation such as

$$\Pr_{\text{NMT}}(y_t \mid \mathbf{y}_{<\mathbf{t}}, \mathbf{x}) \quad \text{and} \quad \Pr_{\text{SMT}}(y_t \mid \mathbf{y}_{<\mathbf{t}}, \mathbf{x})$$

or possibly

$$\mathrm{P}_{\text{NMT}}(y_t \mid \mathbf{y}_{<\mathbf{t}}, \mathbf{x}) \quad \text{and} \quad \mathrm{P}_{\text{SMT}}(y_t \mid \mathbf{y}_{<\mathbf{t}}, \mathbf{x})$$

depending on your preference.

Problem 5 The snippet is

> *Definition 1:* A *linearization* of a concurrent history ρ is a sequential history ρ' such that (1) $\rho' \sim \rho$, meaning that they share the same set of events, and (2) $\forall e_i, e_j : e_i <_\rho e_j$ implies $e_i <_{\rho'} e_j$. In other words, the non-overlapping method calls in ρ retain their execution order in ρ', whereas the overlapping method calls may take effect in any order.
>
> – Zhang et al. *Round-Up: Runtime Checking Quasi Linearizability of Concurrent Data Structures*, 2013.

Rewriting the content after (2):

> (2) for any e_i and e_j, if $e_i <_p e_j$ then $e_i <_{p'} e_j$.

Or alternatively, we might write:

> (2) if $e_i <_p e_j$ for some e_i and e_j, then $e_i <_{p'} e_j$.

More succinct is either of the following:

> (2) if $e_i <_p e_j$ then $e_i <_{p'} e_j$.
> (2) $e_i <_{p'} e_j$ whenever $e_i <_p e_j$.

Any of the above is okay, however, the following is problematic:

> (2) for all e_i and e_j, if $e_i <_p e_j$ then $e_i <_{p'} e_j$.

This would likely be misinterpreted as:

> (2) if $e_i <_p e_j$ for all e_i and e_j, then $e_i <_{p'} e_j$.

These have a different meaning than what is intended by the authors. I was a bit lenient with marks here: the difference is subtle.

The authors' use of blackboard shortcuts leads to this problem: the reader has to (a) read the sentence, then (b) deduce which meaning is correct and which is incorrect—this takes effort and makes the sentence difficult to read. It is much better to write the correct meaning unambiguously, so the correct meaning directly comes from reading the sentence.

Several students wrote the following (or something similar):

> (2) for all e_i and e_j, $e_i <_p e_j$ implies $e_i <_{p'} e_j$.

This is not a good choice as it's far too easy to think that "e_j, $e_i <_p e_j$" belongs together. In fact, this confusion is exactly why we avoid separating mathematical expressions with commas.

Problem 6 The snippet is

> during normal execution. Given n threads each sending r requests, we calculate the availability of the protected system as:
>
> $$\mathcal{A} = 1 - \bar{t}/(r * n) \quad (1)$$
>
> where $\bar{t}$ = Average timeout number,
> r = Requests per thread,
> n = Thread number.
>
> – He et al. *Reverse Replication of Virtual Machines (rRVM) for Low Latency and High Availability Services*, 2016.

I would change it to

> Given n threads each sending r requests, the availability $\mathcal{A}$ of the protected system is defined as
>
> $$\mathcal{A} = 1 - \frac{\bar{t}}{rn}$$
>
> where $\bar{t}$ is the average number of timeouts per thread.

This fixes the following:

- Both n and r are defined twice in the original snippet; we only do this once.
- The n and r at the start of the sentence now correctly use math mode.
- We remove the equation number, which is not used in the paper.
- We correct "Average timeout number" to "average number of timeouts".
- We no longer have the problem of "X, Y, and Z" being written "X, Y, Z". (This is called asyndeton[1].)
- There is no indent before "where".
- We don't write "is" using an equals sign ("=").

In addition:

- By adding $\mathcal{A}$ after the word "availability", we unobtrusively emphasize how availability is denoted $\mathcal{A}$.
- We change the wording "we calculate" (which suggests a method of computation) to "is defined as". (In fact, the author choosing to write "calulate" instead of "defined" indicates that it's not a definition; perhaps an computational approximation instead.)

[1] `https://en.wikipedia.org/wiki/Asyndeton`

- I use `\overline{t}` to give $\overline{t}$ (instead of `\bar{t}` which gives $\bar{t}$). In this case, it's barely noticeable, but generally `\overline` looks better than `\bar`.
- Using `\frac{...}{...}` eliminates the need for brackets around rn, reducing the amount of notation.