Specialist English: Assignment 3 (solutions)

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

November 3, 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 relevant snippets are the following:

> development of parallel algorithm simple and convenient. MapReduce[5]is a parallel programming model proposed by Google, which is used for large scale data sets. Its operating environment consists of the client, the main node and the working
>
> – He and Zhang, *A structure of intelligent grain network ...*, Proc. ICC, 2017.

In this snippet the space before and after "[5]" is incorrectly omitted. Correct is:

MapReduce [5] is a parallel programming model ...

> **1. INTRODUCTION**
>
> Crowd-Based Software Engineering(CBSE) allows anyone to participate in software development tasks including documentation, design, coding and testing. CBSE is widely used
>
> – He and Zhang, *Recommending Relevant Projects via User Behaviour ...*, Proc. CrowdSoft, 2014.

In this snippet the space before "(CBSE)" is incorrectly omitted. Correct is:

Crowd-Based Software Engineering (CBSE) allows anyone ...

Pedantically speaking "Engineering(CBSE)" means Engineering is a function of CBSE (like $f(x)$).

> **HLBPR:A Hybrid Local Bayesian Personal Ranking Method**
>
> – Chen et al., *HLBPR:A Hybrid Local Bayesian ...*, Proc. WWW, 2016.

In this snippet the space after the colon ":" is incorrectly omitted. Correct is:

HLBPR: A Hybrid ...

Problem 2 The snippet is:

> **Zhang et al. [33]:** This method constructs three local graphs for a candidate set based on coauthors and document similarity. A graph embedding is learned for each candidate set by sampling triplets
>
> ⋮
>
> [33] Baichuan Zhang and Mohammad Al Hasan. 2017. Name disambiguation in anonymized graphs using network embedding. In *CIKM'17*. 1239–1248.
>
> – Zhang et al., *Name Disambiguation in AMiner ...*, Proc. KDD, 2018.

The abbreviation "et al." means "and others" (plural), while Zhang et al. have used it to mean "and other" (since there is only one additional author: Mohammad Al Hasan). Moreover, it is pointless replacing a single author's name with "et al." since it saves virtually no space, and just makes it more difficult to read. Correct is:

Zhang and Al Hasan [33]: This method ...

I've seen arguments that it should be typeset:

Zhang and al-Hasan [33]: This method ...

There's some debate about how to typeset Arabic names.

Problem 3 The snippet is:

> encoded using 1-of-N and 1-of-M coding separately. In the hidden layer, we use the identity function as the activation function:
>
> $$\mathbf{h}_u = f(\mathbf{x}_u\mathbf{W}_u) = \mathbf{x}_u\mathbf{W}_u \quad , \quad \mathbf{h}_v = f(\mathbf{x}_v\mathbf{W}_v) = \mathbf{x}_v\mathbf{W}_v$$
>
> and the output unit is the dot product of two hidden layers:
>
> $$y = \mathbf{h}_u \cdot \mathbf{h}_v = \sum_i h_{ui} \times h_{vi} \qquad (3)$$
>
> The learning objective is to minimize reconstruction error
>
> – Lian et al., *CCCFNet: A Content-Boosted ...*, Proc. WWW, 2017.

It has the following punctuation errors:

- There is a space before the comma after the equation beginning "$\mathbf{h}_u = ...$": there should be no space. However, it would be better to write "and" instead of a comma, and I feel the two equations would appear better on separate lines.
- There is no full stop at the end of the sentence, i.e., at the end of equation (3).

I would also delete the colons: colons clash when we have more than one in a single sentence.

Problem 4 The snippet is:

> lized fixation coordinates. Considering the mean minimum time to acquire the full meaning of a word is 151ms [10], we filter out those clustered fixation coordinates that last less than 151 ms. After we get the stabilized fixation coordi-
>
> – Chen et al., *A Real-Time Eye Tracking Based ...*, Proc. CIKM, 2015.

In this snippet "151ms" (with a space) and "151 ms" (without a space) are inconsistent. Either way is okay, but we need to be consistent throughout the paper.

Problem 5 The snippet is:

> We design the experiment with 4 goals:(1) To evaluate how our proposed algorithm performs compared with other base-line algorithms;(2)To test how different features we considered affect the recommendation performance;(3) To examine how different ranking functions affect the results;(4) To consider how new challenges like the recommendation length restriction and recommendation overload affect the performance of our algorithm.
>
> – Zhang et al., *Whom to Mention ...*, Proc. WWW, 2013.

The punctuation-related errors with this snippet are:

- There is no space between ":" and "(1)".
- There is no space between ";" and "(2)".
- There is no space after "(2)".
- There is no space between ";" and "(3)".
- There is no space between ";" and "(4)", and there should also be an "and" before "(4)", since it uses the structure "(1) ...; (2) ...; (3) ...; and (4) ...".

Other recommendations:

- I prefer "four" to "4". I aim to use numbers (like "4") when they're parameters, but in this case, it's not a parameter.
- I prefer lettered enumerations ("(a)", "(b)", etc.) over numbered enumerations ("(1)", "(2)", etc.) since "(1)" can be mistaken for "equation (1)".
- I prefer "the proposed algorithm", rather than "our proposed algorithm", as it's more impartial.
- I prefer starting each enumerated item with a lowercase letter, so it reads more naturally as a sentence. In particular "... and (4) To ..." is problematic as the capitalization indicates the start of a new sentence, but we are still within a sentence.
- The word "other" in "other base-line algorithms" is inaccurate: it implies the proposed algorithm is a base-line algorithm.
- The phrase "different features" should be "the different features", but I recommend never using "different" to mean "various" (since "different" has multiple meanings, and the reader has to put in effort to determine which meaning is intended).
- We should use present tense, i.e., not "considered". (In fact, the authors use both past and present tense the the snippet.)
- I would add commas to make the fourth item more readable. (Thus, we should use semicolons to separate the list items.)

Thus, I would change it to:

> We design the experiment with four goals: (a) to evaluate how the proposed algorithm performs compared with base-line algorithms; (b) to test how the various features we consider affect the recommendation performance; (c) to examine how different ranking functions affect the results; and (d) to consider how new challenges, like the recommendation length restriction and recommendation overload, affect the performance of the proposed algorithm.

I would also be more precise than "the results" and "the performance": these terms are vague, and it's better to be more specific.

Problem 6 The snippet is:

Algorithm 3 Master(int K)

```
1:  MessageSet S = null
2:  int layer_no=0
3:  while TRUE do
4:      Message m = waitForNextMessage()
5:      S.add(m)
6:      if S.size==K then
7:          var, mean = getStatistic(S)
8:          for i=0 to K do
9:              Matrix V = doNormalization(var, mean, S[i])
10:             SendToGPUWorker(V, S[i])
11:         layer_no++
12:         S.removeAll()
```

– Wu et al., *A New Approach to Compute CNNs …*, Proc. CIKM, 2017.

The following is suitable for code, but suboptimal for pseudocode (where we communicate with a human):

- Writing "int K", and "int $layer\_no$=0" is suitable when programming for declaring an integer as defined by the compiler (not a mathematical integer). It also has the problem of being equal to $l \times a \times y \times e \times r \cdots$.
- Writing "==" is suitable when programming, for distinguishing between equality and assignment. When writing pseudocode, we're better off using $\leftarrow$ (typeset `$\gets$`) or :=. In the case of Line 6, we can simply say

  > if the size of S is equal to K then ...

  or

  > if $|S| = K$ then ...

- Writing "++" is suitable when programming, increasing a variable by 1. When writing pseudocode, we're better off using $\leftarrow$ (typeset `$\gets$`) or :=. But better is writing in words, e.g.:

  > increase $layer\_no$ by 1

  (putting aside how $layer\_no$ is not suitable notation).

- Writing "S.removeAll()", "waitForNextMessage(), etc., is code: we should write in English using words and sentences, not code!
- Writing "for $i = 0$ to K" is more suitable for programming—mathematical writing tends to be static: if we say "$i = 0$", it means $i = 0$ is true (not "sometimes true" or "initially true"), we should not change it later on. Better is

  > for i from 0 to K

There are other examples; I don't list them all.

What's important here:

- The default is writing sentences, using words. We use notation to simplify sentences.
- It's not always bad to use code, but this is "breaking the rules". We need to learn when to "break the rules".
- One problem with writing "pseudocode" as "code" is that the reader might be familiar with a different compiler to you, and therefore expect a different syntax.