Specialist English: Assignment 7 (solutions)

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

December 15, 2017

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

Problem 1 *Identify four (or more) distinct structural elements used by Wang et al. in Sections 1 and 2, and briefly explain how they improve the paper's readability.*

- Sections and subsections. These help the reader broadly identify what material is where.
- Bold in the Introduction for "Regularity and Conformity" and "Sparsity and Homogeneity". This helps the reader identify the two major challenges in the paper, and where to find their descriptions.
- Bold in the Introduction for the acronym "**R**egularity and **C**onformity, and employing **H**eterogeneous mobility data for location prediction". This helps the reader identify how the acronym RCH is derived (although, this is not particularly important).
- Italics. The authors use italics to highlight the first usage of "regularity" and "conformity", making it easy to identify where these concepts are introduced.
- Bullet points. The bullet points itemize the major contributions, thereby allowing the reader to easily identify these contributions, and also understand these points one at a time.
- Tables. Table 1 provides a side-by-side comparison of many methods, which is far more structured (and therefore far more understandable) than the same information provided in sentences. Table 2 gives a table of notation, which provides a useful aid for the reader when they need to recall notation later in the paper; there's far too much notation for the reader to memorize.
- Footnotes. The authors use footnotes in order to include URLs without distracting the reader from the primary material.

The authors also use paragraphs, and it was fine to pick any one and describe how its structure assists the reader. For example, the first paragraph is comprised of two sentences: the first sentence immediately identifies the topic (location-aware services) and its widespread use; the last sentence gives a broad (although somewhat vague) description of what the paper is about: understanding user mobility. This helps the reader understand what the paper is about and how it fits into a broader context.

There's probably many other things that are unlisted.

Problem 2 *Why have I described this sentence as illogical?*

> To bridge the gap between actual prediction performance and [the] theoretical limit, many challenges still remain to be addressed: ...

In order "to bridge a gap", we need to actually do something. Compare this to: "to drive around China, one challenge is not having a car", or "to cook a pizza, we haven't yet bought the

cheese". More logical would be "To bridge the gap between actual prediction performance and [the] theoretical limit, many challenges still need to be addressed: ...".

Many students commented that achieving the theory is impossible. While correct, this is an underlying assumption throughout science. "To bridge the gap ..." is metaphorical, so it might be better to say "To reduce the gap ..." or "To minimize the disparity ..." or something along those lines, but I wouldn't consider it illogical.

Problem 3 *Why have I underlined these and described them as "too far apart"?*

> In this paper, we tackle the above challenges by proposing a hybrid model called **RCH**, combining both **R**egularity and **C**onformity, and employing **H**eterogeneous mobility data for location prediction.

The problem here is that the phrase "for location prediction" describes the "model". It is difficult for the reader to make this connection when they're spaced so far apart.

Problem 4 *There's something important missing from Section 2.1. (I mention it among my annotations.)*

1. *What is missing from Section 2.1?*

   Section 2.1 (Predictability of Human Mobility) describes related work, but fails to describe how the work is related to the present paper.

2. *Where does it occur in Section 2.2 (before the start of Section 2.2.1), and Sections 2.2.1 and 2.2.2?*

   In Section 2.2, Table 1 compares location prediction models, including the proposed model, thus showing how the references are related to the present paper.

   In Section 2.2.1, the relationship is described in the text beginning with "Instead, our hybrid approach ...".

   In Section 2.2.2, the relationship is described in the text beginning with "In contrast to the above collaborative filtering approaches ...".

Problem 5 *Identify eight (or more) distinct problems with these two sections, excluding those in Problems 1 to 4. Among these, please include examples the following:*

- *A grammar error.*
  - Quotation marks should not be used for "searching for the closest subway station", since it's not a quote.
  - The sentence containing "accomplish tasks without their querying" is ungrammatical; it should just be "accomplish tasks without querying".
  - "Song et al. [33] show ..." would be better as past tense "showed".
  - The snippet "is the so-called social conformity" should be "is called social conformity"; here "so-called" does not have the same meaning as "called".
  - The word "touched" (which suggests physical touching) would be better as "touched on".
  - Writing both "touched" and "to a certain extent" is unnecessary: if it wasn't only "to a certain extent", it would be incorrect to say "touched".
  - In "the main endeavor of these methods still focuses on a single factor", the word "endeavor" clashes with the word "focuses"—is it a focus, or an endeavor? We should simply write "these methods still focus on a single factor".
  - The authors write "while mobility data captured by low-energy sensing technologies is typically sparse", but "data" is plural, so it should be "are" and not "is".
  - The "that" in the following sentence is incorrect, and should be deleted:

> Song et al. [33] demonstrated that a 93% potential predictability of mobility patterns of mobile phone users.

  The sentence is incomplete: correct would be "[someone] demonstrated that [something] results in [something else]."
- As used, the noun "uniqueness" is uncountable, thus it's not "the mobility uniqueness" but just "mobility uniqueness".
- The phrase "periodic returning" should be "periodically returning" since "periodically" is used as an adverb to describe the verb "returning".
- The phrase "high frequency datasets" should use a compound adjective "high-frequency datasets", otherwise we're referring to "frequency datasets" (i.e., datasets which contain frequencies) which happen to be "high".
- "An approach based on non-linear time series is applied for mobility prediction in [27] ..." should use "was" (past tense) instead of "is".
- In "which focused on predicting most important places", we should say "the most important places".
- The phrase "the best match path" should be "the best matched path".
- In "for next place prediction", we should use the compound adjective "next-place" to describe the type of "prediction".
- The word "the" is missing in "considering location histories of users' friends".
- The phrase "has evolved as a" does not mean "has evolved to become a" (which the authors intend).
- Missing "the" in "and geographical influence of locations".
- It should be "recommendation" in "... has also been adopted for location recommendations ..." since we're talking about the process of location recommendation (and not the recommendations themselves).
- The use of "first employed" in the following sentence is ambiguous (was Lian et al. the first to use this approach, or was it the first step in their approach?).

  > Lian et al. [19] first employed collaborative filtering approaches for location prediction.
- The following sentence is grammatically incorrect and too wordy; it needs rephrasing.

  > Users' location visits are separated to explorations of novel or regular places based on a binary classification.

  It could say "User location visits are classified as either novel or regular."
- It looks like "confronts" is the wrong word in "confronts the risk of two-layer errors"; it should be "incurs" or "encounters".
- The phrase "learns spatial influence on venues" is grammatically incorrect; it's unclear how to correct this without studying the method in depth.
- "Notation" is uncountable, and therefore it should say "notation" in Table 2's caption.
- In Table 2, "incoming" and "outgoing" are words, and we don't need to write "incoming" and "out-going".
- In Table 2, "of time slot t" should be "at time t".
- In Table 2, "submatrix" is a word, and we don't need to write "sub-matrix".

- *A spelling error.*
  - "Combining" is misspelled "combing".
  - In Table 2, "spacial" is misspelled "sptial".
- *A capitalization problem.*
  - In "learned through a Gravity model", we should either use a lowercase G, or it should be "the Gravity Model" if there's a unique named model.

  - On page 1275, the authors write "WiFi", while on page 1276 the authors write "wifi", and also "Wifi" in Table 1; this is inconsistent.
  - The capitalization of "a supervised learning model", "Dynamic Bayesian Network model", and "hidden Markov model" is inconsistent. (I mistakenly flagged "Gaussian Mixture model" in my annotations; "Gaussian" should be capitalized since it's named after Gauss.)
  - We shouldn't write "the conformity theory"; it should be "Conformity Theory" since it's a proper noun and uncountable.
  - The capitalization of "Notation", "Size", and "Description" in Table 2 is inconsistent with "methods", "target", and "feature" in Table 1.
  - The capitalization Table 1's caption is different to the capitalization Table 2's caption.

- *A punctuation problem.*
  - The slash in "advertisers/marketers" is unnecessary; the straightforward word "and" works fine.
  - The comma after "[19]" in "... or a constraint [19], Thus, the interdependency ..." should be a full stop.
  - There should be no space before the full stop in "... a unified prediction model .".
  - The authors inconsistently use the Oxford comma, e.g. it is used in "shopping malls, gyms, and restaurants", but not used in "is regular, predictable and unique".
  - In Table 1, there is no space between "M5Tree" and "[25]".
  - In Table 1, there is no space between "RCH" and "(Our Model)".
  - In Table 2, items should be separated using a semicolon, e.g., "grid set; d is a grid in C with length I". Here, a comma does not make sense.
  - The abbreviation "w.r.t." in Table 2 would be better written "with respect to" (although there is a space consideration).

- *A problem related to LaTeX.*
  - Many references are out of order, e.g. [28, 4, 19]. This can be quickly fixed with `\usepackage{cite}`.
  - The sentences exceed the margins multiple times, e.g. "... is often energy-intensive ...".

- *An example where a complicated (or colloquial, or over-the-top) word is used in place of a simple word.*
  - The word "overwhelming" in the first sentence in the Introduction is over-the-top. A straightforward word like "great" would suffice.
  - The word "exploits" has some inherent sneakiness. We wouldn't say "we exploit a saw to cut wood", but instead simply say "we use a saw to cut wood". Likewise, we shouldn't write "which exploits the interdependent patterns" but "which utilizes the interdependent patterns".
  - The word "marvelous" in "provides marvelous potential" feels over-the-top and emotionally involved. Something simple like "great" or "enormous" would suffice.
  - The word "haunts" is colloquial (although this might be different in US English). Instead of "highly visited haunts", more formal is "highly frequented locations".
  - The word "excavates" in "our hybrid approach excavates" seems peculiar. The straightforward word "utilizes" suffices.
  - The word "excavate" in "excavate novel mobility patterns" could be replaced by the simple "identify" (although "excavate" is arguably better in a data mining context, since it plays on the word "mining").
  - It's unclear what "the antecedent division of mobility types" means, due to the word "antecedent".

On top of these:

- The word "appealing" in "appealing proactive experiences" feels inappropriate.
- Footnote 3 does not describe stampede prevention, as one would expect from the sentence it's referred from.
- The word "meanwhile" is used to mean something other than "at the same time".
- Writing "auxiliary information" reads more naturally than "side information".
- The authors use "Besides", which has multiple meanings and the process of determining which one is correct slows down the reader.
- Saying "with the help of" is clunky; better is "by utilizing".
- Writing "users' hub visit patterns" is clunky, and some might argue that "user's hub visit patterns" is the correct apostrophe usage. In this case, the word "users" is unnecessary since it's unambiguous who the "hub visit patterns" belong to. (Likewise for "users' visited venues".)
- In the phrase "... spatial influence to users' visited venues ...", it is not clear what is influencing the "users' visited venues".
- "Our main contributions are summarized as follows" is passive; better is "We summarize the main contributions as follows".
- Describing specific dataset details and the experimental results is not appropriate for the introduction.
- The acronym LBSN is not defined.
- In Table 1, it is inconsistent to label one column "methods" (plural) while also using the labels "target" (singular) and "feature" (singular).
- "An approach based on non-linear time series ..." would be more succinctly rephrased "A non-linear time-series approach ...".
- In "... and successfully inferred ...", we don't need the word "successfully", since "unsuccessfully inferred" is self-contradictory.
- In "... inferred unknown alighting/boarding stops ..." we don't need the word "unknown", since we don't need to infer something that is known. We can also use the simple word "or" instead of the slash.
- The word "considering" in "considering [the] location histories of users' friends" seems like the wrong choice of word; it should be "incorporating" or "which accounts for". (This occurs four times.)
- Table 2 is not referenced in the main text.
- The notation $\mathbf{H}(t)$ appears in two different ways in Table 2. (I'm not sure why.)
- The use of the wildcard asterisk * is poorly formatted in $^*\in \mathcal{P}$, instead of "$* \in \mathcal{P}$".
- It's unclear what "sub-matrix of group g" means.