Specialist English: Assignment 8

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

Date due: 11 December 2017 (no class December 4)

In this eighth assignment (worth 5% of the final mark), we will look at presenting solutions (i.e., what we've done) in a computer science paper.

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.

Problem 1

1. Rephrase the following sentence to remove "As shown in Figure 1(a) ...", which I recommend to avoid as it's difficult for the reader to parse. [1 mark]

> As shown in Figure 1(a), the baseline CMP system consists of multiple cores, where the L1 caches are private to each core and the lower level caches are shared by the cores.
>
> — Zhao, Xu, and Xie, *Bandwidth-Aware Reconfigurable Cache Design with Hybrid Memory Technologies*, ICCAD, 2011.

2. Rewrite the following snippet to eliminate "evidences" and "besides". [1 mark]

> A comparative summary should contain as many comparative evidences as possible. Besides, it should convey important information in the original document.
>
> — Huang, Wan, and Xiao, *Comparative News Summarization Using Linear Programming*, HLT, 2011.

3. While the following sentence is grammatically correct, "apart from" has multiple commonly used meanings, which slows down the reader who has to identify which meaning the authors intend. Replace "apart from" with a more direct synonym. [1 mark]

> In this paper, apart from horizontal handoff and vertical handoff, terminal handoff is also considered, which is performed between two different terminals.
>
> — Ma et al., *Research on Load Balancing Mechanism ...*, ICHIT, 2009.

(Continued on next page.)

Problem 2 In pseudocode, as with most things in a paper, we use words and sentences. We strive to be consistent in our writing (including capitalization, punctuation, notation, etc.).

When presenting pseudocode, we should minimize notation. This includes eliminating unused notation, replacing notation with words whenever possible, and replacing infrequently used notation by what it's defined as. If notation is used, we should choose sensible and simple notation, and add worded descriptions "the point p" and so on.

Algorithm 1 ELIMINATE-POINTS (m)

Input: $p_1, p_2, ..., p_{n'}$ (in order) where n' is the number of points in the stream.
Output: Skyline points S'

Let $x = 24m$.
Pass 1: For $j = 1, 2, ..., x$, let p'_j be a point picked uniformly at random from the stream. Let S be the set of such points.
Pass 2:
for $i = 1..n'$ **do**
  For any p'_j, if p_i dominates p'_j then $p'_j := p_i$
end for
Let $S' = \{p'_1, p'_2, ...p'_x\}$.
Pass 3: Delete from stream all points in S' and all points dominated by any point in S'.
return S'

— Sarma et al., *Randomized Multi-pass Streaming Skyline Algorithms*, VLDB, 2009.

Rewrite the above pseudocode algorithm to optimize its presentation. Explain at least four ways you've made the pseudocode more readable, more consistent, or have otherwise improved the pseudocode. [4 marks]

You will need to know the definitions of m and n below:

4. ALGORITHMS

In the rest of this paper, we use n for the number of points, m for the number of skyline points, and d for the number of dimensions of each point. We measure the performance of algorithms in terms of random I/O's, sequential

— Sarma et al., *Randomized Multi-pass Streaming Skyline Algorithms*, VLDB, 2009.

(Continued on next page.)

Problem 3 The following visualization is decent, but it has many minor problems which together give a poor ("it could be better") impression.

It contains inconsistencies, minor grammar errors, and inconsistent capitalization. The font is too small, the font is different to the main text, and the text is part of the image. The image pixelates if we zoom in. The icon for "legitimate campaign" is also used to mean "account". The icons have colored backgrounds. The lines underneath "Tweets" are not horizontal. The legend in the top-left is the first thing the reader sees, but the information it contains is only needed in the last step.

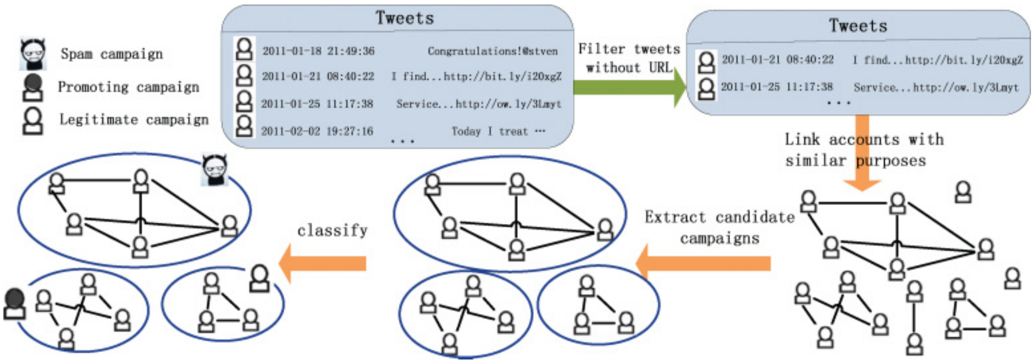

Fig. 4. The work flow of the proposed detection framework.

— Zhang et al., *Detecting Spam and Promoting Campaigns in Twitter*, ACM T. Web, 2016.

The task here is to redraw this figure to make it look polished and worthy of publication. The purpose is not to make it look identical, but to fix the problems and make it visually more appealing. [4 marks]

I recommend using the tikz LaTeX package (`\usepackage{tikz}`) to redraw this figure. Icons, if you choose to use them, are available from the public-domain repository `openclipart.org` (i.e., they're free to use without acknowledgements). Other icons are at `commons.wikimedia.org` but they're not all in the public domain (which means they need attribution).

(Continued on next page.)

Problem 4 The toy example on the right is not self contained. This forces the reader to refer to the main text to understand its content. Moreover, despite its caption saying "an example of destination prediction", no example is given.

Explain how to edit this figure and/or caption to make it self contained, and to actually show an example of destination prediction.

You don't need to redraw the figure, just give a description. (However, it's fine to redraw the figure if you prefer.)

[2 marks]

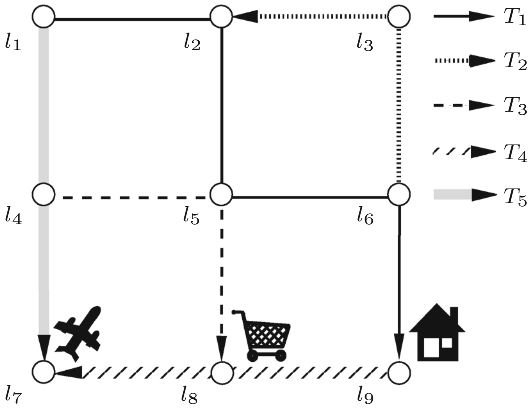

Fig. 1 An example of destination prediction

available from trajectory sharing websites [9,23] or large sets of taxi trajectories [19]. If an ongoing trip partially matches a popular route derived from historical trajectories, the destination of the popular route is very likely to be the destination of the ongoing trip (we refer to the ongoing trip as the *query trajectory*). Shown in Fig. 1 are five historical trajectories: $T_1 = \{l_1, l_2, l_5, l_6, l_9\}$, $T_2 = \{l_6, l_3, l_2\}$, $T_3 = \{l_4, l_5, l_8\}$, $T_4 = \{l_9, l_8, l_7\}$, and $T_5 = \{l_1, l_4, l_7\}$. Each trajectory is represented by a different type of line. For instance, a trip is taken from l_1 to l_4, and this query trajectory $\{l_1, l_4\}$ matches part of the historical trajectory T_5. Therefore, the destination of T_5 (i.e., l_7) is the predicted destination of the query trajectory. In practice, each trajectory here may be associated with

— Xue et al., *Solving the data sparsity problem in destination prediction*, VLDB J., 2015.