Review: Velaga, Quddus & Bristow, "Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems", TRC 09

Sam Toyer

April 2, 2015

This paper introduces a weighted map matching algorithm and an accompanying process for optimising its weights. The algorithm is structured into three distinct phases: identification of the initial link on which the vehicle is travelling, identification of subsequent vehicle locations within a given link, and identification of new links by scoring each road link using a weighted sum of features. Optimising the constants used in MM algorithms like this one is challenging because the most meaningful loss function—the number of links matched incorrectly—requires a complete run through the available data to evaluate, and as such is difficult to manipulate numerically. The paper addresses this problem by learning a simple representation of this loss function, and then choosing the weights which minimise this representation.

The strategy of automatically optimising weights in an MM algorithm is quite novel, and the authors are to be commended for achieving relatively high MM accuracy with such a straightforward algorithm. Demonstrating the effect of the vehicle's environment on the optimal choice of weights was also insightful, as this factor is commonly overlooked in other papers where testing is limited to small urban data sets.

However, there are also some weaknesses which need to be addressed, particularly in the optimisation sections.

The iterative removal of coefficients with a low t-stat during the optimisation process needs further justification, especially in light of the fact that it results in all features related to one of the weights being eliminated entirely when an intercept term is included in the regression for $\mathrm{MM}_{\mathrm{error}}$. In Section 5, the disappearance of this weight is used to justify removal of the intercept, but it seems to me that the "disappearing weight" phenomenon instead argues for removal of the entire coefficient elimination process: after all, removing features from the regression can only hurt its ability to fit the available data. If the motivation for this process was to combat overfitting, then it should be explained why regularisation was not used instead.

The choice to clamp each weight in $[1, 100]$, rather than $[0, 100]$ or even $(-\infty, \infty)$ also lacks justification. The fact that the lower constraint is reached for two of the weights when training on data from a rural setting suggests that these weights are useless—or possibly even harmful—for map matching in that setting, in which case it would be better to allow them to go to zero or become negative. In the same vein, it's not clear what effect the decision not to train on weights which fail to produce a correct initial match has had. Intuitively, it seems like this should introduce significant bias into the $\mathrm{MM}_{\mathrm{error}}$ representation, and it would be helpful for this concern to be addressed in Section 3.3.

Another issue is the absence of cross validation for the learnt representation of $\mathrm{MM}_{\mathrm{error}}$. Ideally, the results of k-fold cross-validation (or similar) should be included for the representation in order to evaluate how well it generalises, rather than just how well it fits the available training data.

Finally, it would be helpful if the algorithm were evaluated using publicly available data in order to facilitate performance comparisons by other researchers. As the authors point out in Section 2, map matching algorithms are currently evaluated on proprietary data sets which often cover only one type of environment or lack DGPS traces, so the extensive data sets used in this paper would be a valuable contribution to the public domain if the rightsholders could be persuaded to release them.

Aside from these shortcomings, this paper is well organised and contributes a useful method for improving the performance of weighted map matching algorithms. As far as future research is concerned, learning $d_{\mathrm{threshold}}$, $h_{\mathrm{threshold}}$ and the scaling constant in $f(D)$ would be an interesting extension, as would applying the optimisation procedure given in this paper to preexisting weighted algorithms—for example, that of Greenfeld (2002) or Quddus et al. (2003), as cited in the introduction—in order to validate its effectiveness compared to hand-tuning.

COMP2550-specific notes

The following context, requested in Part A of the task sheet, may be useful for interpreting this review, but was omitted from the previous page as it would not be appropriate for a real paper review:

- For two of the authors, Abigail Bristow and Nagendra Velaga, this paper appears to be a first foray into map matching; on the other hand, the third author, Mohammed Quddus, was already an authority in the field before producing this paper.
- "Transportation Research Part C: Emerging Technologies" (abbreviated to TRC in the title of this review) is a popular journal for transport-related informatics research, and Google Scholar ranks it highly amongst transportation journals.
- This paper has been relatively highly cited, but I can only find one work which actually builds upon it (Velaga, Quddus & Bristow, "Improving the performance of a topological map-matching algorithm through error detection and correction", Journal of Intelligent Transport Systems, 2012), and that work presented only a minor improvement to the original technique. This suggests that this paper's actual impact on the field has been modest.