Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2015-539-AC1, 2016


Hydrology and Earth System Sciences
Open Access
Discussions

HESSD

Interactive comment

# *Interactive comment on* "Case-based formalization and reasoning method for knowledge in digital terrain analysis – Illustrated by determining the catchment area threshold for extracting drainage networks" *by* C.-Z. Qin et al.

**C.-Z. Qin et al.**

qincz@lreis.ac.cn

Received and published: 19 June 2016

The authors thank Dr. M. Chen for the constructive comments which are helpful for improving the final version of this paper. We answer these comments as below.

Comment 1: In your paper, as determining the CA is a simple function related to DTA and DTA is also just a part of modeling method, how to deal with the complexity problem when conducting comprehensive research and analysis, i.e., how to formalized the complex knowledge about some complex problems, the semantic problem, the structure to represent the knowledge, ect. Do you have some preliminary ideas? I think this is the key step to promote your idea into application in a broader field.

Response: This study explores how to formalize application-context knowledge in DTA and apply it to DTA modeling, when other two types of DTA knowledge (i.e., task knowledge and algorithm knowledge) have been formalized by means of rule or semantic networks (Russell and Norvig, 2009) and hence can be used in existing DTA-assisted tools. Combining the propose method with existing methods for using other two types of DTA knowledge, automated DTA modeling could be implemented to make DTA easy to use for users (especially non-expert users) and ensure that the result model is reasonable comparatively. For other geographic modeling domains, normally the modeling knowledge could also be classified into these three types, i.e., task knowledge, algorithm knowledge, and application-context knowledge. The task and algorithm knowledge in some domains (e.g., watershed modeling) which are more complex than those in DTA have been explored for formalization and inference methods and corresponding tools, such as Gregersen et al. (2007) and Škerjanec et al. (2014) in automated watershed modeling domain. For those geographic modeling domains in which the application context knowledge is also largely non-systematic and tacit knowledge, the case-based idea proposed in this manuscript could also be available to combining with the existing automated modeling methods of using the task and algorithm knowledge in these domains. We will revise the manuscript to include the discussion on this issue.

Comment 2: In page 2, you mentioned that "However, current DTA-assisted tools. . .provide very limited support during the DTA application modeling process". The conclusion is somewhat arbitrary, you may need to provide more arguments here.

Response: Currently, there is no well-established formalization method for application-context knowledge. Existing DTA-assisted tools, which have used the task knowledge and algorithm knowledge, consequently cannot use this type of knowledge to provide more effective support to DTA application modeling process. This situation exists mainly because this type of DTA knowledge is largely non-systematic and tacit knowledge, and often exists only in documents for specific case studies (DTA application instances) or even just in the experience of domain experts. We will revised the manuscript to state this point.

Comment 3: page 3, line 6, “largely inaccurate”

Response: The application-context knowledge of DTA is is largely non-systematic and tacit knowledge. We will revise the manuscript accordingly.

Comment 4: page 4, line 19, “is not necessary to participate”, why? please explain it clearly.

Response: Only the problem part of each case is used to calculate the similarity between the case and the new application problem. The solution of the case with the highest similarity is retrieved as the solution for the new DTA application problem. Thus the solution part of a case does not participate in the reasoning procedure. We will revise the manuscript to state this point.

Comment 5: Page 7, part 4.1, I think maybe you need to provide a table here to explain your quantitative attributes, not just some sentences.

Response: Table 2 lists the attributes used to formalize a case problem and the corresponding similarity functions used in the proposed method.

Comment 6: Page 13, line 2, how to realize your “automatic program” to derive other attributes? Do you mean that these attributes have been processed into a dataset? Otherwise, I think it is hard to automatic match these attributes in their text manual.

Response: In this study, we manually selected the peer-reviewed papers related to the drainage network extraction applications which were published in mainstream journals of related domains. After the study area of each case was set, an automatic program was applied to SRTM DEM or ASTER GDEM of the study area to derive attributes (such as area, total relief, elevation-slope cumulative frequency distribution, and hypsometric curve) for each case. The results were recorded in the case base. We will make it clear in the revised manuscript.

Comment 7: Page 14, line23, 0.4->0.43.

Response: We will revise the manuscript to correct it.

Comment 8: Page 20, table 2, do you consider some other parameters? For example, I think the characteristics of study area are somewhat simple.

Response: The method proposed in current study focuses on DTA domain and considers the area and the terrain condition through a few simple attributes for describing the study area characteristics of a DTA application case. Preliminary evaluation results show the reasonableness of the proposed method. The design of the attributes used to describe the problem part of a case could be improved to describe the domain-specific application-context information in an all-round and efficient manner, which needs additional research. We will revise the manuscript to discuss this issue.

Comment 9: Page 23, the overall similarity, can it be calculated using weighting?

Response: In current method proposed, the overall similarity between a case and a new application problem is determined by applying a minimum operator to synthesizing the similarity values on every attribute in a cautious manner. In the geographical modeling domain, a minimum operator based on the limiting factor principle is often used to synthesize similarity values on multiple attributes (Zhu and Band, 1994). The overall similarity result by a minimum operator is lower (i.e., higher uncertainty of reasoning result) than those from other synthesis means such as weighted-average. Based on the experiment shown in the original manuscript, we also test the effect of calculating the overall similarity by a simple average operator (a representative of weighted-average) instead of the minimum operator. The evaluation results show that the overall similarity for every case increased and the lowest overall similarity among results for 50 evaluation cases increased from 0.47 to 0.68 when the minimum operator was replaced by the simple average operator. Among 50 evaluation cases, the solutions for 13 evaluation cases from the proposed method changed because the cases with the highest similarity resulted by the simple average operator were different from those resulted by the minimum operator. Due to the synthesis by the simple average operator instead of the minimum operator, the relative deviation of river density (E) increased for 10 of these 13 evaluation cases with different solutions, when E slightly decreased for other 3 evaluation cases. The increase of E even reached 20∼80 times for some cases with the overall similarity values larger than 0.8. Because the overall similarity values were larger than 0.8 for most of evaluation cases, there is no a reasonable relationship between the overall similarity value and the E. This shows that the proposed method performed poorly when the simple average operator was used instead of the minimum operator. Note that the simple average is the common representative of weighted-average, and currently it is difficult to choose a more complex weighted-average for synthesizing similarity values on multiple attributes. Therefore the synthesis by a minimum operator is proposed for current method in this study. Additional research is needed to evaluate the similarity calculation method through further test with more types of DTA applications. We will revise the manuscript to include above discussion.

Comment 10: Figure 4, part b. is it right?

Response: Fig. 4b is correct. For this case, the CA threshold resulted from the proposed method is larger than it recorded in the evalution case, which means that the drainage network extracted by using the the CA threshold result is shorter than the original drainage network of this case. The situation shown in Fig. 4a is contrary.

References

Gregersen, J. B., Gijsbers, P. J. A., and Westen, S. J. P.: OpenMI: Open modelling interface, J. Hydroinfo., 9(3), 175-191, 2007.

Škerjanec, M., Atanasova, N., Cerepnalkoski, D, Dzeroski, S., and Kompare, B.: Development of a knowledge library for automated watershed modeling, Environ. Modell. Softw., 54, 60-72, 2014.

Russell, S. and Norvig, P.: Artificial Intelligence: a Modern Approach (3rd Edition), Prentice Hall, 2009.

Zhu, A-X. and Band, L.: A knowledge-based approach to data integration for soil mapping, Can. J. Remote Sens., 20, 408-418, 1994.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2015-539, 2016.