

Position Statement for Workshop on STEC in NLG

Nancy L. Green

University of North Carolina at Greensboro

nlgreen @ uncg.edu

According to the call for participation of this workshop, a shared-task evaluation campaign (STEC) is a competitive approach to research funding where “different approaches to a well-defined problem are compared based on their performance on the same task”. The proposed benefits of this approach are “enhancing the wider NLP community’s view of work in NLG, and in providing a focus for research in the field”. These benefits may not justify the risk.

NLG researchers should be careful that promoting STECs does not have a reductionist effect on the field and does not lead to the marginalization of other important NLG research areas. As many have noted, NLP starts with a well-defined input – text; while NLG does not. Thus, it is possible for the NLP community to use STECs to attack certain well-defined problems, e.g. in text classification, without having to first solve harder computational problems such as understanding all the nuances of meaning in a text. (However, even some computational linguists have complained that this trend in NLP has resulted in neglect of other key research.)

The starting point of an NLG system is not as well-defined since it is often non-linguistic, e.g., a Bayesian network for tumor classification used in an existing decision support system, a database about museum artifacts, or quantitative data requiring further computational analysis to detect trends and other significant features. A STEC providing common inputs might enable researchers to focus on problems in

subsequent stages of the NLG “pipeline”. However, use of a common starting point in the STEC may limit the general applicability of the solutions. Also, it may result in decreased support for NLG research on “what to say”, e.g., reasoning required of an animated agent designed to engage in persuasive conversation with a user about the user’s diet; deciding what to say may require not only nutrition information and dialogue history, but also a model of emotion and argument schemes.

At the other end of the NLG pipeline, application-independent research on how variation in surface generation, rhetorical features, and physical presentation features (such as layout) influences communicative effectiveness is needed. While several NLG systems for generating text variants for use in this kind of experimentation have been developed, the experimentation itself does not fit into a STEC funding model. A STEC could show that one generated result was more successful for a particular task than that of the competitors, but would not address the more fundamental questions whose answers could inform design of many different systems. Also, it is not clear how the narrow focus of a STEC could support the multi-disciplinary research required for multimedia generation, i.e., generation of integrated text and paralinguistic features (speech and gesture) or graphics (pictures, maps, diagrams, data graphics). It would be a mistake to limit the scope of NLG research to the medium of print.