Learning spatial referential words with mobile robots

Simon Dobnik
Centre for Linguistics and Philology
Walton Street, Oxford, OX1 2HG, UK
name.surname@clg.ox.ac.uk

The meaning of spatial words can only be evaluated by establishing a reference to the properties of the environment in which the word is used. For example, in order to evaluate how *near* is *near* or what how *fast* is *fast* in a given context, we need to evaluate properties such as the configuration and position of other objects in the scene, their properties such as animacy and move-ability, the perspective at which the scene is viewed, the size of the scene, and focus on part of the object or its function (Herskovits, 1986).

Rather than encoding the semantic rules that define spatial expressions by hand, we developed a system where such rules are learned (Witten and Frank, 2005) from descriptions produced by human commentators and information that a mobile robot has about itself and its environment (Dissanayake et al., 2001). We concentrate on two scenarios and words that are used in them. In the first scenario, the robot is moving in an enclosed space and the descriptions refer to its motion (*You're going forward slowly. Now you're turning right.*). In the second scenario, the robot is static in an enclosed space which contains real-size objects such as desks, chairs and walls. Here we are primarily interested in prepositional phrases that describe relationships between objects (*The chair is to the left of you. The table is further away than the chair*). The perspective can be varied by changing the location of the robot. Following the learning stage, which is performed offline, the system is able to use this domain specific knowledge to generate new descriptions in new environments or to 'understand' these expressions by providing feedback to the user, either linguistically or by performing motion actions.

If a robot can be taught to 'understand' and use such expressions in a manner that would seem natural to a human observer, then we can be reasonably sure that we have captured at least something important about their semantics. Two kinds of evaluation were performed. First, the accuracy of machine learning classifiers was tested on independent test sets using 10-fold cross-validation. Second, the classifiers were tested live against a human evaluator who judged the acceptability of a generated move or description. The accuracies obtained from the second evaluation were higher (mean accuracy of 86.63%) than those from the first one (mean accuracy of 62.80%) for a particular set of classifiers used in the comparison. The results show that we can learn the semantics of referential expressions. The differences between the evaluations indicate that a single scene or motion may be described by alternative descriptions which were not stipulated in the learning set.

References

M. W. M. Gamini Dissanayake, Paul Newman, Steven Clark, Hugh F. Durrant-Whyte and M. Csorba. 2001. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotic and Automation*, 17(3):229–241.

Annette Herskovits. 1986. *Language and spatial cognition: an interdisciplinary study of the prepositions in English.* Studies in natural language processing. Cambridge: Cambridge University Press.

Ian H. Witten and Eibe Frank. 2005. *Data mining: practical machine learning tools and techniques.* 2nd edition. Morgan Kaufmann.