2010

An Empirically-based Model for Perspective Selection in Route-Finding Dialogues

Robert Ross
Dublin Institute of Technology, robert.ross@dit.ie

Kavita E. Thomas
University of Bremen

Recommended Citation

Ross, R. J. and Thomas, K. E. An Empirically-based Model for Perspective Selection in Route-Finding Dialogues. Poster paper presented at Spatial Cognition, 2010. Portland, Oregon. USA.

An Empirically-based Model for Perspective Selection in Route-Finding Dialogues

Robert J Ross[1] and Kavita E. Thomas[2]

[1] Dublin Institute of Technology, Ireland
robert.ross@dit.ie
[2] University of Bremen, Germany
kathomas@uni-bremen.de

1 Introduction

In this work we aim to computationally model the extent to which certain empirical factors affect spatial perspective selection as used in route-finding dialogues. In such dialogues, both interlocutors need to adopt a spatial perspective in which to describe movement direction. In map-based tasks such as the one we are concerned with, two perspective choices are typically available, i.e., route perspective, where projective terms are defined with respect to the perspective of the route follower themselves, e.g., "go to your right", or survey perspective, where projective terms are defined with respect to a global or allocentric perspective, e.g., "go down", or "go toward the top of the screen". Addressees must be able to assign perspectives to a given spatial term in order to correctly interpret the utterance it is contained in. However the most frequent directional terms, i.e., left and right, can be used in either route or survey perspective, and perspective is not typically marked explicitly at the lexical level. Generally addressees do correctly assign perspective to projective terms, even when perspective is not indicated explicitly in language, but misunderstandings may occur and clarification is often necessary[3].

To develop computational systems which can adequately assign perspective to spatial terms which do not describe perspective explicitly, we need computational models which account for the factors which influence perspective choice. Physical orientation of the instructee and the intended direction of movement described by a spatial term are two such potential factors. However, while orientation and instruction direction would seem to be important factors in perspective use, it is well known that people are far from consistent in their use of perspective, and that a great many other factors can influence perspective use. For example, Watson et al found that partners tend to align reference frames in dialogue, both within and between spatial axes in a task where they describe locations to each other [1]. Such an influence of recent perspective on current perspective can also occur between utterances in monologue i.e., with a speaker's

[3] In the corpus presented later in the paper, perspective-querying clarification requests composed 14.29% of all clarification requests for the whole corpus.

own earlier contributions [2]. Moreover, perspective choice may also be influenced by discourse function, i.e., current or previous dialogue acts; for example, Striegnitz et al show that perspective use in gesture is related to the type of linguistic dialogue acts communicated in current and preceding turns [3]. Goeschler et al have also observed that for a data set similar to the one which we consider in our own work, the mean percentage use of survey perspective shows a weak positive correlation with the number of times participants perform a basic route instruction task [4].

2 Data Collection & Annotation

To empirically estimate the influence of spatial and discourse factors on perspective choice, we annotated a human-human route instruction corpus with respect to a number of features. The corpus we used was collected for a scenario in which two humans interact via a chat box while observing a screen which depicts a shared environment and the location of one participant within that environment. One participant, the route giver had knowledge of the goal location and could see the location of the route follower at any given point in the interaction, but could not directly move the route follower. The route follower on the other hand had a joystick to move their avatar around the shared map, but had no knowledge of the final location. In total there were 15 dyads and each dyad performed a basic route instruction task up to 11 times. We retained the first 6 of 11 trials for each dyad for annotation. The resultant corpus contained a total of 693 utterances, 339 of which (48.91%) indicated spatial perspective. Full details of the corpus collection procedure, samples of the interactions, and a basic analysis of the language used in that corpus has been provided in Tenbrink et al [5]. For our current work, the corpus was annotated for perspective use as well as a number of empirical factors predicted to play a role in determining perspective, i.e., orientation of the avatar, the intended direction underlying a given instruction, previously used perspective for both speakers and dialogue act. Part of the data-set was coded by a second annotator to assess the reliability of annotation. Cohen's Kappa scores (κ) of 0.77, 0.86, and 0.57 were found for the features perspective, orientation, and instruction direction respectively.

3 Data Analysis

Since our goal is to produce computational models which describe the factors which affect perspective use, we first assessed the effect of individual factors on perspective choice. For this analysis, we considered only utterances in which a perspective was identified. From this set we eliminated all cases of mixed (e.g., "on your right, that's up") and unclear perspectives, resulting in a data set consisting of 290 utterances of which 15.86% were conflated (i.e., the same linguistic expression maps to the same spatial direction for both perspectives), 67.59% were route, and 16.55% were survey. We then assessed the independence of perspective choice with respect to predictor variables. Chi-square and Fisher

Model	Type	Predictors	Accuracy	κ
1	MLR	Ori*Dir+PPSS+TN	80.69	0.57
2	MLR	Ori*Dir+PPSS	79.65	0.55
3	MLR	Ori*Dir	77.24	0.43
4	NB	Ori Dir PPSS TN	82.41	0.62
5	NB	Ori Dir PPSS	82.65	0.62
6	NB	Ori Dir	77.57	0.42

Table 1. Results of model evaluation.

tests for independence showed that a null hypothesis assuming independence of perspective and predictor value should be rejected at the 95% confidence threshold for orientation (p=9.836e-27), instruction direction (p = 3.307e-10), previous perspective of the same speaker (p = 1.139e-06), the dyad (p = 2.315e-05), and trial number (p = 3.918e-06). Independence of other predictor factors with respect to perspective choice, i.e., dialogue role (p=0.49), dialogue act direction (p = 0.21), and the previous perspective of the other speaker (p = 0.65), could not be rejected however.

In order to arrive at a classifier which enables us to predict perspective given the annotated empirical factors discussed above, we trained and evaluated both a Naive Bayes classifier [6] and a classification model based on Multinomial Logit Regression [7] for a range of predictor combinations. Multinomial Logit Regression (MLR) is a statistical regression technique which generalizes logistic regression to more than two levels of response variable, while a Naive Bayes (NB) classifier is a machine learning technique based on the Bayes Theorem. Both MLR and NB may be applied to data consisting of mixed predictor variables and a (categorical) multinomial response variable, and as such are well suited to the perspective use data. However, both models also make a number of additional assumptions which must be considered. The main assumption of the Naive Bayes classifier, and an assumption of MLR to a lesser extent, is that predictor variables are independent of each other. The MLR technique also presupposes that response categories are mutually exclusive (i.e., that the independence assumption holds).

Each classification technique was trained and evaluated through 10-fold cross validation. Table 1 shows accuracy and Kappa scores calculated from the confusion matrices for a selection of MLR models and NB classifiers trained on our corpus for combinations of significant perspective predictor variables. The terms used for describing predictive models include: *Ori*, the annotated orientation of the speaker; *Dir*, the intended instruction direction; *PPSS* the previous perspective of the same speaker; and *TN*, the trial number. For MLR models, individual factors can either be considered independently or we can consider the interaction between factors. This is noted in the model description through the use of the addition symbol (+) for the addition of independent items to the model and the multiplication symbol (*) for interactions of factors. It should be noted that we started with a fully interacting model of all predictor variables and refined that model through stepwise elimination of non-significant predictor variables to the set of models shown in Table 1.

Results show that NB based methods slightly outperformed the MLR models for all investigated models; this is likely due to former's better handling of noisy data. As can be seen in the best performing models (models 4 and 5), turn number does not significantly influence model accuracy and can be removed from the predictive model. It should also be noted that due to inter-dyad variability, dyad was also shown to significantly increase model performance, but we omit this factor in our models as we are interested in producing models which are generalized to new dyads. Finally, all models perform better than a simplistic *route-always* predictive model (Accuracy=67.59,κ=0).

4 Conclusions and Future Directions

The main contribution of this work is an empirically based method for spatial perspective selection in natural language dialogues. Orientation, intended direction, and previous self perspective were found to correlate with particular perspective choices, and predictive models based on these factors achieved higher predictive power than the selection of one perspective only.

Determining perspective from a corpus can easily be influenced by noisy data as well as inconsistent speakers. In order to verify the significance of the factors presented here, we are currently running empirical studies to estimate single variable effects on perspective selection under controlled conditions. Moreover, since our goal is to improve the quality of spoken interaction with situated spatial applications, we also plan to incorporate these perspective choice models into dialogue system applications and evaluate whether using such models do indeed provide any benefit to human-computer interaction.

References

1. Watson, M.E., Pickering, M.J., Branigan, H.P.: Alignment of reference frames in dialogue. In: Proceedings of the 26th Annual Conference of the Cognitive Science Society, Mahwah, NJ, Lawrence Erlbaum Associates (2004)
2. Vorwerg, C.: Consistency in Successive Spatial Utterances. In: Spatial Language and Dialogue. Oxford Linguistics (2009)
3. Striegnitz, K., Tepper, P., Lovett, A., Cassel, J.: Knowledge representation for generating locating gestures in route directions. In: Spatial Language in Dialogue. Oxford University Press. (2008)
4. Goschler, J., Andonova, E., Ross, R.J.: Perspective use and perspective shift in spatial dialogue. In: Spatial Cognition VI: Learning, Reasoning and Talking about Space, Springer (2008)
5. Tenbrink, T., Ross, R., Thomas, K., Andonova, E.: Efficient negotiation of routes in map-based dialogues between humans and dialogue systems. Journal of Visual Language and Computation (In Press)
6. Mitchell, T.: Machine Learning. McGraw-Hill (1997)
7. Hosmer, D.W., Lemeshow, S.: Applied Logistic Regression. 2nd edition edn. Wiley (2000)