USP-EACH: Improved Frequency-based Greedy Attribute Selection

Diego Jesus de Lucena
University of São Paulo
São Paulo - Brazil
diego.si@usp.br

Ivandré Paraboni
University of São Paulo
São Paulo - Brazil
ivandre@usp.br

Abstract

We present a follow-up of our previous frequency-based greedy attribute selection strategy. The current version takes into account also the instructions given to the participants of TUNA trials regarding the use of location information, showing an overall improvement on string-edit distance values driven by the results on the Furniture domain.

1 Introduction

In previous work (Lucena & Paraboni, 2008) we presented a frequency-based greedy attribute selection strategy submitted to the TUNA Challenge 2008. Presently we further the issue by taking additional information into account - namely, the trial condition information available from the TUNA data - and report improved results for string-edit distance as required for the 2009 competition.

2 Background

In Lucena & Paraboni (2008) we presented a combined strategy based on attribute frequency and certain aspects of a greedy attribute selection strategy for referring expressions generation. A list P of attributes sorted by frequency is the centre piece of the following selection strategy:

- select all attributes whose relative frequency falls above a threshold value t (t was estimated to be 0.8 for both Furniture and People domains.)
- if the resulting description uniquely describes the target object, then finalizes.
- if not, starting from the most frequent attribute in P, search exhaustively for an attribute g such that g, if selected, would rule out all remaining distractors in the context.

The overall effect obtained is twofold: on the one hand, in a complex situation of reference (in which many attributes may rule out many distractors, but more than one will be required to achieve uniqueness) the algorithm simply selects *frequent* attributes. This may be comparable to a human speaker who has to single out the target object but who does not have the means to come up with the 'right' attribute straight away.

On the other hand, as the number of distractors decreases, a single attribute capable of ruling out all distractors will eventually emerge, forcing the algorithm to switch to a *greedy* strategy and finalize. Once again, this may be comparable to what a human speaker may do when an appropriate attribute becomes sufficiently salient and all distractors in the context can be ruled out at once.

The above approach performed fairly well (at least considering its simplicity) as reported in Lucena & Paraboni (2008). However, there is one major source of information available from the TUNA data that was *not* taken into account in the above strategy: the *trial condition* represented by the +/- LOC feature. Because this feature distinguishes the very kinds of instruction given to each participant to complete the TUNA task, the information provided by -/+ LOC is likely to have a significant impact on the overall results. This clear gap in our previous work represents an opportunity for improvement discussed in the next section.

3 Algorithm

The present work is a refined version of the original frequency-based greedy attribute selection strategy submitted to the TUNA Challenge 2008 (Lucena & Paraboni, 2008), now taking also the trial condition (+/-LOC) into account.

Proceedings of the 12th European Workshop on Natural Language Generation, pages 189–190,
Athens, Greece, 30 – 31 March 2009.

In the TUNA data, +LOC indicates the instances of the experiment in which participants were told that they were allowed to refer to the X,Y coordinates of the screen (i.e., selecting the X- and/or Y-DIMENSION attributes), whereas -LOC indicates the trials in which they were discouraged (but not prevented) to do so. In practice, references in +LOC trials are more likely to convey the X- and Y-DIMENSION attributes than those in which the -LOC condition was applied.

Our modified algorithm simply consists of computing separated frequency lists for +LOC and -LOC trial conditions, and then using the original frequency-based greedy approach with each list accordingly. In practice, descriptions are now generated in two different ways, depending on the trial condition, which may promote the X- and Y-DIMENSION attributes to higher positions in the list P when +LOC applies.

Using the TUNA Challenge 2009 development data set, the attribute selection task was performed as above. For the surface realisation task, we have reused the English language surface realisation module provided by Irene Langkilde-Geary for the TUNA Challenge 2008.

4 Results

The following Figure 1 shows mean sting-edit distance and BLEU-3 scores computed using the evaluation tool provided by the TUNA Challenge team. For ease of comparison with our previous work, we also present Dice and MASI scores computed as in the previous TUNA Challenge, although these scores were not required for the current competition.

The most relevant comparison with our previous work is observed in the overall string-edit distance values in Figure 1: considering that in Lucena & Paraboni (2008) we reported 6.12 edit-distance for Furniture and 7.38 for People, the overall improvement (driven by the descriptions in the Furniture domain) may be explained by the fact that the current version makes more accurate decisions as to when to use these attributes according to the instructions given to the participants of the TUNA trials (the trial condition +/-LOC.)

On the other hand, the divide between +LOC and -LOC strategies does not have a significant effect on the results based on the semantics of the description (i.e., Dice and MASI scores), which remain the same as those obtained previously. This may be explained by the fact that using location information inappropriately counts as one single error in Dice/MASI calculations, but it may have a much greater impact on the wording of the surface string (e.g., one single use of the X-DIMENSION attribute may be realized as “on the far left”, adding four words to the descriptions.)

	Overall	Furniture	People
String-edit distance	6.03	4.78	7.50
BLEU-3	0.19	0.31	0.04
Dice	0.74	0.82	0.65
MASI	0.53	0.63	0.41

Figure 1. Results (TUNA Challenge 2009 development data set)

5 Conclusion

We have presented a refined version of our previous frequency-based greedy attribute selection strategy. The current version takes into account the instructions given to the participants of TUNA trials regarding the use of location information (the trial condition +/-LOC.)

Results obtained using the TUNA Challenge 2009 development data set show improvements on string-edit distance, suggesting that the generated descriptions resemble more closely those seen in the TUNA corpus.

Acknowledgments

This work has been supported by CNPq-Brazil (484015/2007-9) and FAPESP (2006/03941-7).

References

Lucena, Diego Jesus de, and Ivandré Paraboni (2008) *USP-EACH Frequency-based Greedy Attribute Selection for Referring Expressions Generation*. Proc. of INLG-2008 (TUNA Challenge 2008). Salt Fork, US, pp.219-220.