USP-EACH Frequency-based Greedy Attribute Selection for Referring Expressions Generation

Diego Jesus de Lucena
Escola de Artes, Ciências e Humanidades
University of São Paulo – USP
Av. Arlindo Bettio, 1000 - São Paulo, Brazil
diego.si@usp.br

Ivandré Paraboni
Escola de Artes, Ciências e Humanidades
University of São Paulo – USP
Av. Arlindo Bettio, 1000 - São Paulo, Brazil
ivandre@usp.br

Abstract

Both greedy and domain-oriented REG algorithms have significant strengths but tend to perform poorly according to humanlikeness criteria as measured by, e.g., Dice scores. In this work we describe an attempt to combine both perspectives into a single attribute selection strategy to be used as part of the Dale & Reiter Incremental algorithm in the REG Challenge 2008, and the results in both Furniture and People domains.

1 Introduction

Minimality and Humanlikeness in REG are often conflicting goals. Greedy algorithms tend to favour shorter descriptions, but in doing so their output may look unnatural. On the other hand, domain-oriented algorithms that arguably favour more "human-like" strategies (e.g., selecting the most typical attributes first) pay little or no attention to minimality, and as a result the generated descriptions may become overly long or clumsy.

Which strategy might a human speaker favour? In this work we describe an algorithm that disregards minimality entirely and attempts to select 'typical' attributes based on two simple assumptions: first, when facing a complex context with a large number of objects, an attempt to compute the precise attribute capable of ruling out the largest possible number of distractors is not only hard

(from the computational point of view), but also less natural than simply using typical (e.g., frequent) attributes. On the other hand, as the number of distractors decreases, it may become gradually clearer for the speaker which attributes are most helpful to achieve uniqueness, up to the point in which she may naturally switch to a 'greedy' strategy and finalize the description. These assumptions are implemented as an attribute selection strategy to be used with the Incremental algorithm (Dale & Reiter, 1995) described below.

2 System Description

We take a simple view of humanlikeness in which the list of preferred attributes is sorted by relative frequency[1] as seen in the training data. The resulting list P is the centre piece of the following attribute selection strategy:

(1) select all attributes whose relative frequency falls above a trainable threshold value t (in our experiments t is estimated to be 0.8 for both Furniture and People domains.)

(2) if the resulting description uniquely describes the target object, then finalizes.

(3) 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.

---

[1] This contrasts the work in Kelleher (2007), which takes into account *absolute* counts seen in the training data.

(4) if such attribute *g* exists, then *g* is selected and the algorithm finalizes.

(5) if not, select the most frequent attribute *f* that can rule out at least one distractor, and repeat steps (3-5).

The selection of attribute *g* stands for the greedy component of our approach, whilst the initial attributes in step 1 and the attribute *f* account for our 'humanlikeness as frequency' assumption. The overall effect attempted is the following:

- Highly frequent attributes are always selected. In our tests this means that the attributes *type* and *colour* were always included in Furniture descriptions, and *type* was always included in People descriptions (in both cases this is so regardless of discriminatory power.) As a result, we can only produce minimal descriptions by chance.

- 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 the most frequent attributes, perhaps not unlike a human speaker who has to single out the target object but who does not have the time or re-sources to come up with the 'best' attribute straight away.

- 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 might be just what humans do when a suitable (i.e., economical) attribute becomes sufficiently salient and all distractors in the context can be ruled out at once.

3 Results

Below we summarize our results for Task 1 (Attribute Selection) and also for Task 3 (Attribute Selection and Surface Realisation combined) for the REG 2008 development data set (80 instances for Furniture and 68 instances for People.) As expected, our algorithm is heavily penalized in the Minimality criteria but performs reasonably well in Humanlikeness (Dice and MASI.) if compared to the systems presented in the previous GRE Challenge.

	Overall		Furniture		People	
	Mean	SD	Mean	SD	Mean	SD
Dice	0.75	0.25	0.82	0.22	0.66	0.26
MASI	0.53	0.39	0.62	0.39	0.42	0.35
Accuracy	0.37	0.48	0.49	0.50	0.24	0.43
Uniqueness	1.00	-	1.00	-	1.00	-
Minimality	-	-	-	-	-	-
String-edit distance	6.70	3.09	6.13	3.28	7.38	2.72
String-accuracy	0.02	0.14	0.04	0.19	-	-

Figure 1. Attribute Selection and Surface Realisation results

Acknowledgments

References

Dale, Robert and Ehud Reiter. 1995. *Computational interpretations of the Gricean maxims in the generation of referring expressions*. Cognitive Science (19).

Kelleher, J.D. (2007) *DIT - Frequency Based Incremental Attribute Selection for GRE*. MT Summit XI Workshop Using Corpora for Natural Language Generation: Language Generation and Machine Translation, pp. 90-91.