Learning finite state transducers using bilingual phrases[1]

Jorge González, Germán Sanchis, and Francisco Casacuberta

Instituto Tecnológico de Informática
Universidad Politécnica de Valencia
{jgonzalez, gsanchis, fcn}@dsic.upv.es

Abstract. Statistical Machine Translation is receiving more and more attention every day due to the success that the phrase-based alignment models are obtaining. However, despite their power, state-of-the-art systems using these models present a series of disadvantages that lessen their effectiveness in working environments where temporal or spacial computational resources are limited. A finite-state framework represents an interesting alternative because it constitutes an efficient paradigm where quality and realtime factors are properly integrated in order to build translation devices that may be of help for their potential users. Here, we describe a way to use the bilingual information in a phrase-based model in order to implement a phrase-based ngram model using finite state transducers. It will be worth the trouble due to the notable decrease in computational requirements that finite state transducers present in practice with respect to the use of some well-known stack-decoding algorithms. Results for the French-English EuroParl benchmark corpus from the 2006 Workshop on Machine Translation of the ACL are reported.

1 Introduction

Machine translation (MT) is an important area of Information Society Technologies in different research frameworks of the European Union. While the development of a classical MT system requires a great human effort, *Statistical machine translation* (SMT) has proved to be an interesting framework due to being able to automatically build MT systems if adequate parallel corpora are provided [1].

Given a sentence $\mathbf{s}$ from a source language, it is commonly accepted that a convenient way to express the SMT problem is through the Bayes' rule [1]:

$$\hat{\mathbf{t}} = \underset{\mathrm{t}}{\operatorname{argmax}} \Pr(\mathbf{t}|\mathbf{s}) = \underset{\mathrm{t}}{\operatorname{argmax}} \Pr(\mathbf{t}) \cdot \Pr(\mathbf{s}|\mathbf{t}) \tag{1}$$

where $\hat{\mathbf{t}}$ stands for the most likely hypothesis, according to the model, from all the possible output sentences $\mathbf{t}$. $\Pr(\mathbf{t})$ is frequently approached by a *language model*, which assigns high probability to well formed target sentences, and $\Pr(\mathbf{s}|\mathbf{t})$ is modelled by a *translation model* that is based on stochastic dictionaries and alignment models [2, 3].

[1] This work has been partially supported by the EC (FEDER) and the Spanish projects TIN2006-15694-C02-01 and the Consolider Ingenio 2010 CSD2007-00018