

Primeiros Experimentos na Investigação e Avaliação da Tradução Automática Estatística Inglês-Português

Israel Aono Nunes¹, Helena de Medeiros Caseli¹

¹Departamento de Computação, Universidade Federal de São Carlos (UFSCar)
Rod. Washington Luís, Km 235, CP 676, CEP 13565-905 São Carlos-SP

israelaono@gmail.com, helenacaseli@dc.ufscar.br

Abstract. *This paper describes some experiments carried out to train and test statistical machine translation models for Brazilian Portuguese and English parallel texts. Here we present BLEU and NIST values for the translation from English to Portuguese with and without a tuning step. Satisfactory results were obtained considering the small corpus size used for training the models. In the near future we will compare these results with those obtained from training and testing the statistical factored translation models.*

Resumo. *Esse artigo descreve alguns experimentos desenvolvidos com o intuito de treinar e testar os modelos de tradução automática estatística para textos paralelos escritos em português do Brasil e inglês. Neste artigo são apresentados os valores de BLEU e NIST na tradução de inglês para português com e sem a realização de uma etapa de otimização (tuning). Os resultados obtidos podem ser considerados satisfatórios para o pequeno corpus usado no treinamento. Em breve, esses resultados serão usados como base para a compração com os valores obtidos no treinamento e teste de modelos de tradução fatorada.*

1. Introdução

Este artigo relata os primeiros experimentos na investigação de métodos estatísticos para a tradução automática envolvendo o Português do Brasil. A Tradução Automática (TA), uma das principais áreas de Processamento de Língua Natural, pode ser descrita como o processo em que uma língua fonte é traduzida para uma língua alvo com a utilização de programas de computador. A TA estatística começou a ser pesquisada com sucesso em meados dos anos 80 com o projeto Candide da IBM [Koerner e Asher 1995].

Inicialmente, a tradução era realizada mapeando-se palavra por palavra, mais tarde foi desenvolvida uma abordagem mais eficiente que se caracteriza pelo mapeamento de frases: *phrase-based machine translation*.¹ Os sistemas de TA estatística baseados em frases como [Koehn et al. 2003] e [Och e Ney 2004] são considerados o estado da arte na TA de acordo com as medidas de avaliação automática BLEU [Papineni et al. 2002] e NIST [Doddington 2002].

Mais recentemente, uma nova estratégia de TA estatística, a tradução fatorada, surgiu com a proposta de utilização de informações adicionais – sintáticas, morfológicas, de *part-of-speech* ou uma combinação dessas – juntamente com os fatores estatísticos para gerar os modelos de tradução [Koehn e Hoang 2007].

¹Vale dizer que na TA estatística uma frase (*phrase*) é uma sequência de palavras independente de ser um sintagma propriamente dito.

Nesse contexto, este artigo apresenta os resultados dos primeiros experimentos com TA estatística de textos escritos em inglês para textos em português do Brasil com o intuito de comparar a tradução baseada em frases tradicional com a aplicação de uma etapa de otimização (*tuning*). Porém, antes da apresentação desses experimentos e de seus resultados, na seção 3, a seção 2 apresenta um breve relato sobre a TA estatística. Por fim, a seção 4 traz algumas considerações finais e propostas de trabalhos futuros.

2. Tradução Automática Estatística

Resumidamente, o processo de TA estatística envolve a construção de dois modelos: um modelo de língua e um modelo de tradução. Esses modelos são construídos a partir de um corpus paralelo² de treinamento com base nas probabilidades de uma dada palavra (ou sequência de palavras) fonte ser traduzida para uma palavra (ou sequência de palavras) alvo.

Para o cálculo dessas probabilidades são utilizados os alinhamentos lexicais gerados com base nos modelos IBM [Brown et al. 1993] e HMM [Vogel et al. 1996] implementados, por exemplo, pelo alinhador estatístico automático GIZA++³ [Och e Ney 2000]. GIZA++ faz parte do *toolkit* de código aberto Moses⁴ [Koehn et al. 2007], utilizado nos experimentos descritos neste artigo.

Moses pode ser executado com diversos parâmetros que especificam, por exemplo, qual a configuração que deve ser seguida para se determinar os alinhamentos lexicais usando GIZA++. Os principais parâmetros usados nos experimentos descritos nesse artigo foram: 5 iterações dos modelos IBM-1 e HMM e 3 iterações dos modelos IBM-3 e IBM-4 para GIZA++, tamanho máximo de frase igual a 7 com a opção de reordenamento.

Além da TA estatística baseada em frases, Moses permite a realização da tradução fatorada, ou seja, a tradução de textos com informações adicionais (traços morfológicos, lemas, *part-of-speech* etc.) cujos primeiros resultados para o português do Brasil serão divulgados em breve.

3. Experimentos e Resultados

Os experimentos apresentados neste artigo foram realizados no sentido inglês-português utilizando Moses, ou seja, a língua de origem é o inglês enquanto a língua alvo é o português. O corpus paralelo de treinamento está composto por 646 artigos da versão *online* da revista científica *Pesquisa Fapesp*⁵, com 17.397 pares de sentenças paralelas num total de 1.026.512 *tokens*.

Para teste foi utilizado um conjunto de 649 sentenças paralelas, obtidas da mesma revista, porém diferentes daqueles do corpus de treinamento. Para cada uma dessas sentenças de teste coletou-se, também a tradução correspondente na revista para servir de referência na avaliação automática usando as medidas BLEU [Papineni et al. 2002] e NIST [Doddington 2002].

Entre as várias configurações testadas nos primeiros experimentos, duas são apresentadas neste artigo: (i) TA estatística baseada em frases sem a realização de *tuning*

²Um corpus paralelo é um conjunto de sentenças fonte e suas traduções para outra língua.

³<http://code.google.com/p/giza-pp/>

⁴<http://www.statmt.org/moses>

⁵<http://www.revistapesquisa.fapesp.br/>

Tabela 1. Valores de BLEU e NIST para a tradução de inglês para português com e sem o *tuning*

	BLEU	NIST
sem <i>tuning</i>	0,3589	7,8312
com <i>tuning</i>	0,3209	7,5745

Tabela 2. Exemplo de uma sentença em inglês traduzida para português do Brasil

fonte	the research center into vaccines of the national institute of health of the united states managed the first victory against the ebola virus .
referência	o centro de pesquisa de vacinas do instituto nacional de saúde dos estados unidos conseguiu a primeira vitória contra o vírus ebola .
sem <i>tuning</i>	o centro de pesquisa <u>em</u> vacinas do instituto nacional de saúde dos estados unidos , conseguiu <u>o</u> primeiro vitória contra o vírus ebola .
com <i>tuning</i>	_ centro de <u>pesquisas em</u> vacinas do instituto nacional de saúde <u>de</u> estados unidos conseguiu <u>o</u> primeiro vitória contra o _ ebola .

(otimização) do treinamento e (i) outra com essa etapa de otimização. Para permitir a realização do *tuning*, foi considerado um terceiro conjunto de textos paralelos, com cerca de 46.000 palavras, obtidos da mesma revista científica porém diferentes dos textos de treinamento e teste. Os valores de BLEU e NIST obtidos após o treinamento e teste de acordo com as duas configurações citadas estão na Tabela 1 e um exemplo de tradução realizada por cada configuração é apresentado na Tabela 2.

Esperava-se que a realização de *tuning* melhorasse o desempenho do tradutor aumentando as pontuações obtidas anteriormente, porém os resultados não foram os esperados talvez devido a limitações de tamanho do corpus de *tuning* já que a partir da sexta iteração do *script* de otimização os resultados não tiveram melhoras significativas.

Como se pode perceber pelo exemplo da Tabela 2, a versão sem o *tuning* apresentou cerca de 4 *tokens* (sublinhados) diferentes em relação à referência, enquanto na versão obtida após a realização do *tuning* as diferenças subiram para 7. Também vale notar que alguns erros ocorreram em ambas as configurações como a falta de concordância entre *o primeiro* e *vitória*, o que pode ser um indício das limitações da abordagem de TA estatística.

4. Considerações Finais e Trabalhos Futuros

Os experimentos descritos neste artigo servirão de base na investigação de abordagens mais complexas como a tradução fatorada, cujos experimentos já estão em desenvolvimento e os resultados serão publicados em breve.

Novos experimentos também serão realizados com *tuning* na tentativa de se obter melhores (e não piores como foi o caso apresentado neste artigo) resultados para BLEU e NIST. Além desses, experimentos com outros sentidos de tradução (português-inglês, por exemplo) e pares de línguas (português-espanhol, por exemplo) deverão ser realizados.

Agradecimentos

Agradecemos o apoio do Programa Integrado de Apoio ao Docente Recém Doutor PI-ADRD/PUIC/UFSCar e da FAPESP (na construção do corpus paralelo).

Referências

- Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra e Robert. L. Mercer (1993). The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 19:263–311.
- G. Doddington (2002). Automatic evaluation of machine translation quality using n-gram cooccurrence statistics. Em *Proceedings of HLT 2002*, páginas 128–132.
- P. Koehn e H. Hoang (2007). Factored translation models. Em *Proceedings of EMNLP-2007*, páginas 868–876, Prague.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin e Evan Herbst (2007). Moses: Open Source Toolkit for Statistical Machine Translation. Em *Proceedings of ACL 2007*, páginas 177–180, Prague, Czech Republic.
- P. Koehn, F. J. Och e D. Marcu (2003). Statistical phrase-based translation. Em *Proceedings of the Human Language Technology (HLT/NAACL 2003)*, páginas 127–133.
- E. F. K. Koerner e R. E. Asher, editores (1995). *Concise history of the language sciences: from the Sumerians to the cognitivists*, páginas 431–445. Pergamon Press, Oxford.
- F. J. Och e H. Ney (2000). Improved statistical alignment models. Em *Proceedings of ACL 2000*, páginas 440–447, Hong Kong, China.
- Franz Josef Och e Hermann Ney (2004). The Alignment Template Approach to Statistical Machine Translation. *Computational Linguistics*, 30(4):417–449.
- K. Papineni, S. Roukos, T. Ward e W. J. Zhu (2002). BLEU: a method for automatic evaluation of machine translation. Em *Proceedings of ACL 2002*, páginas 311–318.
- S. Vogel, H. Ney e C. Tillmann (1996). HMM-based word alignment statistical translation. Em *Proceedings of COLING 1996*, páginas 836–841, Copenhagen.