Modelos de Linguagem N-grama para Reconhecimento de Voz com Grande Vocabulário

Ênio Silva, Marcus Pantoja, Jackline Celidônio e Aldebaro Klautau

[1]Laboratório de Processamento de Sinais – Universidade Federal do Pará
DEEC-CT, Belém, PA, 66075-900, Brasil
http://www.laps.ufpa.br
E-mail: a.klautau@ieee.org

Abstract. *This work describes preliminary results on N-gram language models applied to Brazilian Portuguese. The project is part of an effort to develop a large vocabulary continuous speech recognition system, where language modeling plays a fundamental role. We present a brief summary of state-of-art techniques, including the recently proposed interpolated additive (AI) model. We also describe simulation results, which show that the AI model is competitive with some well-established techniques.*

Resumo. *Este trabalho apresenta resultados preliminares acerca do uso de modelos estatísticos N-grama para o português brasileiro. O mesmo se insere no âmbito do desenvolvimento de um sistema de reconhecimento de voz com suporte a grandes vocabulários, onde a modelagem da linguagem é um aspecto fundamental. Apresentamos um breve sumário das técnicas do estado-da-arte, dentre as quais o modelo aditivo interpolado, recentemente proposto. Descrevemos também, de forma comparativa, os resultados obtidos por essas técnicas.*

1. Introdução

A modelagem da linguagem é ingrediente essencial de muitos sistemas computacionais, tais como reconhecimento de voz. Geralmente, os sistemas de reconhecimento de voz (SRV) são baseados em cadeias escondidas de Markov (HMMs, de *hidden Markov models*) [Huang et al., 2001]. Esses sistemas convertem o sinal de voz digitalizado em uma matriz $\mathbf{X}$ de *parâmetros*, e buscam a seqüência de palavras W que maximiza a probabilidade condicional

$$\hat{W} = \arg\max_{W} p(W|\mathbf{X}).$$

Na prática, usa-se a regra de Bayes para implementar a busca através de:

$$\hat{W} = \arg\max_{W} p(W|\mathbf{X}) = \arg\max_{W} \frac{p(\mathbf{X}|W)p(W)}{p(\mathbf{X})} = \arg\max_{W} p(\mathbf{X}|W)p(W),$$

com $P(\mathbf{X})$ sendo desprezado pois não depende de W. Para cada W, os valores de $P(\mathbf{X}|W)$ e $P(W)$ são fornecidos pelos *modelos acústico* e *de linguagem* (ou *língua* [Pessoa et al., 1999b]), respectivamente. Ambos modelos são imprescindíveis em SRV, mas esse trabalho concentra-se nos modelos de linguagem.

Modelos estatísticos de linguagem fornecem a probabilidade de uma seqüência de palavras $W = w_0 \dots w_l$, a qual também será representada por w_0^l e chamada genericamente de *sentença*. Nós assumimos que w_0 é um símbolo para o início da sentença consistindo de $l-1$ palavras, e w_l é um símbolo para o final da sentença. O modelo de linguagem mais utilizado para aplicações em reconhecimento de voz usa a aproximação *n*-grama , a qual assume que a distribuição de probabilidade para a palavra atual depende somente das $n-1$ palavras precedentes:

$$p(w_1^l|w_0) = \prod_{i=1}^{l} p(w_i|w_0^{i-1}) \approx \prod_{i=1}^{l} p(w_i|w_{i-n+1}^{i-1}).$$

Ressalta-se que a probabilidade para o símbolo final da sentença será avaliada no fim da sentença como se fosse uma outra palavra, enquanto que o começo da sentença é tratado apenas como uma informação do histórico (ou *contexto*).

Na criação do modelo de linguagem é desejável então encontrar estimativas ótimas para probabilidades condicionadas a cada contexto. A principal dificuldade em encontrar essas estimativas provém da esparsidade dos dados do treinamento. Uma vez que muitas palavras são nunca ou raramente observadas, suas estimativas não são confiáveis. Para um reconhecedor de voz, palavras que possuem probabilidade zero nunca serão reconhecidas nem que elas sejam acusticamente plausíveis. Isso é chamado de *problema da freqüência zero*. Existem muitas técnicas de suavização que buscam assegurar que todas as palavras, mesmo as que não apareçam no conjunto de treino, possuam probabilidade positiva.

Para melhor estabelecer os objetivos do presente trabalho, alguns conceitos importantes são descritos a seguir. Um texto T é uma coleção de sentenças e sua probabilidade $p(T)$ é o produto da probabilidade de sentenças individuais (assume-se independência estatística entre as sentenças). Para avaliar a qualidade de um modelo de linguagem em T, pode-se usar a entropia cruzada (também chamada *per-word coding length* ou *cross-entropy*)

$$H_p(T) \stackrel{\text{def}}{=} \frac{1}{W_T} \log_2 \left(\frac{1}{p(T)} \right),$$

onde W_T denota o número de palavras em T. Note que se uma probabilidade zero é atribuída a uma palavra que aparece no texto, $H_p(T)$ é infinita. A partir de $H_p(T)$, pode-se definir a *perplexidade* como

$$\text{PP} \stackrel{\text{def}}{=} 2^{H_p(T)}.$$

A perplexidade pode ser entendida como o número médio de diferentes (e equiprováveis) palavras que podem seguir uma dada palavra, de acordo com o modelo de linguagem adotado. Por exemplo, $\text{PP} = 10$ em um SRV para dez dígitos (0 a 9). Para SRV da língua inglesa, com vocabulários de tamanho superior a 20.000 palavras, PP costuma variar entre 100 e 250. Para uma dada tarefa de reconhecimento de voz, objetiva-se encontrar modelos de linguagem que conduzam a baixas perplexidades e custo computacional reduzido.

Considerando-se o SRV como um todo, a medida mais comum de avaliação é a taxa de palavras erradas (WER, de *word error rate*). Pode-se avaliar modelos de linguagem mantendo-se o modelo acústico fixo, e observando-se como as diferentes técnicas impactam a WER. Contudo, essa estratégia possui um custo computacional alto, sendo

comum a utilização da perplexidade nos estágios iniciais do desenvolvimento de modelos de linguagem para SRV. Isso é justificado pela forte correlação entre WER e PP ou, equivalentemente, $H_p(T)$, como indica a expressão[1]

$$\text{WER} \approx -12.37 + 6.48 \log_2(\text{PP}) = -12.37 + 6.48 H_p(T).$$

Assim, o principal objetivo desse trabalho é a obtenção de bons modelos de linguagem para o português brasileiro, e a avaliação será feita através do decréscimo de PP ou $H_p(T)$.

Ressalta-se que há diversos grupos de pesquisa desenvolvendo SRV em universidades como UFSC [Seara et al, 2003], PUC-RJ [Santos and Alcaim, 2002], INATEL [Ynoguti and Violaro, 1999], e PUC-RS [Fagundes and Sanches, 2003], mas há relativamente poucos trabalhos publicados acerca de modelos de linguagem para SRV usando o português brasileiro [Pessoa et al., 1999a, Pessoa et al., 1999b].

Este artigo encontra-se organizado da seguinte forma. Na Seção 2 faz-se uma breve revisão dos mais importantes modelos de linguagem adotados em reconhecimento de voz. Essa revisão é fortemente baseada no trabalho de nossos colaboradores [Jevtic and Orlitsky, 2003]. Na Seção 3 são apresentados resultados de simulação para algumas das técnicas discutidas, comparando-as de acordo com a abordagem adotada em [Chen and Goodman, 1999]. Na Seção 4 são apresentadas as conclusões do trabalho.

2. Estimação dos Modelos de Linguagem N-grama

Entre as primeiras aproximações para o *problema da freqüência zero* encontra-se a *suavização aditiva*, que remonta da época de Laplace [de Laplace, 1816]. Dado um conjunto de símbolos V, denotamos $c(v)$ o número de vezes que o símbolo $v \in V$ foi gerado. Esses estimadores atribuem para cada símbolo $w \in V$ a probabilidade

$$p_{add}(w) \stackrel{\text{def}}{=} \frac{c(w) + \delta}{\sum_{v \in V} (c(v) + \delta)}.$$

Essa equação é conhecida como lei de sucessão de Laplace (veja, e.g., [Jeffreys, 1939, Witten and Bell, 1991]). A regra "add-one" usa a lei de Laplace de sucessão com $\delta = 1$ para estimar a probabilidade da próxima palavra. Este foi um dos primeiros métodos empregados na modelagem da linguagem, mas em [Gale and Church, 1994], os autores mostraram experimentalmente que a mesma tem uma baixa performance. Um método que supera a regra "add-one"na tarefa de modelagem da linguagem é a regra "add-small-delta". Nesse caso, usa-se um subconjunto dos dados de treino (chamada *validação*) para encontrar o δ que maximiza a probabilidade desse subconjunto.

Em geral, as regras "add-small-delta"e "add-one"são eficientes quando todas as probabilidades são diferentes de zero, o que não é o caso em modelagem da linguagem para reconhecimento de voz. Para driblar a esparsidade dos dados, a maioria dos modelos populares de linguagem usa o conceito de *back-off*. Ao invés de remanejar a probabilidade, a distribuição do contexto mais amplo é usada, pois os mesmos possuem estimativas

[1]Obtida por W. Fisher a partir do estudo de diversos SRV, e divulgada em reunião organizada pelo NIST / EUA em 2000. Veja *http://www.isip.msstate.edu/publications/courses/ece_8463/*.

mais robustas. Por exemplo, de um contexto com as $n-1$ palavras mais recentes, recorre-se a um contexto com as $n-2$ palavras mais recentes. A recursão poderia finalizar em uma distribuição uniforme.

Chen e Goodman [Chen and Goodman, 1999] distinguem duas implementações de back-off, a *estrita* e a *interpolada*, e concluem que a interpolada leva a melhores resultados do que a estrita. Assim, neste artigo considera-se somente a variante interpolada:

$$p(w_i|w_{i-n+1}^{i-1}) = \overline{\lambda} \cdot p_0(w_i|w_{i-n+1}^{i-1}) + \lambda \cdot p(w_i|w_{i-n+2}^{i-1}),$$

onde λ é o parâmetro de interpolação e p_0 a estimativa inicial para a probabilidade desejada. Há vários métodos para balanceamento da distribuição do contexto total e seu back-off. A seguir, nós apresentamos alguns dos mais populares. Mais detalhes podem ser encontrados em [Chen and Goodman, 1999].

2.1. Modelo de Jelinek-Mercer

Jelinek e Mercer [Jelinek and Mercer, 1980] descreveram uma classe geral de modelos N-grama que interpolam diferentes cadeias de Markov:

$$p(w_i|w_{i-n+1}^{i-1}) = \sum_{j=0}^{n-1} \lambda_j \cdot p_{ML}(w_i|w_{i-j}^{i-1}),$$

onde $\sum_{j=0}^{n-1} \lambda_j = 1$, $p_{ML}(w_i|w_{i-n+1}^{i-1}) = \frac{c(w_{i-n+1}^{i})}{c(w_{i-n+1}^{i-1})}$ e $c(w_{i-n+1}^{i})$ representa quantas vezes w_{i-n+1}^{i} foi observada durante o treino. Os dados para treinamento são divididos em dois subconjuntos disjuntos: *treino* e *validação*. A estimativa de máxima verossimilhança (MLE, de *maximum likelihood estimation*) dos dados do conjunto *treino* é usada para obter as probabilidades p_{ML} de cada nível, e os parâmetros λ de interpolação são otimizados para maximizar a probabilidade do conjunto *validação*. A performance de suavização de Jelinek-Mercer é relativamente fraca quando se usa o mesmo λ para todos os contextos, mas inviável caso se adote um λ diferente para cada contexto. Uma solução de compromisso particularmente útil é a interpolação de forma *hierárquica* [Brown et al., 1992]:

$$\begin{aligned} p_{interp}(w_i|w_{i-n+1}^{i-1}) = \quad & \lambda_{w_{i-n+1}^{i-1}} p_{ML}(w_i|w_{i-n+1}^{i-1}) + \\ & (1 - \lambda_{w_{i-n+1}^{i-1}}) p_{interp}(w_i|w_{i-n+2}^{i-1}). \end{aligned}$$

A definição hierárquica permite agrupar λ's para vários contextos similares, separadamente a cada nível. Após isso, a mesma estima conjuntamente os valores ótimos através do algoritmo "expectation-maximization" (EM). O critério original usado para agrupamento em [Brown et al., 1992] foi o número de vezes que o contexto foi observado (contagem total). Assumia-se que um contexto que ocorre um grande número de vezes conduz a uma estimativa mais confiável. O parâmetro crítico que deve ser escolhido é o número dos contextos que serão agrupados, ou o número de parâmetros de interpolação livres que devem ser estimados. Ressalta-se contudo, que estes números dependem do tamanho dos dados do treinamento. Chen mostra em sua tese [Chen, 1996] que é melhor usar a contagem média da palavra como um critério para se aglomerar. Este critério é também muito menos sensível à mudança do tamanho do conjunto, comparado ao critério da contagem total.

2.2. Desconto Linear e Absoluto

Ney, Essen e Kneser [Ney and Essen, 1991, Ney et al., 1994] discutiram que todas as palavras em contextos mais longos são superamostradas ("oversampled") e que há duas maneiras gerais de descontar (ou de reduzir suas probabilidades) para compartilhar a probabilidade com as palavras não observadas do back-off: *desconto linear* e *absoluto*. No desconto linear a MLE dos contextos são descontadas proporcionalmente às probabilidades (escaladas) e a probabilidade descontada total é dada ao back-off (isso corresponde à suavização de Jelinek-Mercer com um único conjunto).

No desconto absoluto, todas as palavras são descontadas por uma constante aditiva igual:

$$p_{abs}(w_i|w_{i-n+1}^{i-1}) = \begin{cases} \frac{c(w_{i-n+1}^{i})-D}{c(w_{i-n+1}^{i-1})}, & \text{se } c(w_{i-n+1}^{i}) > 0 \\ \frac{D \cdot N_{1+}(w_{i-n+1}^{i-1}\cdot)}{c(w_{i-n+1}^{i-1})} p_{abs}(w_i|w_{i-n+2}^{i-1}), & \text{senão} \end{cases} \quad (1)$$

Assume-se que $0 < D < 1$ e $N_{1+}(w_{i-n+1}^{i-1}\cdot)$ é o número de palavras diferentes que foram observadas uma ou mais vezes seguindo w_{i-n+1}^{i-1}. Os autores mostrarm que o desconto absoluto (Equação 1) tem um desempenho melhor do que o desconto linear. Entretanto, quando agrupamentos de contextos são usados para o desconto linear, o desempenho é semelhante.

2.3. Modelo Kneser-Ney

Kneser e Ney [Kneser and Ney, 1995] aperfeiçoaram o modelo de desconto absoluto, impondo uma restrição à distribuição do back-off, forçando distribuições de ordem mais alta a terem as mesmas marginais dos dados de treinamento

$$\sum_{w_{i-n+1}} p_{KN}(w_{i-n+1}^{i}) = \frac{c(w_{i-n+2}^{i})}{N}.$$

De acordo com o modelo, a distribuição back-off é proporcional não ao número de vezes que a palavra foi observada no contexto, mas sim ao número de diferentes contextos nos quais foi observada

$$p_{KN}(w_i|w_{i-n+2}^{i-1}) = \frac{N_{1+}(\cdot w_{i-n+2}^{i})}{N_{1+}(\cdot w_{i-n+2}^{i-1}\cdot)}.$$

Isto produziu uma grande melhoria de desempenho.

2.4. Variação do Modelo de Kneser-Ney

Ney *et al.* [Ney et al., 1997] sugeriram uma variação do desconto absoluto que basicamente usa dois descontos: D_1 para os símbolos observados uma vez, e D_{2+} para aqueles observados duas ou mais vezes. Chen e Goodman em [Chen and Goodman, 1999] mostraram que três constantes D_1, D_2 e D_{3+} tem desempenho consistentemente superior. Ressaltamos então que o modelo de linguagem de Kneser-Ney com três parâmetros de descontos é muitas vezes considerado o melhor algoritmo para estimar um modelo N-grama.

2.5. Modelo Aditivo Interpolado

Em [Jevtic and Orlitsky, 2003], foi proposto o modelo *aditivo interpolado* (AI). Para cada contexto w_{i-n+1}^{i-1} de palavras observadas nos dados de treino, usa-se uma constante aditiva $\delta \in (-1, +\infty)$ para suavizar a distribuição.

$$p(w_i|w_{i-n+1}^{i-1}) = \begin{cases} \frac{c(w_{i-n+1}^{i})+\delta}{c(w_{i-n+1}^{i-1})+N_{1+}(w_{i-n+1}^{i-1}\cdot)\delta}, & c(w_{i-n+1}^{i}) > 0 \\ 0, & c(w_{i-n+1}^{i}) = 0. \end{cases}$$

Já para as palavras não observadas nos dados de treino, usa-se a aproximação interpolada de Jelinek-Mercer:

$$\begin{aligned} p_{opt}(w_i|w_{i-n+1}^{i-1}) = & \quad \lambda_{w_{i-n+1}^{i-1}} p(w_i|w_{i-n+1}^{i-1}) \\ & +(1-\lambda_{w_{i-n+1}^{i-1}}) p_{opt}(w_i|w_{i-n+2}^{i-1}). \end{aligned}$$

O final da recursão é uma distribuição uniforme. Esta formulação permite usar o algoritmo EM para estimar λ's em níveis diferentes, da mesma maneira usada no modelo de Jelinek-Mercer.

3. Resultados

Nesta seção são apresentados os resultados de simulações. Foi utilizado um corpus[2] do português brasileiro constituído majoritariamente por textos de um jornal e formatado usando XML. O corpus tem aproximadamente 30 milhões de linhas, das quais foram retirados os tags XML. A pontuação foi substituída por tags especiais, tais como <EXCLAMAÇÃO>, <VÍRGULA>, etc. Os dados foram separados em três conjuntos disjuntos para treino, validação e teste. Tanto o conjunto de *validação* quanto o de teste foram mantidos em 2500 sentenças.

Na Figura 1 encontram-se os resultados de um experimento preliminar usando bigramas e trigramas estimadas através do método "default" do software HTK (*http://htk.eng.cam.ac.uk/*). Esses resultados são compatíveis com os obtidos para a língua inglesa em simulações semelhantes, onde a perplexidade situa-se em torno de 100 a 250.

No intuito de aperfeiçoar os modelos obtidos com o HTK, lançamos mão do software desenvolvido por Nikola Jevtic (também usado em [Jevtic and Orlitsky, 2003]). De forma similar à metodologia em [Chen and Goodman, 1999], comparamos as técnicas mais populares em função do aumento no tamanho da seqüência de treino. Para os métodos que requerem "clustering" dos parâmetros de interpolação (Jelinek-Mercer e AI), os modelos foram construídos para diversos tamanhos de cluster e foi escolhido o que melhor se adapta a um segundo conjunto *validação* (também de 2500 sentenças).

Os resultados das novas simulações com trigramas são mostrados na Figura 2. Seguindo o formato adotado em [Chen and Goodman, 1999, Jevtic and Orlitsky, 2003], todos os gráficos mostram a diferença relativa na entropia $H_p(T)$ quando o método é comparado com o resultado mostrado Figura 1 (obtido com o HTK). Pode-se observar

[2]Gentilmente fornecido pelo Professor Ticiano Monteiro do CESUPA-PA.

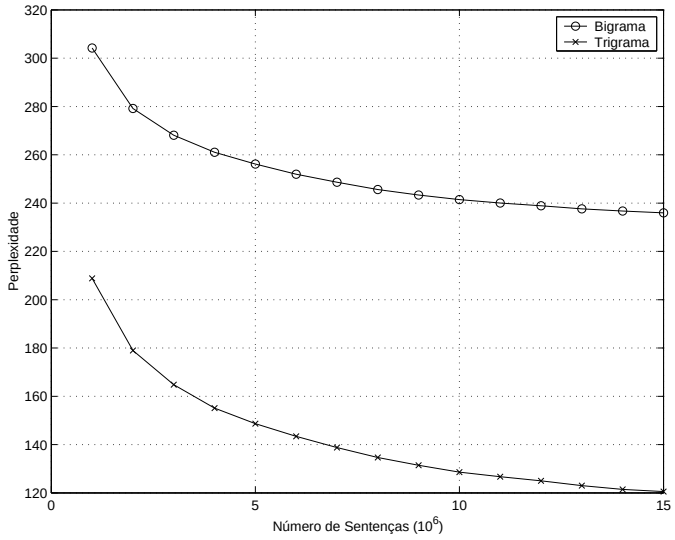

Figura 1: Evolução da perplexidade PP com o aumento dos dados para treino, para bigramas e trigramas.

uma melhoria de desempenho, com exceção do método de desconto absoluto. Verifica-se também que o método AI apresenta desempenho bem próximo ao do Kneser-Ney modificado. Ressalta-se contudo, que o AI apresenta melhor escalabilidade quando se aumenta a duração do contexto (ou seja, usa-se n-gramas com maior n), de acordo com [Jevtic and Orlitsky, 2003].

4. Conclusões

Este trabalho apresentou um breve sumário das técnicas do estado-da-arte em modelagem de linguagem, dentre as quais o modelo AI, recentemente proposto. Apresentou-se também, de forma comparativa, os resultados obtidos por várias das técnicas mais importantes quando aplicadas a um corpus de português brasileiro. Foi constatado que o modelo AI atinge bons resultados, com um custo computacional relativamente baixo quando comparado a métodos de desempenho similar.

Como todo sistema *data-driven*, o reconhecimento de voz se beneficia da disponibilidade de corpora com grande volume de dados. Existe uma quantidade razoável de textos para estudos de modelagem de linguagem para a língua inglesa, português europeu e outras (vide catálogo do LDC em *http://www.ldc.upenn.edu/*). Todavia, há poucos recursos acessíveis quando se trata do português brasileiro. Essa lacuna é ainda maior quando se trata de voz digitalizada para treinamento do modelo acústico. A inexistência dessas bases de dados não só atrasa as pesquisas em reconhecimento de voz e áreas correlatas, mas também impede que os resultados obtidos por diferentes grupos de pesquisa sejam comparados diretamente.

Futuros desenvolvimentos desse trabalho incluem o aumento da base de dados, melhoria dos algoritmos de estimação de n-gramas e uma ampla comparação entre os algoritmos no tocante à perplexidade, WER e custo computacional.

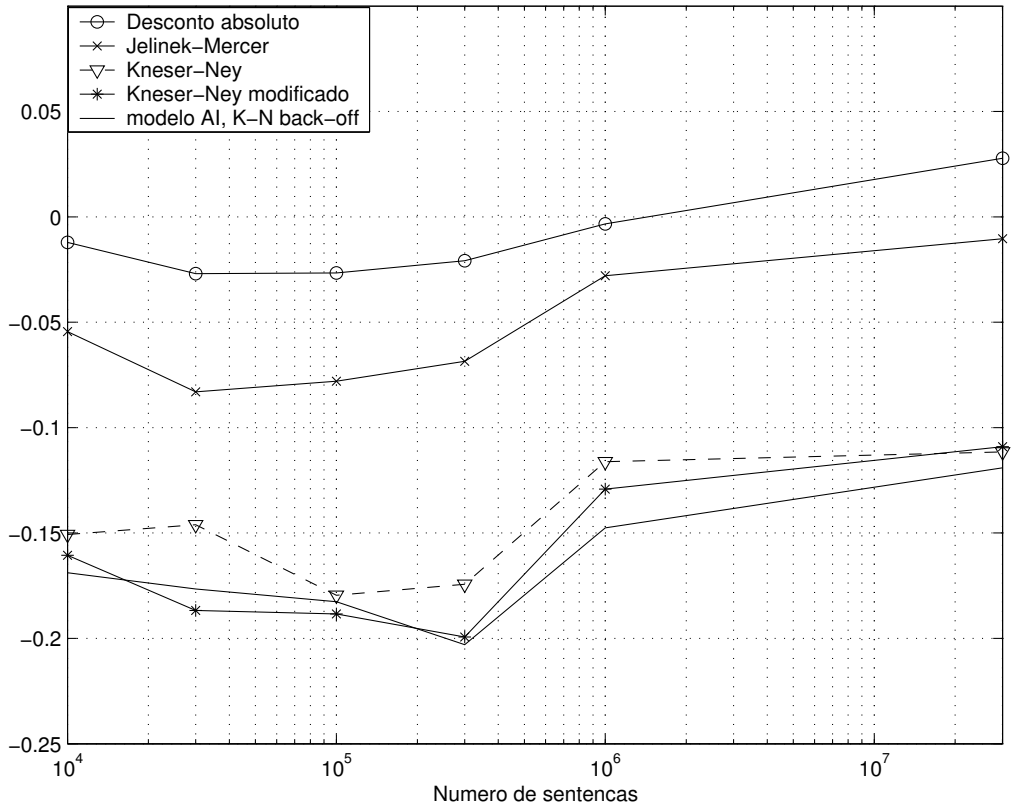

Figura 2: Resultados para trigramas: diferença relativa na entropia $H_p(T)$ em relação ao algoritmo do HTK usado para gerar a Figura 1. Quanto mais negativo o gráfico (menor entropia), melhor o resultado.

Referências

Brown, P. F., Pietra, S. A. D., Pietra, V. J. D., Lai, J. C., and Mercer, R. L. (1992). An estimate of an upper bound for the entropy of english. *Computational Linguistics*, 18:31–40.

Chen, S. F. (1996). *Building Probabilistic Models for Natural Language*. PhD Thesis.

Chen, S. F. and Goodman, J. (1999). An empirical study of smoothing techniques for language modeling. *Computer Speech and Language*, 13:359–394.

de Laplace, P. S. (1816). *Essay Philosophique sur la Probabilités*. Courcier Imprimeur, Paris.

Fagundes, R. and Sanches, I. (2003). Uma nova abordagem fonético-fonológica em sistemas de reconhecimento de fala espontânea. *Revista da Sociedade Brasileira de Telecomunicações*, 95.

Gale, W. A. and Church, K. W. (1994). What's wrong with adding one. *Corpus-Based Research Into Language (Oosdijk, N. and de Haan, P., eds)*.

Huang, X., Acero, A., and Hon, H.-W. (2001). *Spoken language processing*. Prentice-Hall.

Jeffreys, H. (1939). *Theory of Probability*. Clarendon, Oxford.

Jelinek, F. and Mercer, R. L. (1980). Interpolated estimation of markov source parameters from sparse data. *Proceedings of the Workshop on Pattern Recognition in Practice*, pages 381–397.

Jevtic, N. and Orlitsky, A. (2003). On the relation between additive smoothing and universal coding. *IEEE ASRU*.

Kneser, R. and Ney, H. (1995). Improved backing-off for m-gram language modeling. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1:181–184.

Ney, H. and Essen, U. (1991). On smoothing techniques for bigram-based natural language modeling. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2:825–829.

Ney, H., Essen, U., and Kneser, R. (1994). On structuring probabilistic dependences in stochastic language modeling. *Computer Speech and Language*, 8:1–38.

Ney, H., Martin, S., and Wessel, F. (1997). Statistical language modeling using leaving-one-out. *In Corpus Based Methods in Language and Speech Processing*, pages 174–207.

Pessoa, L., Violaro, F., and Barbosa, P. (1999a). Modelo de língua baseado em gramática gerativa aplicado ao reconhecimento de fala contínua. In *XVII Simpósio Brasileiro de Telecomunicações*, pages 455–458.

Pessoa, L., Violaro, F., and Barbosa, P. (1999b). Modelos da língua baseados em classes de palavras para sistema de reconhecimento de fala contínua. *Revista da Sociedade Brasileira de Telecomunicações*, 14(2):75–84.

Santos, S. and Alcaim, A. (2002). Um sistema de reconhecimento de voz contínua dependente da tarefa em língua portuguesa. *Revista da Sociedade Brasileira de Telecomunicações*, 17(2):135–147.

Seara et al, I. (2003). Geração automática de variantes de léxicos do português brasileiro para sistemas de reconhecimento de fala. In *XX Simpósio Brasileiro de Telecomunicações*, pages v.1. p.1–6.

Witten, I. H. and Bell, T. C. (1991). The zero frequency problem: Estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory*, 37(4):1085–94.

Ynoguti, C. A. and Violaro, F. (1999). Influência da transcrição fonética no desempenho de sistemas de reconhecimento de fala contínua. In *XVII Simpósio Brasileiro de Telecomunicações*, pages 449–454.