P.O.Box 8718, Beijing 100080, China	Journal of Software, Feb. 2007,18(2):196-204
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2007 by *Journal of Software*

语言建模中最小化样本风险算法的研究和改进

袁 伟, 高剑峰, 步丰林

Full-Text PDF Submission Back

袁 伟1, 高剑峰2, 步丰林1

1(上海交通大学 计算机科学与工程系,上海 200230)

2(Natural Language Processing Group, Microsoft Research, Redmond 98052,USA)

作者简介: 袁伟(1981－),男,硕士,主要研究领域为自然语言处理,信息检索.高剑峰(1971－),男,博士,研究员,主要研究领域为自然语言处理,信息检索,机器 翻译.步丰林(1961－),男,副教授,主要研究领域为软件工程,自然语言处理.

联系人: 袁 伟 Phn: +86-21-64836410, E-mail: weiyuan1981@gmail.com

Received 2006-01-04; Accepted 2006-06-12

Abstract

Most existing discriminative training methods adopt smooth loss functions that could be optimized directly. In natural language processing (NLP), however, many applications adopt evaluation metrics taking a form as a step function, such as character error rate (CER). To address the problem, a newly-proposed discriminative training method is analyzed, which is called minimum sample risk (MSR). Unlike other discriminative methods, MSR directly takes a step function as its loss function. MSR is firstly analyzed and improved in time/space complexity. Then an improved version MSR-II is proposed, which makes the computation of interference in the step of feature selection more stable. In addition, experiments on domain adaptation are conducted to investigate the robustness of MSR-II. Evaluations on the task of Japanese text input show that: (1) MSR/MSR-II significantly outperforms a traditional trigram model, reducing CER by 20.9%; (2) MSR/MSR-II is comparable to the other two state-of-the-art discriminative algorithms, Boosting and Perceptron; (3) MSR-II outperforms MSR not only in time/space complexity but also in the stability of feature selection; (4) Experimental results of domain adaptation show the robustness of MSR-II. In all, MSR/MSR-II is a quite effective algorithm. Given its step loss function, MSR/MSR-II could be widely applied to many fields of NLP, such as spelling check and machine translation.

Yuan W, Gao JF, Bu FL. A study and improvement of minimum sample risk methods for language modeling. *Journal of Software*, 2007,18(2):196-204.

DOI: 10.1360/jos180196

http://www.jos.org.cn/1000-9825/18/196.htm

摘要

目前,一些主流的判别学习算法只能优化光滑可导的损失函数,但在自然语言处理(natural language processing,简称NLP)中,很多应用的直接评价标准(如字符转换错误数(character error rate,简称CER))都是不可导的阶梯形函数.为解决此问题,研究了一种新提出的判别学习算法——最小化样本风险(minimum sample risk,简称MSR)算法.与其他判别训练算法不同,MSR算法直接使用阶梯形函数作为其损失函数.首先,对MSR算法的时空复杂性作了分析和提高;同时,提出了改进的算法MSR-II,使得特征之间相关性的计算更加稳定.此外,还通过大量领域适应性建模实验来考察MSR-II的鲁棒性.日文汉字输入实验的评测结果表明:(1) MSR/MSR-II显著优于传统三元模型,使错误率下降了20.9%;(2) MSR/MSR-II与另两类主流判别学习算法Boosting和Perceptron表现相当;(3) MSR-II不仅在时空复杂度上优于MSR,特征选择的稳定性也更高;(4) 领域适应性建模的结果证明了MSR-II的良好鲁棒性.总之,MSR/MSR-II是一种非常有效的算法.由于其使用的是阶梯形的损失函数,因此可以广泛应用于自然语言处理的各个领域,如拼写校正和机器翻译.

References:

[1] Jelinek F. Self-Organized language modeling for speech recognition. In: Waibel A, Lee KF, eds. Readings in Speech Recognition. San Mateo: Morgan-Kaufmann Publishers, 1990. 450-506.

[2] Brown PF, Cocke J, Pietra SAD, Pietra VJD, Jelinek F, Lafferty JD, Mercer RL, Roossin PS. A statistical approach to machine translation. Computational Linguistics, 1990,16(2):79-85.

[3] Gao JF, Suzuki H, Wen Y. Exploring headword dependency and predictive clustering for language modeling. In: Hajic J, Matsumoto Y, eds. Proc. of the Empirical Methods in Natural Language Processing (EMNLP). MACL, 2002. 248-256.

[4] Collins M. Discriminative reranking for natural language parsing. In: Langley P, ed. Proc. of the 17th Int'l Conf. on Machine Learning (ICML 2000). San Francisco: Morgan Kaufmann Publishers, 2000. 175-182.

[5] Collins M. Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In: Hajic J, Matsumoto Y, eds. Proc. of the Empirical Methods in Natural Language Processing (EMNLP). MACL, 2002. 1-8.

[6] Gao JF, Yu H, Yuan W, Xu P. Minimum sample risk methods for language modeling. In: Proc. of the Empirical Methods in Natural Language Processing (EMNLP). 2005. 209-216. http://research.microsoft.com/~jfgao/

[7] Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd ed., Wiley-Interscience, 2000. 117-120.

[8] Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical Recipes in C: The Art of Scientific Computing. 2nd ed., Cambridge: Cambridge University Press, 1992. 412-419.

[9] Quirk C, Menezes A, Cherry C. Dependency tree translation: Syntactically informed phrasal SMT. In: Proc. of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL). 2005. 271-279. http://www.cs.ualberta.ca/~colinc/papers/ ms_acl05.pdf

[10] Och FJ. Minimum error rate training in statistical machine translation. In: Proc. of the 41st Annual Meeting of the Association for Computational Linguistics (ACL). 2003. 160-167. http://acl.ldc.upenn.edu/acl2003/main/pdfs/Och.pdf

[11] Theodoridis S, Koutroumbas K. Pattern Recognition. 2nd ed., Academic Press, 2003. 182-183.

[12] Yu H, Gao JF, Bu FL. One new discriminative training method for language modeling. Chinese Journal of Computers, 2005,28(10): 1708-1715 (in Chinese with English abstract).

附中文参考文献:

[12] 于浩,高剑峰,步丰林.一种新的语言模型判别训练方法.计算机学报,2005,28(10):1708-1715.