Major Symbols Used in the Book and Their Descriptions

Symbols	Descriptions
R	number of training samples (tokens or strings)
$r = 1, \ldots, R$	index of individual training samples
X	aggregate of all training samples
χ	a random variable in the space of X
S	aggregate of training reference labels
s	aggregate of hypothesis labels (reference or otherwise)
X_r	rth training sample. It may represent a sequence of observation vectors with a variable length: $X_r = x_{r,1}, x_{r,2}, \ldots, x_{r,T}$
χ_r	a random variable in the space of X_r
S_r	reference label of the rth training sample. It may represent a sequence of words with variable length: $S_r = w_{r,1}, w_{r,2}, \ldots, w_{r,M}$
s_r	hypothesis label of the rth training sample (reference or otherwise)
$x_{r,t}$	tth feature vector of the rth training observation sequence.
$\chi_{r,t}$	A random variable in the space of $x_{r,t}$.
$w_{r,i}$	ith word of the reference label of the rth training sample
q	hidden Markov model (HMM) state sequence
θ	natural parameters of the exponential family distributions
Λ	Aggregate of model parameter sets
Λ'	model parameter sets obtained from the immediately previous iteration in an iterative learning algorithm
$a_{i,j}$	HMM transition probability from state i to state j
$b_i(x)$	HMM emitting probability for observation x at state i
μ	mean vector of Gaussian distribution
Σ	covariance matrix of Gaussian distribution

Mathematical Notation

1. Superscript T denotes the transpose of a matrix or vector; for example, $\mathbf{x}^{\mathrm{T}}$ will be a row vector.
2. $(w_1, \ldots, w_M)$ denotes a row vector with M elements, and the corresponding column vector is denoted as $\mathbf{w} = (w_1, \ldots, w_M)^{\mathrm{T}}$.
3. $[a, b]$ denotes the closed interval from a to b (i.e., the interval including the values a and b themselves). (a, b) denotes the corresponding open interval (i.e., the interval excluding a and b). $[a, b)$ denotes an interval that includes a but excludes b.
4. The expectation of a function $f(x)$ with respect to a random variable x is denoted by $\mathbb{E}_{p(x|\lambda)}[f(x)]$ or $\mathbb{E}_x[f(x)]$, assuming that the distribution of x is described by $p(x|\lambda)$, where λ is the parameter set in this distribution. In situations where there is no ambiguity on which a variable is being averaged over, this will be simplified by removing the suffix.

Bibliography

[1] M. Afify, "Extended Baum–Welch re-estimation of Gaussian mixture models based on reverse Jensen's inequality," *Proc. INTERSPEECH*, 2005, 1113–1116.

[2] S. Amari, "A theory of adaptive pattern classifiers," *IEEE Trans. Electron. Comput.*, 16, 1967, 299–307.

[3] S. Axelrod, V. Goel, R. Gopinath, P. Olsen, and K. Visweswariah, "Discriminative estimation of subspace constrained Gaussian mixture models for speech recognition," *IEEE Trans. Audio, Speech Lang. Proc.*, 15, 2007, 172–189.

[4] L. Baum and G. Sell, "Growth transformations for functions on manifolds," *Pac. J. Math.*, 27(2), 1968, 211–227.

[5] E. Birney, "Hidden Markov models in biological sequence analysis," *IBM J. Res. Dev.*, 45, 2001, 449–454.

[6] P. Brown, The acoustic modeling problem in automatic speech recognition, Ph.D. thesis, Carnegie Mellon University, 1987.

[7] W. Chou and L. Li, "A minimum classification error framework for generalized linear classifier in machine learning and its application to text categorization/retrieval," *Proc. Int. Conf. Mach. Learning Appl.*, Dec. 2004, 382–390.

[8] W. Chou, "Minimum classification error approach in pattern recognition," in W. Chou and B.-H. Juang (eds.), *Pattern Recognition in Speech and Language Processing*, CRC Press, Boca Raton, 2003, pp. 1–49.

[9] M. Collins, "Discriminative training methods for hidden Markov models: theory and experiments with perceptron algorithms," *Proc. Empirical Methods of Natural Language Processing (EMNLP)*, 2005, 1–8.

[10] L. Deng and D. O'Shaughnessy, *Speech Processing — A Dynamic and Optimization-Oriented Approach*, Marcel Dekker Publishers, New York, NY, 2003.

[11] L. Deng, D. Yu, and A. Acero, "A generative modeling framework for structured hidden speech dynamics," Neural Information Processing System (NIPS) Workshop, Whistler, BC, Canada, Dec. 2005.

[12] L. Deng, J. Wu, J. Droppo, and A. Acero, "Analysis and comparison of two feature extraction/compensation algorithms," *IEEE Signal Process. Lett.*, 12(6), June, 2005, 477–480.

Q1 [13] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press, 1998.

[14] P. Gopalakrishnan, D. Kanevsky, A. Nadas, and D. Nahamoo, "An inequality for rational functions with applications to some statistical estimation problems," *IEEE Trans. Inf. Theory*, 37, January 1991, 107–113.

[15] Y. Gao and J. Kuo, "Maximum entropy direct models for speech recognition," *IEEE Trans. Speech Audio Process.*, 2006.

Q2 [16] V. Goel and W. Byrne, "Minimum Bayes risk methods in automatic speech recognition," in W. Chou and B.-H. Juang (eds.), *Pattern Recognition in Speech and Language Processing*, CRC Press, Boca Raton, 2003.

[17] A. Gunawardana and W. Byrne, "Discriminative speaker adaptation with conditional maximum likelihood linear regression," *Proc. EUROSPEECH*, 2001.

[18] A. Gunawardana, M. Mahajan, A. Acero, and J. Platt, "Hidden conditional random fields for phone classification," *Proc. Interspeech*, Lisbon, 2005.

[19] X. He and W. Chou, "Minimum classification error linear regression for acoustic model adaptation of continuous density HMMs," *Proc. ICASSP*, April 2003.

[20] X. He, L. Deng, and W. Chou, "A novel learning method for hidden Markov models in speech and audio processing," *Proc. IEEE Workshop on Multimedia Signal Processing*, Victoria, BC, October 2006.

Q3 [21] X. Huang, A. Acero and H. Hon, *Spoken Language Processing*, Prentice Hall, 2001.

[22] T. Jebara, Discriminative, Generative and imitative learning, PhD Thesis, Media Laboratory, MIT, December 2001.

Q4 [23] T. Jebara and A. Pentland, "On reversing Jensen's inequality," *Neural Inf. Process. Syst.*, 13, December 2000.

[24] B.-H. Juang and S. Katagiri, "Discriminative learning for minimum error classification, *IEEE Trans. Signal Process.*, 40(12), 1992, 3043–3054.

Q5 [25] B.-H. Juang, W. Chou, and C.-H. Lee, "Minimum classification error rate methods for speech recognition," *IEEE Trans. Speech Audio Process*, 5, May 1997.

[26] D. Kanevsky, "Extended Baum transformations for general functions," *Proc. ICASSP*, 2004.

[27] D. Kanevsky, "A generalization of the Baum algorithm to functions on non-linear manifolds," *Proc. ICASSP*, 1995.

[28] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," *Proc. Int. Conf. Mach. Learning*, 2001, 282–289.

[29] Y. Li, L. Shapiro, and J. Bilmes, "A generative/discriminative learning algorithm for image classification," *IEEE Int. Conf. Computer Vision*, Beijing, China, 2005, 1605–1612.

[30] A. McCallum, D. Freitag, and F. Pereira, "Maximum entropy Markov models for information extraction and segmentation." *Proc. Int. Conf. Mach. Learning*, 2000, 591–598.

[31] W. Macherey, L. Haferkamp, R. Schlüter, and H. Ney, "Investigations on error minimizing training criteria for discriminative training in automatic speech recognition," *Proc. Interspeech*, 2005, Lisbon, 2133–2136.

[32] E. McDermott, "Discriminative training for automatic speech recognition using the minimum classification error framework," *Neural Inf. Process. Syst.* (NIPS) Workshop, Whistler, BC, Canada, Dec. 2005.

[33] E. McDermott, T. Hazen, J. L. Roux, A. Nakamura, and S. Katagiri, "Discriminative training for large vocabulary speech recognition using minimum classification error," *IEEE Trans. Speech Audio Process.*, 15(1), 2007, 203–223.

[34] Y. Normandin, Hidden Markov models, maximum mutual information estimation, and the speech recognition problem, PhD. dissertation, McGill University, Montreal, 1991.

[35] Y. Normandin, "Maximum mutual information estimation of hidden Markov models," in C.-H. Lee, F.K. Soong, and K.K. Paliwal (eds.), *Automatic Speech and Speaker Recognition*, Kluwer Academic Publishers, Norwell, MA, 1996. Q6

[36] F. Och, "Minimum error rate training in statistical machine translation," *Proc. 41st Meeting Assoc. Comp. Linguistics*, 2003, 160–167.

[37] F. Pereira, "Linear models for structure prediction," *Proc. Interspeech*, Lisbon, 2005, 717–720.

[38] D. Povey and P.C. Woodland, "Minimum phone error and I-smoothing for improved discriminative training," *Proc. ICASSP*, 2002.

[39] D. Povey, M.J.F. Gales, D.Y. Kim, and P.C. Woodland, "MMI-MAP and MPE-MAP for acoustic model adaptation," *Proc. Eurospeech*, 2003.

[40] D. Povey, Discriminative training for large vocabulary speech recognition, Ph.D. dissertation, Cambridge University, Cambridge, UK, 2004.

[41] D. Povey, B. Kingsbury, L. Mangu, G. Saon, H. Soltau, and G. Zweig, "fMPE: discriminatively trained features for speech recognition," *Proc. DARPA EARS RT-04 Workshop*, Nov. 7–10, 2004, Palisades, NY, Paper No. 35, 5 pp.

[42] C. Rathinavalu and L. Deng, "Speech trajectory discrimination using the minimum classification error learning," *IEEE Trans. Speech Audio Process.*, 6, 1998, 505–515.

[43] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993. Q7

[44] J. Roux and E. McDermott, "Optimization for discriminative training," *Proc. INTERSPEECH*, Lisbon, 2005.

[45] R. Schlüter, Investigations on discriminative training criteria, Ph.D. dissertation, RWTH Aachenm University of Technology, Aachen, Germany, 2000.

[46] R. Schlüter, W. Macherey, B. Muller, and H. Ney, "Comparison of discriminative training criteria and optimization methods for speech recognition," *Speech Commun.*, 34, 2001, 287–310.

Q8 [47] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Elsevier Science, Academic Press, 2003.

Q9 [48] V. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, 1998.

[49] V. Valtchev, Discriminative methods for HMM-based speech recognition, Ph.D. thesis, Cambridge University, England, 1995.

Q10 [50] V. Valtchev, J.J. Odell, P.C. Woodland, and S.J. Young, "MMIE training of large vocabulary recognition systems," *Speech Commun.*, 22(4), 1997.

[51] V. Valtchev, J.J., Odell, P.C., Woodland, and S.J. Young, "Lattice-based discriminative training for large vocabulary speech recognition," *Proc. ICASSP*, 1996.

[52] P. Woodland and D. Povey, "Large scale discriminative training for speech recognition," *Proc. ITRW ASR*, 2000.

[53] Y. Xiong, Q. Huo, C. Chan, "A discrete contextual stochastic model for the offline recognition of handwritten Chinese characters," *IEEE Trans. PAMI*, 23, 2001, 774–782.

[54] R. Yan, J. Zhang, J. Yang, and A. Hauptmann, "A discriminative learning framework with pairwise constraints for video object classification," *IEEE Trans. PAMI*, 28(4), 2006, 578–593.

[55] J. Yang, Y. Xu, and C.S. Chen, "Hidden Markov model approach to skill learning and its application in telerobotics, " *IEEE Trans. Robotics Autom.*, 10(5), 1994, 621–631.

[56] C. Yen, S. Kuo, and C.-H. Lee, "Minimum error rate training for HMM-based text recognition," *IEEE Trans. Image Process.*, 8, 1999, 1120–1124.

[57] D. Yu, L. Deng, and A. Acero, "A* lattice search algorithm for a long-contextual-span hidden trajectory model and phonetic recognition," *Proc. Interspeech*, Lisbon, 2005, 553–556.

[58] D. Yu, L. Deng, X. He, A. Acero, "Large-margin minimum classification error training for large-scale speech recognition tasks," *Proceedings of ICASSP*, Honolulu, Hawaii, April 2007.

Author Biography

Xiaodong He received his bachelor's degree from Tsinghua University, Beijing, China, in 1996, and earned his master's degree from the Chinese Academy of Sciences in 1999, and his doctoral degree from the University of Missouri–Columbia in 2003. He joined the Speech and Natural Language group of Microsoft in 2003, and the Natural Language Processing group of Microsoft Research, Redmond, WA, in 2006, where he currently serves as researcher. His research areas include statistical machine learning, automatic speech recognition, natural language processing, machine translation, signal processing, nonnative speech processing, and human–computer interaction. In these areas, he has authored/coauthored more than 30 refereed papers in leading international conferences and journals. He has filed more than 10 U.S. or international patents in the areas of speech recognition, language processing, and machine translation. He served as a reviewer for major conferences and journals in the areas of speech recognition, natural language processing, signal processing, and pattern recognition. He also served on program committees of various conferences in these areas. He is a member of ACL, IEEE, ISCA, and Sigma Xi.

Li Deng received his bachelor's degree from the University of Science and Technology of China and his Ph.D. degree from the University of Wisconsin–Madison. In 1989, he joined the Department of Electrical and Computer Engineering, University of Waterloo, Ontario, Canada, as assistant professor; he became tenured full professor in 1996. From 1992 to 1993, he conducted sabbatical research at the Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, MA, and from 1997 to 1998, at the ATR Interpreting Telecommunications Research Laboratories, Kyoto, Japan. During 1989–1999, he taught a wide range of electrical and computer engineering courses, both at undergraduate and graduate levels. In 1999, he joined Microsoft Research, Redmond, WA, as senior researcher; he currently serves as principal researcher for the same institution. He has also been affiliate professor in the Department of Electrical Engineering at University of Washington since 2000 after moving to Seattle. His past and current research areas include automatic speech and speaker recognition, statistical methods and machine learning, neural information processing, machine intelligence, audio and acoustic signal processing, statistical signal processing and digital communication, human speech production and perception, acoustic phonetics, auditory speech processing, noise robust speech processing, speech synthesis and enhancement, spoken language understanding systems, multimedia signal processing, and multimodal human–computer interaction. In these areas, he has published more than 300 refereed papers in leading international conferences and journals, and 14 book chapters, and has given keynotes, tutorials, and lectures worldwide. He has been granted more than 20 U.S. or international patents in acoustics, speech/language technology, and signal processing. He has likewise authored two recent books on speech processing.

He served on Education Committee (as a founding member) and Speech Processing Technical Committee of the IEEE Signal Processing Society 1996–2000, and was associate editor of *IEEE Trans. Speech and Audio Processing* for 2002–2005. He currently serves on the Society's Multimedia Signal Processing Technical Committee and on the Society's Board of Governors. He is editor-in-chief of the *IEEE Signal Processing Magazine* and associate editor of *IEEE Signal Processing Letters*. He was a technical chair of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP04), and the general chair of the 2006 IEEE Workshop on Multimedia Signal Processing. He will be general chair of the ICASSP-2013 in Vancouver, Canada. He is a Fellow of the Acoustical Society of America and a Fellow of the IEEE.

• • • •

Author Query Form

(Queries are to be answered by the Author)

He – BM

The following queries have arisen during the typesetting of your manuscript. Please answer these queries.

Query Marker	Query	Reply
Q1	Please provide the page range for this reference.	
Q2	Please provide the page range for this reference.	
Q3	Please provide the publisher location and page range for this reference.	
Q4	Please provide the page range for this reference.	
Q5	Please provide the page range for this reference.	
Q6	Please provide the page range for this reference.	
Q7	Please provide the publisher location and page range for this reference.	
Q8	Please provide the publisher location and page range for this reference.	
Q9	Please provide the publisher location and page range for this reference.	
Q10	Please provide the page range for this reference.	

Thank you very much.