5
Referências

1 Linde, Y., Buzo, A., Gray, R. M., "An algorithm for vector quantizer design", IEEE Trans. Communications (January), vol. COM-28, issue 1, pp. 84-95, 1980.

2 Pedreira, C. E., "Learning vector quantization with training data selection", IEEE Trans. Pattern Analysis and Machine Intelligence (January), vol. 28, issue 1, pp. 157-162, 2006.

3 Lehn-Schioler, T., Hedge, A., Erdogmus D., Principe, J. C., "Vector Quantization using Information Theoretic Concepts", Natural Computing, nº4, pp.39-51, 2005.

4 Cover, T.M., Thomas J.A., "Elements of Information Theory", Wiley Series in Telecommunications, 1991.

5 Haykin, S., "Neural Networks: A Comprehensive Foundation", Prentice-Hall, 1998.

6 Principe, J. C., Xu, D., Fisher, J., "Information Theoretic Learning", in Unsupervised Adaptive Filtering, S. Haykin, Ed. New York: Wiley, 2000.

7 Duda, R.O., Hart, P.E., Stork, G., "Pattern Recognition", 2nd. Ed., Wiley, 2001.

8 Jensen, R., "An Information Theoretic Approach to Machine Learning", Dissertation for the Degree of Doctor Scientiarum, Dept. of Physics, University of Tromso, 2005.

9 Li, L., Pratap, A., Lin, H.-T., Abu-Mostafa, Y. S., "Improving generalization by data categorization," in PKDD, LNAI 3721, Springer-Verlag, pp. 157-168, 2005.

10 Plutowski, M., White, H., "Selecting concise training sets from clean data," IEEE Trans. Neural Networks, vol. 4, issue 2, pp. 305-318, March 1993.

11 Hasenjäger, M., Ritter, H., Obermayer, K., Kohonen Maps. In E.Oja & S. Kaski editors, "Active learning in self-organizing maps," pp. 57-70, Elservier, 1999.

12 D. Xu. Energy, Entropy and Information Potential for Neural Computation. PhD thesis, University of Florida, Gainesville, FL, USA, 1999.

13 Yen, C-W., Young, C-N, Nagurka, M.L., "A vector quantization method for nearest neighbor classifier design", Pattern Recognition Letters, 25, pp. 725-731, 2004.

14 Pedreira, C. E., Macrini, L., Costa, E. S., "Input and data selection applied to heart disease diagnosis", Proc. IEEE-INNS-ENNS International Joint Conference on Neural Networks, Montreal, 2005.

15 Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., Guppy, K., Lee, S. e Froelicher, V., "International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease", Am. J. Cardiology, pp. 304-310, 1989.

16 Gray, R.M., Neuhoff, D.L., "Quantization", IEEE Trans. Information Theory (October), vol. 44, issue 6, pp. 2325-2383, 1998.

17 Gersho, A., Gray, R.M., "Vector Quantization and Signal Compression", Kluwer Academic Publishers, 1992.

18 Zhang, G.P., "Neural Networks for Classification: A Survey". IEEE Trans. Systems, Man, and Cybernetics – Part C: Applications and Reviews, vol. 30, nº 4, pp. 451-462, November, 2000.

19 Morejon, R. A., Principe, J.C., "Advanced Search Algorithms for Information Theoretic Learning with Kernel-Based Estimators", IEEE Transactions Neural Networks, vol.15, issue 4, pp.874–884, 2004.

20 Kohonen, T., "Self-Organizing Maps", third ed. Springer, 2001.

21 Lloyd, S.P., "Least Squares Quantization in PCM", Bell Laboratories, Technical Note, 1957.

22 Nakagaki, R., Katsaggelos, A.K., "A VQ-Based Blind Image Restoration Algorithm", IEEE Trans. Image Processing (September), vol. 12, issue 9, pp. 1044-1053, 2003.

23 Lopes, W.T.A., Madeiro, F., Neto, B.G.A., Alencar, M.S., "Combining Modulation Diversity and Index Assignment to Improve Image VQ for a Rayleigh Fading Channel", Learning and Nonlinear Models, vol.1, nº3, pp.168-179, 2004.

24 Xie, Q., Laszlo, C.A., Ward, R.K., "Vector Quantization Technique for Nonparametric Classifier Design", IEEE Trans. on Pattern Analysis and Machine Intelligence (December), vol. 15, issue 12, pp. 1326-1330, 1993.

25 Soong, F.K., Rownberg, A.E., Rabiner, L.R., Juang, B.H., "A vector quantization approach to speaker recognition," AT&T Tech. J. vol. 66, pp. 14-26, 1987.

26 Wu, X., Guan, L., "Acceleration of the LBG Algorithm", IEEE Trans. Communications (February/March/April), vol. 42, issue 2/3/4, pp. 1518-1523, 1994.

27 Fritzke, B., "The LBG-U Method for Vector Quantization – an Improvement over LBG Inspired from Neural Networks", Neural Processing Letters, nº 5, pp. 35-45, 1997.

28 Gokcay, E., Principe, J.C., "Information Theoretic Clustering", IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.24, no. 2, pp. 158-171, February 2002.

29 Erdogmus, D., Principe, J.C., "An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems," IEEE Trans. Signal Process., vol. 50, no. 7, pp. 1780–1786, Jul. 2002.

30 Jain, A.K., Duin, R., Mao, J., "Statistical Pattern Recognition: A Review", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, pp, 4-37, Jan. 2000.

31 Xu, R., Wunsch II, D., "Survey of Clustering Algorithms", IEEE Trans. Neural Networks, vol. 16, no. 3, pp. 645-678, May. 2005.

32 Bishop, C.M., "Pattern Recognition and Machine Learning", Springer, 2006.

33 Vapnik, V. N., "Statistical Learning Theory", New York: Wiley, 1998.

34 Müller, K.R., Mika, S., Rätsch, G., Tsuda, K., Schölkopf, B., "An Introduction to kernel based learning algorithms", IEEE Transactions on Neural Networks, vol. 12, no. 2., pp.181-202, 2001.

35 Mitra, P., Pal S.K., "A probabilistic active support vector learning algorithm", IEEE Trans. Pattern Analysis and Machine Intelligence (March), vol. 26, issue 3, pp. 413-418, 2004.

36 Li, M., Sethi, I. K., "Confidence-Based Active Learning", IEEE Trans. Pattern Analysis and Machine Intelligence (August), vol. 28, issue 8, pp. 1251-1261, 2006.

37 Hwang, J. N., Choi, J. J., Oh, S., Marks II, R. J., "Query-based learning applied to partially trained multi-layer perceptrons", IEEE Trans. Neural Networks, vol.2, issue 1, pp.131-136, January, 1991.

38 Faraway, J. J., "Sequential design for the nonparametric regression of curves and surfaces", in Proceedings of the 22nd Symposium on the Interface between Computing Science and Statistics, Springer, pp. 104-110, 1990.

39 Burges, C. J. C., "A tutorial on support vector machines for pattern recognition", Knowledge Discovery and Data Mining, vol. 2, no. 2, pp.121–167, 1998.

40 Johnson, R.A. and Wichern, D.W., "Applied Multivariable Statistical Analysis", Prentice Hall, 1998.

41 Silverman, B.W., "Density Estimation for Statistics and Data Analysis", Monographs on Statistics and Applied Probability 26, Chapman & Hall/CRC, 1986.

42 Cover, T., Hart, P., "Nearest Neighbor Pattern Classification", IEEE Trans. Information Theory, vol.13, issue 1, pp.21-27, January, 1967.

43 Vinga, S., Almeida, J., "Rényi continuous entropy of DNA sequences", J. Theor. Biol., vol. 231 (3), pp. 377 – 388, 2004.

44 Fariñas, M., Pedreira, C.E., Medeiros, M.C., "Local-Global Neural Networks: A New Approach For Nonlinear Time Series Modelling", Journal of the American Statistical Association, vol. 468, p. 1092-1107, 2004.

45 Pedreira, C.E., Farinas, M., Pedroza, L.C., "Redes Neurais Locais-Globais Uma Aplicação ao Problema de Dados Faltantes", Learning and nonlinear models, 2002.

46 Peres, R. T., Pedreira, C.E., "Seleção de Dados para LVQ através de Aprendizado Exaustivo", VII Congresso Brasileiro de Redes Neurais, 2005, Natal. Anais, 2005.

47 Peres, R. T., Pedreira, C.E., “Preliminary Results on Noise Detection and Data Selection for Vector Quantization”, IEEE World Congress on Computational Intelligence, 2006, Vancouver. Proc., 2006.
48 Hastie, T., Tibshirani, R., Friedman, J., “The Elements of Statistical Learning”, Springer Series in Statistics, 2001.
49 Theodoridis, S., Koutroumbas, K., “Pattern Recognition”, Third Edition, Elsevier, 2006.
50 Bishop, C.M., “Neural Networks for Pattern Recognition”, Oxford, 1995.
51 G. Tutz, H. Binder. “Localized Classification”, Statistics and Computing, 15, pp. 155-166, 2005.
52 Ripley, B.D., “Pattern Recognition and Neural Networks”, Cambridge University Press, 1996.
53 Huang, K., Yang, H., King, I., Lyu, M.R., Chan, L., “The minimum error minimax probability machine”, Journal of Machine Learning Research, 5, pp. 1253-1286, 2004.
54 Fawcett, T., “An introduction to ROC analysis”. Pattern Recognition Letters, vol. 27, pp. 861-874, 2006.
55 Toussaint, G. T. "Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining", International Journal of Computational Geometry and Applications, vol. 15, no. 2, pp. 101-150, 2005.
56 Aha, D.W., Kibler, D., Albert, M., “Instance-based learning algorithms”, Machine Learning, 6, pp. 37-66, 1991.
57 Wilson, D. R., Martinez, T. R., “Reduction Techniques for instance-based learning algorithms”, Machine Learning, 38, pp. 257-286, 2000.
58 Hong, Z. Q., Yang, J. Y., "Optimal Discriminant Plane for a Small Number of Samples and Design Method of Classifier on the Plane", Pattern Recognition, vol. 24, no. 4, pp. 317-324, 1991.
59 Breiman, L., Friedman, J. H., Olshen, A. R., Stone, J. C., "Classification and Regression Trees." 1984.
60 Bouguila, N., Ziou, D., “A hybrid SEM algorithm for high-dimensional unsupervised learning using a finite generalized Dirichlet mixture”, IEEE Trans. on Image Processing, vol. 15, issue 9, pp. 2657-2668, 2006.
61 Qin, A. K., Suganthan, P. N., “Initialization insensitive LVQ algorithm based on cost-function adaptation”, Pattern Recognition, 38, 773-776, 2005.
62 MacKay D. J. C., “Bayesian interpolation”, Neural Computation, 4:415-447, 1992.
63 Banerjee, A., Merugu, S., Dhillon, I., Ghosh, J., “Clustering with Bregman distances”, J. Machine Learning Research, 1705-1749, 2005.
64 Peres, R.T., Pedreira, C.E., “Generalized Risk Zone: Selecting Observations for Classification”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Aceito no prelo, 2009.

6
Apêndice A – Notação

X – conjunto de observações;

Y – rótulos das observações;

n – dimensão do espaço;

f – função que estabelece a classificação, também chamada de classificador;

C_1 – Classe 1;

C_2 – Classe 2;

$P(C_i)$ – probabilidades a priori de C_i, i = 1,2;

$p(x \mid C_i)$ – funções de densidade de probabilidade condicionais a classe C_i, i = 1,2;

$P(C_i \mid x)$ – probabilidade posteriori da classe C_i, i = 1,2.

7
Apêndice B – Quantização Vetorial

O termo quantização pode ser definido como a ação de fragmentar uma informação contínua em partes discretas. Se X é uma variável aleatória contínua, então o quantizador é dito escalar. No caso de tratar-se de um vetor aleatório, tem-se um quantizador vetorial. A maior parte da análise matemática destes métodos veio da comunidade de processamento de sinais [7], entretanto algoritmos de quantização vetorial podem ser considerados dentro do contexto estatístico de clusterização. Para uma perspectiva histórica de QV pode-se consultar [16]. Atualmente, QV tem sido extensivamente explorada na literatura de classificação de padrões. A idéia principal é estabelecer uma aproximação quantizada da distribuição da amostra, através de protótipos. Em geral, estes protótipos são associados às observações de acordo com a regra de vizinho mais próximo [1], [16], [17], [4].

Sejam $X \subseteq R^n$ e $P = \{p_1,..., p_r\}$ um conjunto de protótipos tal que $p_k \in R^n$, $\forall k = 1, ..., r$, um procedimento de Quantização Vetorial pode ser definido como uma associação de cada observação $x_i \in X$ a um protótipo p_k. Em geral, p_k é o protótipo mais próximo a x_i de acordo com alguma métrica estabelecida. Assim, se gera uma aproximação quantizada da distribuição da amostra através do conjunto P de protótipos. Formalmente, dado I um conjunto arbitrário contável, α: $X \rightarrow I$ e β: $I \rightarrow P$, define-se um quantizador como a função q: $X \rightarrow P$, $q(x) = \beta o \alpha(x)$. Em geral, a função α pode ser escrita como:

$$\alpha(x_i) = \arg\min_j (d(x_i, p_j))$$

onde 'd' representa uma função de distância. Nesta tese, utiliza-se a distância Euclidiana, porém qualquer outra métrica pode ser utilizada. Já a função β é a associação de 'j' ao protótipo p_j.

Uma forma equivalente de quantização seria considerar a partição do conjunto de observações X em células $S_1,..., S_r$, onde $S_k \equiv \{ x_i \in X \mid \alpha(x_i) = k\}$ e C

= $\{\beta(i) \mid i \in I\}$, onde as funções α e β estão como definidas anteriormente. O quantizador é a função $q(x) = \beta(i)$, dado que $x \in S_i$.

Obviamente, as células geradas pela quantização são mutuamente exclusivas e exaustivas, ou seja, $S_1 \cup ... \cup S_r = X$ e $S_i \cap S_j = \emptyset$, $\forall i, j = 1,...,r, i \neq j$.

Com esta definição de quantizador, garante-se que qualquer observação $x \in X$ será associada ao protótipo p_k mais próximo a ela. Com isso as células $S_1,..., S_r$ formam células de Voronói [4], [5], que tratam-se de células que minimizam a distorção. Em um segundo passo, dadas as células de Voronói obtidas, pode-se encontrar o conjunto de protótipos para estas células que minimizam a distorção (no caso da métrica Euclidiana, são os centróides de cada célula). Com isso é possível obter um algoritmo iterativo que converge para um mínimo local da distorção [4]. Este algoritmo foi proposto em [21] para uma variável aleatória.

Muitos algoritmos podem ser considerados dentro do contexto de QV, como por exemplo, K-means [40], mapas auto-organizáveis [5], [20] e o algoritmo LBG [1]. Nesta tese, utiliza-se o LBG, que é considerado uma extensão da proposta de [21] para um vetor aleatório. A seguir, pode-se encontrar o algoritmo do LBG [1]:

(1) Inicialização: Sejam r = número de protótipos; o valor de corte da distorção $\varepsilon \geq 0$; um conjunto inicial de protótipos P_0; um conjunto de observações $X = \{x_j \mid j = 1,...,n\}$; $m = 0$ e $D_{-1} = \infty$.

(2) Dado $P_m = \{p_i \mid i = 1,...,r\}$, encontre a partição de distorção mínima $Part(P_m) = \{S_i \mid i = 1,...,r\}$ do conjunto de observações, onde $x_j \in S_i$ se $d(x_j, p_i) \leq d(x_j, p_s)$, para todo s. Calcule a distorção média:

$$D_m = \frac{1}{n} \sum_{j=1}^{n} \min_{p \in P_m} \left[d(x_j, p) \right]$$

(3) Se $(D_{m-1} - D_m)/ D_m \leq \varepsilon$, pare com P_m o conjunto final de protótipos. Caso contrário, continue.

(4) Encontre o conjunto ótimo de protótipos para $Part(P_m)$, isto é, os centróides de cada conjunto S_i, e faça deste conjunto P_{m+1}. Troque m por m + 1 e volte para a etapa (1).

Observe que uma vez que o conjunto final de protótipos P_m é obtido, ele pode ser usado para observações fora do conjunto de treinamento através da regra do vizinho mais próximo.

A escolha de P_0 também é levada em consideração em [1]. A forma utilizada nesta tese é a seguinte:

(1) Inicialização: Seja t = 1 e defina $P_0(1)$ como o centróide das observações;

(2) Dado o conjunto de protótipos $P_0(t)$ com t protótipos $\{ p_i \mid i = 1,...,t\}$, divida cada p_i em dois protótipos $p_i + \varsigma$ e $p_i - \varsigma$, onde ς é um vetor de perturbação fixa. O conjunto $\{ p_i + \varsigma, p_i - \varsigma, \mid i = 1,...,t\}$ tem 2t protótipos. Troque t por 2t.

(3) t = r? Se for, $P_0 = P_0(t)$ é o conjunto inicial de protótipos. Caso contrário, execute o algoritmo para os t protótipos e volte para a etapa 1.

Note que, desta forma, o conjunto final de protótipos terá sempre cardinalidade $r = 2^{\xi}$, $\xi = 0,1,...$.

O LBG tem sido largamente utilizado na literatura. Exemplos de aplicações em que QV através do LBG são combinadas com outras técnicas estão em [22], onde um algoritmo para restauração de imagem e restauração de imagem cega é proposto; [23] para transmissão de imagens quantizadas; em uma proposta para obtenção de protótipos para um algoritmo de classificação [24]; e em [25] para um algoritmo de reconhecimento de fala. Dois exemplos de mudanças baseadas no algoritmo original estão em [26], onde uma busca pelo protótipo mais próximo se limita a um subconjunto dos protótipos originais que estejam dentro de uma hiperesfera e em [27], onde uma medida de utilidade é dada a cada protótipo e o protótipo com menor utilidade é movido para outra região do espaço.