



US006397200B1

(12) **United States Patent**
Lynch, Jr. et al.

(10) **Patent No.:** **US 6,397,200 B1**
(45) **Date of Patent:** **May 28, 2002**

(54) **DATA REDUCTION SYSTEM FOR IMPROVING CLASSIFIER PERFORMANCE**

(75) Inventors: **Robert S. Lynch, Jr.**, Groton; **Peter K. Willett**, Coventry, both of CT (US)

(73) Assignee: **The United States of America as represented by the Secretary of the Navy**, Washington, DC (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/285,173**

(22) Filed: **Mar. 18, 1999**

(51) **Int. Cl.**⁷ **G06F 15/18**

(52) **U.S. Cl.** **706/20; 702/181**

(58) **Field of Search** 706/40, 20; 702/181

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,959,870 A *	9/1990	Tachikawa	382/253
5,479,576 A *	12/1995	Watanabe et al.	706/25
5,572,597 A *	11/1996	Chang et al.	382/125
5,633,948 A *	5/1997	Kegelmeyer, Jr.	382/132
5,701,398 A *	12/1997	Glier et al.	706/41
5,790,758 A *	8/1998	Streit	706/33
5,796,924 A *	8/1998	Errico et al.	706/25
5,884,294 A *	3/1999	Kadar et al.	706/10
5,999,893 A *	12/1999	Lynch, Jr. et al.	702/181
6,009,418 A *	12/1999	Cooper	706/15
6,027,217 A *	2/2000	McClure et al.	351/224
6,035,057 A *	3/2000	Hoffman	382/159
6,278,799 B1 *	8/2001	Hoffman	382/159

OTHER PUBLICATIONS

Lynch Jr., et al.; "Classification using dirichlet priors when the training data are mislabeled". 1999 IEEE International Conference on Acoustics, Speech and Signal Processing, Mar. 1999, vol. 5, pp. 2973-2976.*

Lynch Jr., et al.; "Testing the statistical similarity of discrete observations using Dirichlet priors". 1998 IEEE International Symposium on Information Theory, Aug. 1998, p. 144.*

Lynch Jr., et al.; "Bayesian classification and the reduction of irrelevant features from training data". Proceedings of the 37th IEEE Conference on Decision and Control, Dec. 1998, vol. 2, pp. 1591-1592.*

Morris et al.; "Some solution to the missing feature problem in data classification, with application to noise robust ASR". Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, May 1998, vol. 2, pp. 737-740.*

Huang et al.; "An automatic hierarchical image classification scheme". Proceedings of the sixth ACM international conference on Multimedia, Sep. 1998, pp. 219-228.*

Vivarelli et al.; "Using Bayesian neural networks to classify segmented images". Proceedings of the fifth IEE International Conference on Artificial Neural Networks, Apr. 1997.*

* cited by examiner

Primary Examiner—Thomas Black

Assistant Examiner—Kelvin Booker

(74) *Attorney, Agent, or Firm*—Michael J. McGowan; James M. Kasischke; Prithvi C. Lall

(57) **ABSTRACT**

A data reduction method for a classification system using quantized feature vectors for each class with a plurality of features and levels. The reduction algorithm consisting of applying a Bayesian data reduction algorithm to the classification system for developing reduced feature vectors. Test data is then quantified into the reduced feature vectors. The reduced classification system is then tested using the quantized test data. A Bayesian data reduction algorithm is further provided having by computing an initial probability of error for the classification system. Adjacent levels are merged for each feature in the quantized feature vectors. Level based probabilities of error are then calculated for these merged levels among the plurality of features. The system then selects and applies the merged adjacent levels having the minimum level based probability of error to create an intermediate classification system. Steps of merging, selecting and applying are performed until either the probability of error stops improving or the features and levels are incapable of further reduction.

7 Claims, 2 Drawing Sheets

