

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 November 2002 (28.11.2002)

PCT

(10) International Publication Number
WO 02/095534 A3

(51) International Patent Classification?: **G06F 15/18,**
G06E 1/00

(72) Inventors; and

(21) International Application Number: PCT/US02/16012

(75) Inventors/Applicants (for US only): **WESTON, Jason** [GB/GB]; 20 Burhill Way, St. Leonard's on Sea, Sussex TN 38 OXP (GB). **ELLISSEEFF, Andre** [FR/US]; 407 W. 51st Street, Apartment #5E, New York, NY 10019 (US). **SCHOELKOPF, Bernhard** [DE/DE]; Payerstr. 6, 72074 Tübingen (DE). **CHAPELLE, Olivier** [FR/FR]; 147 rue de Charenton, F-75012 Paris (FR). **PEREZ-CRUZ, Fernando** [ES/ES]; Area de Teoria de la Senal y Comunicaciones, Universidad Carolos III de Madrid, C/Butarque 15, E-28911 Madrid (ES).

(22) International Filing Date: 20 May 2002 (20.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/292,133 18 May 2001 (18.05.2001) US
60/292,978 23 May 2001 (23.05.2001) US
60/332,021 21 November 2001 (21.11.2001) US

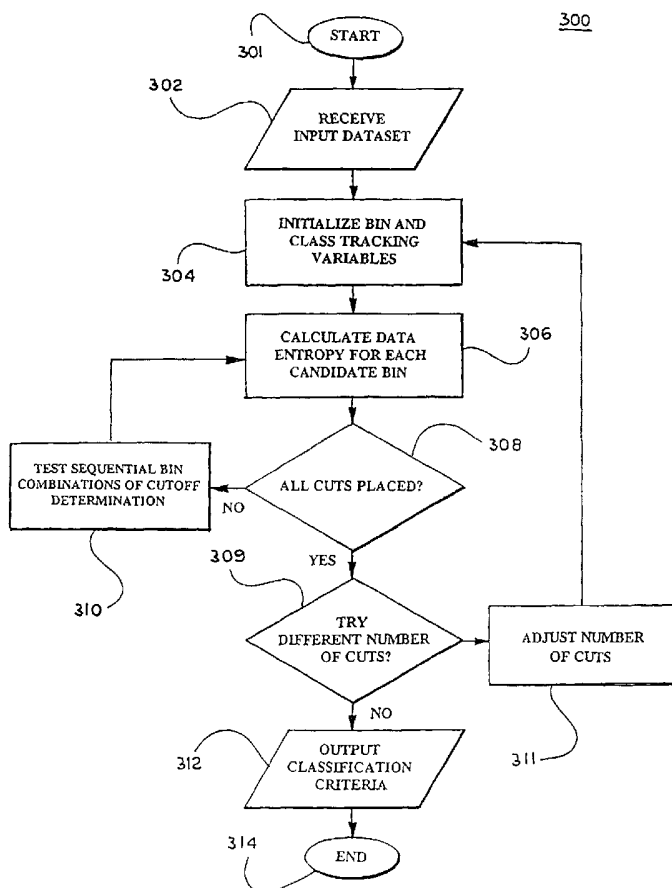
(74) Agents: **PRATT, John, S.** et al.; 1100 Peachtree Street, Suite 2800, Atlanta, GA 30309 (US).

(71) Applicant (for all designated States except US):
BIOWULF TECHNOLOGIES, LLC [US/US]; 7505 Waters Avenue, Suite E4, Savannah, GA 31406 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,

[Continued on next page]

(54) Title: METHODS FOR FEATURE SELECTION IN A LEARNING MACHINE



(57) Abstract: In a pre-processing step prior to training a learning machine, pre-processing includes reducing the quantity of features to be processed using feature selection methods selected from the group consisting of recursive feature elimination (RFE), minimizing the number of non-zero parameters of the system (l_0 -norm minimization), evaluation of cost function to identify a subset of features that are compatible with constraints imposed by the learning set, unbalanced correlation score and transductive feature selection. The features remaining after feature selection are then used to train a learning machine for purposes of pattern classification, regression, clustering and/or novelty detection. (Fig. 3, 300, 301, 302, 304, 306, 308, 309, 310, 311, 312, 314)

WO 02/095534 A3



LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

Published:

— with international search report

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(88) Date of publication of the international search report:

24 April 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/16012

A. CLASSIFICATION OF SUBJECT MATTER				
IPC(7) : G06F 15/18; G06E 1/00				
US CL : 706/16; 706/12				
According to International Patent Classification (IPC) or to both national classification and IPC				
B. FIELDS SEARCHED				
Minimum documentation searched (classification system followed by classification symbols) U.S. : 706/16; 706/12				
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Non-Patented Literature (ACM Portal & IEBE); CITESEER NEC ResearchIndex; www.google.com				
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Non-Patented Literature (ACM Portal & IEBE); CITESEER NEC ResearchIndex; www.google.com				
C. DOCUMENTS CONSIDERED TO BE RELEVANT				
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
A	Terrance S. Furey; Nello Cristianini; Nigel Duffy; David W. Bednarski; Michel Schummer; David Haussler; "Support Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expression Data; USCS-CRL-00-04; April 04, 2000; pps. 1-17	1-19		
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.				
<table style="width: 100%; border: none;"> <tr> <td style="width: 50%; border: none;"> * Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed </td> <td style="width: 50%; border: none;"> "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family </td> </tr> </table>			* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family			
Date of the actual completion of the international search		Date of mailing of the international search report		
26 November 2002 (26.11.2002)		13 DEC 2002		
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230		Authorized officer <i>JM</i> John Follansbee <i>James R. Matthews</i> Telephone No. (703) 308-3900		