
**The 1998 IEEE International Joint Conference on
Neural Networks Proceedings**

**IEEE World Congress
on Computational Intelligence**

May 4 - May 9, 1998
Anchorage, Alaska, USA



Volume 1 Page 1 - 802

The 1998 IEEE International Joint Conference on Neural Networks Proceedings

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. All rights reserved. Copyright © 1998 by the Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Number 98CH36227
ISBN 0-7803-4859-1 (softbound)
ISBN 0-7803-4860-5 (casebound)
ISBN 0-7803-4861-3 (microfiche)
ISBN 0-7803-4862-1 (CD-ROM)
ISSN 1098-7576

Additional copies of this publication are available from

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
1-800-678-IEEE or 1-732-981-0060

Coordinated and produced by



in cooperation with

DOCUMENTATION

Contact ALTEC at 800.997.9921 x 1334

Author's Index

A		
Abad Grau, M.	2334	
Abdelbar, A.	402	
Adachi, N.	1057	
Adachi, S.	651	
Aiken, J.	713	
Aiyad, R.	419	
Akamatsu, N.	686	
Albesano, D.	2190	
Allinson, N.	2277	
Almeida, L.	2224	
Altug, S.	1250	
Alvager, T.	1580	
Amaral, J.	1235	
Amaral, W.	1624	
Amari, S.	39, 1530, 2328	
Amin, A.	1743	
Anand, T.	1466	
Andersen, T.	1920	
Anderson, R.	1559	
Aranki, N.	598	
Arantes do Amaral, J.	1889	
Araujo, A.	474, 2057, 2378	
Ariyanto, A.	310	
Asai, H.	646, 1652, 2157	
Atiya, A.	419, 1101	
Auda, G.	1356	
Aweya, J.	140	
Azimi-Sadjadi, M.	177, 1732	
B		
Babuska, R.	1613	
Back, B.	266	
Baig, A.	160	
Bakker, R.	2483	
Balachander, T.	1804	
Balicki, J.	1646	
Balkarey, Y.	938	
Bandiera, N.	457	
Bao, W.	367	
Barataud, D.	128	
Barcia, R.	425	
Baron, R.	1433	
Barreto, G.	2378	
Barreto, J.	2413	
Barrows, G.	525	
Bartfai, G.	1137	
Bartfi, G.	2352	
Basak, J.	18	
Base, A.	1516	
Bax, E.	1230	
Bebis, G.	1688, 1717	
Behnke, S.	820	
Beiu, V.	1321	
Benedek, S.	1438	
Bennani, Y.	322	
Bennet, K.	2396	
Berger, T.	784, 2175	
Bergius, P.	189	
Bersini, H.	2402	
Berthold, M.	441, 447	
Best, P.	587	
Beuschel, M.	2425	
Bharitkar, S.	1629	
Biachini, M.	1619	
Bianucci, A.	117	
Birattari, M.	2402	
Bischof, H.	2294	
Blasch, E.	2092	
Blue, J.	2396	
Bluff, K.	630	
Blumenstein, M.	1738	
Boa, P.	1667	
Boehme, H.	372	
Bollacker, K.	1404	
Bologna, G.	146	
Bolouri, H.	165	
Bondarenko, V.	1472	
Bonifacio Jr., J.	205	
Bontempi, G.	2402	
Boronowski, D.	1937	
Borschbach, M.	837	
Bortolozzi, F.	1700	
Borzenko, A.	2541	
Bosch, H.	1167	
Boussalis, H.	943	
Brakensiek, A.	372	
Branca, A.	1590	
Branston, N.	326	
Braumann, U.	372	
Braun, J.	2184	
Brdys, M.	958	
Brejil, M.	814	
Brotherton, T.	876	
Brown, V.	625	
Browse, R.	431	
Bruce, J.	2214	
Bucciarelli, P.	2312	
Burkhardt, S.	1810	
Butler, J.	69	
C		
Caldwell, C.	980	
Caloba, L.	1889	
Calvert, D.	2507	
Campbell, S.	1498	
Campolucci, P.	384	
Campos, M.	763	
Cancelo, G.	996	
Cansian, A.	205	
Caponetto, R.	492	
Capparelli, F.	903	
Carlton, A.	1047	
Carpenter, G.	763	
Carvalho, A.	277	
Carvalho, A.	271	
Casasent, D.	1971	
Castillo, O.	106	
Cawley, G.	101	
Cesmeli, E.	2069	
Cha, E.	1162, 1883	
Chadderton, G.	876	
Chae, S.	156	
Chai, C.	2122	
Chakravarti, S.	832	
Chan, L.	237	
Chan, L.-W.	96	
Chan, S.	657	
Chang, H.	735	
Chang, J.	1428	
Chang, K.	695	
Chang, M.	680	
Chassiakos, A.	943	
Chatelet, E.	226	
Chatwin, C.	2034	
Chau, P.	928	
Chehade, H.	128	
Chen, F.	1520	
Chen, K.	367, 619	
Chen, M.	1851	
Cheng, S.	980	
Cheng, Y.	2273	
Cherkassky, V.	843, 2258	
Chew, P.	891	
Chi, H.	367	
Chiarantoni, E.	1877, 2361	
Chiu, W.-M.	2447	
Cho, C.	12	
Cho, J.	581, 1883	
Cho, K.	569, 669	
Cho, S.	1916	
Choi, D.	2028	
Choi, H.	1943, 2098, 2122	
Choi, J.	245, 1799	
Choi, S.	537, 1673	
Choi, W.	2122	
Cholewo, T.	1074	
Chow, H.-F.	96	
Chow, M.	63, 1250	
Chow, M.-Y.	12	
Choy, I.	245	
Chua, L.	503	
Chuieh, T.	1428	
Chun Su, M.	2116	
Cichocki, A.	39	
Cilingiroglu, U.	553	
Ciobanu, A.	1188	
Cipriano, A.	1981	
Ciuca, I.	1218	
Ciungradi, B.	216	
Claesson, I.	2180	
Cloarec, G.	2513	
Cloete, I.	1150	
Cochofel, H.	963	
Coggins, K.	1760	
Cohen, A.	938	
Coleman, S.	625	
Corchardo, J.	713	
Costa, M.	216	
Costin, H.	1188	
Cox, C.	1047	
Craddock, R.	1361, 1382	
D		
D'Arbo Jr., H.	474	
da Silva, I.	1624	
Dagher, I.	1688, 1717	
Dagli, C.	295	
Dahl, M.	2180	
Damper, R.	2196	
Dandolini, G.	425	
Das, S.	1559	
Daud, T.	593	
Daxwanger, W.	2086	
De, R.	18	
de Arruda, L.	1624	
de Azevedo, F.	2413	
de Carvalho, A.	205	
de Carvalho, A.	1723	
De Castro, F.	1235	
De Castro, M.	1235	
de Castro Silva, L.	1932	
de Figueiredo, R.	690, 1002	
De Nicolao, G.	2407	
de Vries, R.	2104	
DeClaris, N.	171	
DeFigueiredo, R.	408	
Delgado, A.	2126	
Deng, C.	221	
Denham, M.	1541, 1547, 1575	
Di Lecce, V.	1877	
Dich, W.	1299	
Diepenhorst, M.	603, 2519	
Diercksen, G.	1299	
Dillon, T.	29	
Diniz, H.	271	
Distante, A.	1590	
Dobeck, G.	177	
Downs, T.	1661	
Doya, K.	1553	
Dracopoulos, D.	2081	
Draghici, S.	547	
Dreyfus, S.	2016	
Drummond, S.	211	
Dunin-Barkowski, W.	640	
Duong, T.	593	
E		
Ebecken, N.	1889	
Eem, J.	1943	
Eki, Y.	2346	
El Moudni, A.	57	
El-Deredy, W.	326	
Elinson, M.	938	
Eltoft, T.	408	
Embrechts, M.	1438	
Emerson, R.	123	
Engelbrecht, A.	1150, 1608	

Euliano, N. 1063
Evitkhov, M. 938

F

Fancourt, C. 2324
Fanelli, S. 1619
Feldkamp, L. 69, 598,
1816, 2262
Feldkamp, T. 598
Feng, M. 2288
Ferguson, A. 165
Fernandez, M. 1450
Fernandez, P. 101
Ferrandez, J. 541
Ferreira, P. 1261
Fiori, S. 854, 1332, 2312
Fischer, I. 441, 447
Fisher III, J. 1712
Fletcher, L. 1608
Fortuna, L. 492
Fracassi, G. 79
Franco, P. 1235
Frangakis, A. 1937
Freisleben, B. 837
French, V. 1580
Frolov, A. 520
Fu, H. 1754
Fu, L. 45
Fu, M. 918, 1727
Fujimura, K. 301
Fujita, A. 1460
Fukao, T. 1057
Fukumi, M. 686
Fukushima, K. 1172
Furuhashi, T. 35
Furukawa, S. 2306
Fyfe, C. 713

G

Gan, Q. 1156, 1293
Gao, X. 1954
Garavaglia, S. 289
Gauthier, F. 425
Geczy, P. 51
Gedeon, T. 343
Gemello, R. 2190
Georgiopoulos, M. 396, 1688,
1717
Ghosh, J. 695, 757, 1404
Gierczak, C. 69
Giles, C. 1051, 1834, 2483
Girau, B. 1433
Gobel, R. 1810
Goddu, G. 63
Gomez, J. 1564
Gomez, P. 541
Gomez-Ramirez, E. 390
Gomide, F. 251
Goncalves, R. 251
Gorban, A. 1271
Gori, M. 1619

Grabec, I. 1960
Grassi, G. 1504
Grbic, N. 2180
Gross, H. 372, 1992
Grudzinski, K. 1299
Gu, X. 1899
Guan, L. 922
Gubina, F. 1966
Guerriero, A. 1877
Guo, H. 1816
Guo, P. 725
Gutta, S. 1766

H

Ha, K. 1893
Haese, K. 1007
Hagen, C. 837
Hagiwara, M. 514, 1107,
1510
Hahnloser, R. 2373
Halgamuge, S. 7
Halilcevic, S. 1966
Ham, B. 134
Hamagishi, H. 1948, 2535
Hamalainen, J. 193
Han, J. 1667
Hangl, F. 2425
Haouani, M. 57
Hara, K. 2247
Hardier, G. 2441
Harkouss, Y. 128
Harris, P. 112
Haselsteiner, E. 1245
Hassoun, M. 486
Hatano, F. 609
Hattori, D. 1948
Hattori, M. 531
He, Y. 1388, 1478
He, Z. 221
Healy, M. 396
Heh, J. 680
Heileman, G. 396, 1688, 1717
Herault, L. 1239, 1398
Hernandez, C. 1450
Hernandez Molinero, L. 2334
Heywood, M. 2034
Hikawa, H. 557
Hikosaka, O. 1553
Hirabayashi, A. 2236
Hirasawa, K. 480, 968, 1482,
1602, 2346, 2453, 2465
Hirose, A. 1078
Ho, T. 1524
Hodge, L. 283
Holmstrom, L. 1305
Honavar, V. 2208, 2318
Hong, D. 537
Horne, B. 1051, 1834
Horst, P. 277
Hoti, F. 1305
Hou, Z. 1667
Howard, A. 231

Hoyer, P. 859
Hrycej, T. 949
Hu, J. 968, 1482, 1602,
2346, 2453, 2465
Hu-Yu, X. 1624
Huang, C. 2116
Huang, D. 1113
Huang, H. 2116
Huang, K. 1840
Huang, Q. 177
Huang, X. 497
Husek, D. 520
Hussain, T. 431
Hwang, H. 1799
Hyvarinen, A. 859, 1350,
2282

I

Ibikunle, F. 2557
Ibnkahla, M. 865
Ichigashi, H. 651
Ichihashi, H. 1224, 2390
Ichimura, T. 1131
Ikeda, K. 2340
Ikeda, M. 306
Ikeda, N. 486
Ikeda, T. 2306
Ilg, W. 1367
Ineyama, N. 1057
Intanagonwiwat, C. 462
Ioannou, P. 943
Ishibuchi, H. 2051
Ishikawa, M. 2328
Ito, K. 2367
Ito, N. 780
Iwanoto, K. 301
Iwasa, T. 992
Izhikevich, E. 2547

J

Jagielska, I. 24
James, J. 598
Janet, J. 2218
Jarvimaki, I. 193
Jayasuriya, A. 7
Jerry Lin, J. 1428
Jesion, G. 69, 598
Jiang, D. 1640
Jie, Z. 359
Jimenez, D. 753
Jo, T. 2531
Johnson, J. 2092
Johnson, J. 1784
Johnson, T. 876
Joshi, A. 211
Joutsensalo, J. 152

K

Kambhampati, C. 1361
Kamel, M. 1356, 2507
Kameyama, K. 1861

Kamimura, R. 1030, 1845
Kamo, A. 2157
Kanada, H. 2471
Kanata, Y. 1002
Kanerva, P. 1416
Kang, M. 1162
Karaf, A. 2247
Karayiannis, N. 2230
Kasatkina, L. 241, 1685
Kaski, S. 413
Katkovink, V. 1608
Kawakami, H. 2300
Kayama, M. 701
Keller, E. 1559
Khalaf, A. 1975
Khan, M. 262
Khorasani, K. 378, 2063
Khosla, R. 29
Kim, D. 1799
Kim, G.-S. 790
Kim, J. 134
Kim, K. 1883
Kim, S. 1943, 2122
Kimura, E. 1172
King, I. 237, 932
Kirby, S. 112
Kisdia, S. 301
Kita, H. 2010
Kitowski, Z. 1646
Kiviluoto, K. 189, 2268
Klopf, A. 2092
Koch, M. 441, 447
Kocher, M. 1047
Koivisto, H. 2419, 2477
Koivo, H. 2419, 2477
Kok, J. 730
Kolb, T. 1367
Konar, A. 747
Konishi, R. 1772
Kosmatopoulos, E. 943
Kosugi, Y. 349, 1861, 2471
Kothari, R. 1778, 1804
Kozek, T. 503
Krabbes, M. 372, 1992
Kruse, R. 468
Kuh, A. 1584
Kulczycki, P. 1344
Kumamaru, K. 2453, 2465
Kumayama, J. 701
Kung, S. 1051, 1834
Kuo, R. 90
Kuriyama, Y. 1131
Kuroe, Y. 2300
Kussul, E. 241, 1685
Kussul, N. 1685
Kusumoputro, B. 310
Kwon, H. 1799

L

L'Abbate, G. 1569
Labbi, A. 1167

Labonte, G.	880	Lu, T.	2131	Mizutani, E.	503, 2016	Onat, A.	2010
Lai, Z.	725	Lu, W.	1857	Mizutani, S.	2551	Ong, S.	537
Lamedica, R.	79	Lu, Y.	1816	Montuno, D.	140	Onishi, H.	1078
Lamers, M.	730	Lucas, S.	1867	Moore, W.	2146	Onozuka, T.	2045
Lampinen, J.	199	Ludik, J.	1926	Moreira, E.	205	Orlov, Y.	938
Lancieri, L.	1454	Luk, A.	1090, 1143	Mori, T.	2300	Osaki, T.	663
Lane, C.	2163	Luk, P.	1095, 1315, 1377	Morishta, T.	609	Osana, Y.	514, 1107, 1510
Lang, S.	2151	Lursinap, C.	1125	Morizumi, N.	992	Ota, Y.	2435, 2459
Lange, J.	1810	Lynn, G.	1488	Morooka, Y.	701	Oten, R.	690
Lappalainen, H.	2384			Moura-Pires, F.	457, 1910	Ouyang, N.	306
Lathia, B.	134			Mouria-Beji, F.	1200	Owen, M.	1019
Lau, T.	932	M		Mozayani, N.	160		
Lebret, E.	730	Macchiarul, L.	216	Murata, J.	480, 968, 1482, 1602, 2453, 2465	P	
Lee, B.	1916	Maeda, Y.	1002, 1772	Murshed, N.	1700	Pacheco, R.	425
Lee, C.	954	Mak, M.	669	Mutata, J.	2346	Pacut, A.	2004
Lee, H.	2357	Makaruk, H.	1321			Padgett, C.	231
Lee, H.-M.	2447	Makikallio, T.	797			Padgett, M.	803, 1784
Lee, H.	1943	Mana, F.	2190	N		Padilha, T.	277
Lee, L.	1799	Mandal, A.	747	Na, K.	156	Pajunen, P.	152, 1350
Lee, J.	2028	Mandelj, S.	1960	Nagano, S.	1553	Pal, S.	18
Lee, K.	537	Manzie, C.	986	Nakabara, H.	1553	Palaniswami, M.	986, 1256
Lee, M.	2098	Mao, J.	1828	Nakajima, Y.	1948	Panchapakesan, C.	1256
Lee, S.	613, 2122	Marino, C.	928	Nakashima, A.	2236	Pap, R.	1047
Lee, S.	581, 2357	Marko, K.	598, 1286	Nakashima, T.	2051	Paplinski, A.	1372
Lees, B.	713	Martin, T.	980	Nakayama, K.	1494, 1975, 2247, 2253, 2340	Paquir, W.	865
Lefebvre, D.	57	Martineli, E.	271	Nan, J.	2075	Paraskevopoulos, V.	2034
Lendaris, G.	1013	Martinelli, G.	79	Nelson, A.	2489	Parekh, R.	2208, 2318
Leonardis, A.	2294	Martinez, T.	509, 1036, 1920	Neville, R.	1095, 1315, 1377	Park, B.	1905
Lepage, R.	1176	Martins, A.	425	Newcomb, R.	541	Park, C.	1883
Lera, G.	2242	Martins, W.	1932	Ng, S.	1090	Park, D.	826, 1673
Leung, K.	2151	Mascioli, F.	79	Ngoya, E.	128	Park, S.	958
Leung, S.	1090	Mascolo, S.	1504	Niestroy, M.	2110	Park, T.	826
Levine, D.	625	Mathur, S.	913	Nijhuis, J.	603, 2519	Park, W.	1943
Levkowitz, H.	2184	Matsuda, S.	1657	Nikulin, V.	809	Park, Y.	569, 2098
Lewis, C.	1047	Matsumoto, T.	1948, 2535	Ninomiya, H.	646, 1652, 2157	Parlos, A.	1101
Li, B.	63	Matsumoto, Y.	634	Nishikage, T.	634	Pasero, E.	216
Li, C.	1119	Matsushita, S.	35	Nishikawa, Y.	2010	Pasquet, S.	226
Li, G.	1857	Matsuyama, Y.	2306	Noda, T.	480	Patel, R.	2063
Li, J.	359, 2092	May, E.	295	Noumeir, R.	1176	Paul, B.	747
Liang, X.	1388	Mayosky, M.	996	Nunes, L.	2224	Peck, M.	101
Liaw, J.	2175	McCabe, S.	1541, 1547, 1575	Nurnberger, A.	468	Pellegrini, C.	146
Liebe, C.	231	Meert, K.	719, 1926			Peng, H.	1293
Lien, S.	1143	Melin, P.	106			Peng, P.	1147
Lin, C.	2116	Mellgren, R.	625	O		Perez-Urbe, A.	2022
Lin, C.	2116	Melnik, O.	1069	Occhipinti, L.	492	Perov, P.	938
Lin, D.	1794	Mendel, J.	1629	Ogawa, H.	2236	Perry, L.	1873
Lin, S.	353	Menozzi, A.	12	Ogawa, T.	2471	Perry, S.	922
Lin, T.	1051, 1834	Mercado III, E.	1584	Oh, C.	2051	Peterson, T.	768
Lindblad, T.	803	Michaelis, B.	707	Oh, H.	134	Piazza, F.	384, 903, 1332, 2312
Lindsey, C.	803	Micheli, A.	117	Oh, S.	2357	Piazza, P.	854
Lindsey, D.	2069	Michiyuki, Y.	1748	Oh, S.	2028	Pinho, A.	870, 1261
Liou, R.	1851	Milanese, R.	1167	Ohbayashi, M.	968, 1482, 1602, 2346, 2465	Pinzolas, M.	2242
Liqing, Z.	437	Minai, A.	587, 1466	Ohbayayshi, M.	2453	Pollack, J.	1069
Lisboa, P.	112, 326	Mirmirani, M.	943	Ohkouchi, K.	1748	Pozarkik, R.	1444
Liu, X.	897	Misawa, J.	1482	Ohugi, M.	2253	Poznyak, A.	390
Liu, Y.	2202	Mishra, Y.	913	Oja, E.	859, 2268	Prampero, P.	1723
Lo, J.	1311, 2429	Misra, M.	1327	Okabe, Y.	2367	Principe, J.	849, 963, 1063, 1084, 1488, 1712, 1760, 2324
Lo, K.	74	Mitsui, M.	2300	Oliveira, I.	1194	Privault, C.	1398
Lobo, V.	457, 1910	Miyazaki, Y.	497	Omatu, S.	891, 909, 992, 2142		
Lopera, F.	1564	Miyoshi, S.	2340				
Lotikar, R.	1778	Miyoshi, T.	651, 1224, 2390, 2390				
Lu, H.	1478						

Prokhorov, D. 2262
 Protasi, M. 1619
 Prudenzi, A. 79
 Puskorius, G. 69, 598
 Putman, G. 1580

R

Rachkovskij, D. 241, 1685
 Radetzky, A. 468
 Ragheb, M. 337
 Raitz, R. 425
 Ralph, D. 1256
 Ramirez, J. 897
 Rampone, S. 774
 Ratto, G. 1569
 Raudys, S. 1530
 Raveendran, R. 891
 Ray, S. 337
 Rees, N. 713
 Regoli, M. 79
 Reinke, D. 1732
 Rezende, S. 271, 277
 Rhee, F. 1905
 Ribeiro, B. 1987
 Riedmiller, M. 1998
 Rijckaert, M. 719, 1926
 Ringwood, J. 2513
 Rizzi, A. 79
 Roa, L. 541
 Rocha, C. 277
 Rodellar, V. 541
 Roisenberg, M. 2413
 Rojas, R. 820
 Roning, J. 797
 Roppel, T. 1784
 Rouat, J. 1524
 Rouhana, R. 1176
 Rousset, J. 128
 Rukonuzzaman, M. 74
 Ryosuke, K. 663

S

Sabala, I. 39
 Sabsich, T. 165
 Saeki, K. 1460
 Saeks, R. 1047
 Saez, D. 1981
 Saini, J. 913
 Saito, M. 1948, 2535
 Sanchez, E. 2022
 Sandberg, I. 1212, 1265
 Sano, T. 2551
 Sanz-Bobi, M. 1981
 Sarajedini, A. 928
 Sarkar, M. 741
 Schioler, H. 1344
 Schmidt, G. 2086
 Schouten, J. 2483
 Schroder, D. 2425
 Schultz, S. 2218
 Schunemann, S. 707

Scoggins, S. 2218
 Sefic, W. 1047
 Sekerkiran, B. 553
 Sekhar, C. 1206
 Sekine, Y. 1460
 Sellami, L. 541
 Seppala, J. 2419
 Sere, K. 266
 Setiawan, E. 832
 Sgro, J. 123
 Sha, L. 1293
 Shaheen, S. 419
 Shaibon, H. 74
 Shalvi, D. 171
 Shannon, T. 1013
 Shao, N. 1602
 Shao, X. 843
 Shepherd, R. 2258
 Sheu, B. 569, 784
 Shi, B. 1857
 Shi, D. 974
 Shibata, K. 2367
 Shimohara, K. 2551
 Shin, C. 1119
 Shin, S. 2357
 Shinchi, T. 1772
 Shishkin, S. 640
 Shouno, H. 1172
 Si, J. 353
 Sick, B. 84
 Simonetti, M. 384
 Singh, S. 1743
 Sio, K. 954
 Sipe, M. 1971
 Sirag, D. 1147
 Sjodin, G. 1410
 Smith, K. 1679
 Snijder, F. 1613
 Someya, K. 1460
 Song, J. 245
 Song, S. 1799
 Song, S. 2075
 Song, W. 2288
 Sonka, M. 814
 Souza, J. 425
 Spaanenburg, L. 603, 2519
 Speer, R. 2146
 Sperduti, A. 117
 Srivastava, M. 913
 Stacey, D. 283, 2507
 Stanton, P. 123
 Starita, A. 117, 1569
 Stassinopoulos, D. 2039
 Stateczny, A. 1646
 Steffens, F. 1608
 Stella, E. 1590
 Stephan, V. 1992
 Stubberud, P. 2214
 Stubberud, S. 1019
 Su, M. 735
 Subramanian, K. 1156
 Suddarth, S. 593

Sudduth, K. 211
 Sugahara, K. 1772
 Suganthan, P. 1706
 Sugarhara, K. 663
 Sugi, J. 1948, 2535
 Sugita, Y. 701
 Sukharev, A. 640
 Sun, K. 613
 Sun, P. 1286
 Sun, R. 1, 768
 Sun, W. 885
 Sundareshan, M. 1024, 2169
 Sung, A. 316
 Sutton III, J. 2218
 Swindell, R. 112
 Swinarski, R. 1910
 Sze, H. 2151

T

T'sou, B. 657
 Tagliarini, G. 402
 Takeda, F. 634, 1748
 Takeda, N. 2306
 Talebi, H. 2063
 Talukder, A. 1971
 Tan, A. 183
 Tan, Y. 221
 Tanaka, H. 1224
 Tanaka, M. 1748
 Tanprasert, T. 1125
 Tavares, V. 1488
 Tawel, R. 598
 Teo, C. 183
 Ter Haseborg, H. 603
 Teramoto, I. 609
 Teyssier, J. 128
 Thakoor, A. 593
 Tham, M. 1361
 Thomas, D. 326
 Thomas, T. 593
 Tian, B. 177, 1732
 Todd, J. 2069
 Todirascu, A. 1188
 Togo, K. 1602
 Tokutaka, H. 301
 Tomikawa, Y. 1494
 Tontini, G. 1694
 Toshimitsu, K. 968
 Tranprasert, C. 1125
 Treadgold, N. 343
 Trecate, G. 2407
 Trussell, H. 1250
 Tsai, M. 1422
 Tsai, R. 784
 Tsai, W. 2447
 Tse, P. 2151
 Tso, S. 1899
 Tsutsui, H. 35
 Tulppo, M. 797
 Tumer, K. 757

U

Uchikawa, Y. 780
 Uluyol, O. 337
 Uncini, A. 384, 854, 903
 Usui, S. 51
 Uto, K. 349

V

Vacca, F. 2361
 Vainamo, K. 797
 Van Bael, P. 719
 van den Bleek, C. 2483
 Van Gorp, J. 2136
 Vanharanta, H. 266
 Vanyachine, A. 803
 Vaucher, G. 160
 Vellido, A. 112
 Venema, R. 2519
 Ventura, D. 509, 1036
 Verhaegen, M. 1613
 Verma, B. 332, 1738, 1790, 2163
 Verzi, S. 396
 Vierira, M. 2057
 Visala, A. 1281
 Voigt, H. 1810
 Von Zuben, F. 251, 1932
 Vonder Haar, T. 1732
 Voronenko, D. 674

W

Wallace, J. 630
 Warns, B. 2104
 Wan, E. 2489
 Wang, C. 1084
 Wang, D. 619, 897, 1182, 1498, 2069
 Wang, F. 363
 Wang, G.-N. 790
 Wang, J. 885, 1422, 1596, 1640
 Wang, L. 1679
 Wang, T. 725
 Wang, Y. 1624
 Wangchao, L. 359
 Ware, J. 1218
 Warner, B. 1327
 Warwick, K. 1361, 1382
 Watanabe, E. 2501
 Watanabe, M. 780
 Watanabe, T. 301
 Watson, H. 986
 Watta, P. 486
 Wazlawick, R. 1194
 Wechler, H. 1766
 Wei, Y. 1293
 Wenjing, L. 359
 White, M. 2218
 White, R. 2352
 Wilamowski, B. 2435, 2459

Wille, J.	1367	Xu, D.	849	Yao, X.	2202	Yu, W.	2131
Wilson, K.	1536	Xu, G.	974	Yasunaga, M.	563	Yudashkin, A.	1392
Wismuller, A.	575	Xu, L. ...	237, 725, 1275, 1822, 2495, 2525	Yazdizadeh, A.	378		
Wong, H.	112	Xu, Y.	1754	Yea, B.	663	Z	
Wong, Y.	1024			Yegnanarayana, B. ...	741, 1206	Zaknich, A.	256
Wood, I.	1661	Y		Yilong, L.	2288	Zanevsky, Y.	303
Wooten, D.	963	Yacoub, M.	322	Yin, H.	2277	Zegers, P.	2169
Wu, H.-C.	849	Yagi, T.	780	Yixin, Z.	2557	Zerhouni, N.	57
Wu, L.	1388	Yamamoto, H.	1652	Yoneyama, T.	646, 2157	Zhang, B.	918, 1727
Wu, P.	613, 1338	Yamamura, M.	2045	Yongbin, W.	359	Zhang, Q.	140, 363
Wunsch, D.	241, 640, 1271, 1685	Yamauchi, K.	306	Yoshida, E.	563	Zhang, W.	1899
		Yan, H.	918, 1727	Yoshida, K.	2328	Zhdanov, A.	1042
X		Yang, J.	2208	Yoshihara, K.	301	Zhu, Z.	974
Xia, Y.	1596	Yang, S.	974	Yoshioka, M.	909, 2142	Zin, A.	74
Xibilia, M.	492	Yao, J.	1156	You, C.	537	Zurada, J.	1074
Xiong, F.	221			Yu, H.	680		

An HMM/MFNN Hybrid Architecture Based on Stacked Generalization for Speaker Identification

Wei-quan Bao Ke Chen and Hui-sheng Chi

Center for Information Science, National Laboratory of Machine Perception

Peking University, Beijing 100871, China

Email: wqbao@usa.net

Abstract

A hybrid architecture based upon Hidden Markov Models (HMMs) and Multilayer Feed-forward Neural Network (MFNN) is presented for speaker identification. Unlike most of the previous combining methods, the proposed architecture uses HMMs to model individual speaker and uses MFNN to deal with the inter-speaker information for improving performance. Learning in the proposed architecture consists of two phases. In particular, only a small amount of data is needed for training. The HMM/MFNN architecture has been applied to text-independent speaker identification. Simulation has shown that the hybrid architecture yields better identifying rate than that of conventional methods and other hybrid architectures.

Keywords: Speaker identification, Stacked generalization, HMM/MFNN hybrid architecture.

1 Introduction

The Hidden Markov Model (HMM) technique has already been applied to both speech and speaker recognition [1, 2]. However, the HMM has weak discriminative ability for classification [3]. The Multilayer Feed-forward Neural Network (MFNN) has been regarded as a powerful tool for static pattern classification [4] without temporal sequential processing. A rational selection is hybrid architectures, among them the HMM and Neural Network hybrids are promising and have been extensively studied [3, 8]. Basically, these hybrid architectures are classified into two categories; i.e. the MFNN is integrated with the HMM either for pre-processing or for post-processing. For pre-processing [4, 9], activation values of output nodes of the MFNN are used as some kind of posterior probability. This approach results in no assumption on a specific observation probability distribution in the HMM and it eliminates the feature independence assumption. How-

ever, much more training data are usually needed for parameter estimation since there exists a great number of adjustable parameters in the MFNN. In another hybrid architecture [10], the MFNN is used as labeler for discrete parameter HMMs. For post-processing, an elaborate representation of outputs of HMMs is sent to the MFNN as its input. A hybrid architecture has been proposed for the on-line handwritten character recognition, in which the vector consisting of all final observation probabilities of a sequence is used as the input of the MFNN to capture the variance of the sequence [5]. However, the method cannot be applied to speaker identification since the representation is unsuitable for conveying speaker-related information [8].

Unlike most of the previous hybrid approaches, we propose a hybrid architecture in which the MFNN is used for post-processing. In speaker identification, each HMM is often used to model an individual speaker's personality, which results in the lack of use of *inter-speaker information*. Empirical studies have shown that the classification errors of HMMs are relatively stable such that an MFNN can be employed to "learn" from HMMs for use of inter-speaker information based upon the principle of *stacked generalization* [6]. For this purpose, an elaborate representation of matching scores produced by observation sequences with HMMs is presented as the input of the MFNN. Thus the dimension of the input vector is just equal to the number of HMMs in the system, which results in a fast training of the MFNN due to the lower dimensionality in the input space. Some simulations with 20 population have been conducted on the KING database for text-independent speaker identification. It is evident that the proposed hybrid architecture yields better identifying rate than the conventional HMM method [8].

2 Stacked Generalization

Stacked generalization is a generic term referring to any scheme for feeding information from one set of generalizers to another before forming the final guess. A generalizer is a mapping $\{(x_k, y_k), q\} \rightarrow \{g\}$, $1 \leq k \leq m$, $x_k \in \text{space } R^n$, $y_k \in R^n$ and $g \in R$, where (x_k, y_k) is a pair of a learning set of m pairs, q is a question and g is a guess [6]. Full generality would have the guess $\in R^p$, not R . However, for most applications one can replace a generalizer making guesses in R^p with the Cartesian product of p separate generalizers making guesses in R . Thus, we can simply take p to equal to 1. The distinguishing feature of stacked generalization is that the information fed up the net of generalizers comes from multiple partitions of the original learning set, all of which split up the learning set into two subsets. Each such pair of subsets is then used to collect information about the biases of the generalizing behavior of the original generalizer(s) with respect to the learning set. Stacked generalization is a means of estimating and correcting for the biases of the constituent generalizer(s) with respect to the provided learning set.

The first step in employing stacked generalization is choosing a set of r partitions, each of which splits a learning set Θ into two (usually disjoint) sets. Then label such set of partitions as Θ_{ij} , where $1 \leq i \leq r$ and $j \in \{1, 2\}$. Such set of partitions is called a *partition set*. For instance, for a *cross-validation partition set* (CVPS) [7], r is equal to m , where m is the total pair number of learning set. In CVPS, for all i , Θ_{i2} consists of a single element of Θ , the corresponding Θ_{i1} consists of the rest of Θ , and $\Theta_{i2} \neq \Theta_{j2}$ for $i \neq j$. For simplicity in the following discussion, we shall only consider the CVPS, thus any set Θ_{i2} consists of merely one element.

Now define the space of original learning set Θ as R^{n+1} or "level 0 space". When generalizing directly from Θ in the level 0 space, any generalizer is called a "level 0" generalizer, and the original learning set Θ is called a "level 0" learning set. For any real-world learning set Θ , there are always N generalizers $\{G_j\}$, where $N \geq 1$ (i.e., we are given a set of N separate sequences of functions $\{g_i\}$, $1 \leq i \leq \infty$), that one can use to extrapolate from Θ . Here we'll just discuss 2 levels stacked generalization. For each of the r partitions of Θ , $\{\Theta_{i1}, \Theta_{i2}\}$, look at a set of k numbers determined by a p subset of the $N\{G_j\}$ working together with that partition. Typically these k numbers can be those like the guesses made by the $\{G_j\}$ when taught with Θ_{i1} and presented as a question of the input component of the element Θ_{i2} , i.e., $G_j(\Theta_{i1}; \text{the input component of } \Theta_{i2})$. Take each such set of k numbers and view it as the input component

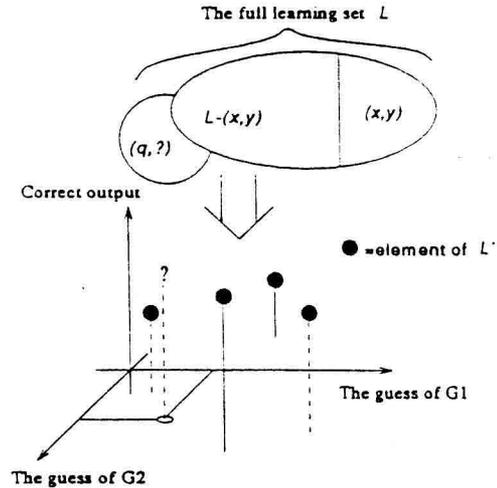


Figure 1: A simple example of how to use stacked generalization to combine generalizers(From [6])

of a point in a space R^{k+1} . The corresponding output value of each such point is calculated from the output component of the corresponding Θ_{i2} . This space R^{k+1} is called the "level 1 space". Since we have r partitions of Θ , we have r points in the level 1 space. Those r points are known as the "level 1" learning set. Now the common idea is to take a question in the level 0 space, pass it through the transformations which produced the input components of the level 1 learning set to get a level 1 question in the level 1 input space, and then answer that level 1 question by generating from level 1 learning set. This level 1 guess is then transformed back into a level 0 guess. Any generalizing process of this form is known as "stacked generalization". The process as a whole can be iterated, resulting in levels $p \geq 1$, i.e., multiple stackings. An example of how to use stacked generalization to combine generalizers is given as Figure 1 [6].

Here we combine two generalizers, G_1 and G_2 . The learning set L is represented figuratively by the full ellipse. A question q lying outside of L is also indicated. Finally, a partition of L into two portions is also indicated; one portion consists of the single input-output pair (x, y) , and the other portion contains the rest of L . Given this partition, we train both G_1 and G_2 on the portion $\{L - (x, y)\}$. Then we ask both generalizers the question x ; their guesses are g_1 and g_2 . Generally, since the generalizers have not been trained with the pair (x, y) , both g_1 and g_2 will differ from y . Therefore, we have just learned something. When G_1 guess

g_1 and G_2 guess g_2 , the correct answer is y . This information can be cast as input-output information in a new space, i.e., as a single point with the 2-dimensional input (g_1, g_2) and the output (y). Choosing other partitions of L gives us other such points. Taken together, these points constitute a new learning set, L' . We now train G_1 and G_2 on all of L and ask them both the question q . Then we take their pair of guesses, and feed that pair as a question to a third generalizer which has been trained on L' . This third generalizer's guess is our final guess for what output corresponds to q . Assuming there is a strong correlation between the guesses made by G_1 and G_2 on the one hand, and the correct guess on the other hand, this implementation of stacked generalization will work well.

According to the principle of stacked generalization, in the speaker recognition procedure with HMM-based method, an HMM can be considered as a generalizer G , and its accumulating matching score can be considered as the guess g . Now there are c HMMs corresponding to c speakers, therefore $G_1 G_2 \dots G_c$ must be considered for the full learning sets corresponding to the guesses $g_1 g_2 \dots g_c$, respectively. The full learning set L can be subdivided into two sets L_1 and L_2 . The subset L_1 is used to train all HMMs. These HMMs are called level 0 generalizers and the subset L_1 is called level 0 learning set. Question q is that correct speaker under the output guess distribution $\{g_1 g_2 \dots g_c\}$. Since it is not enough to obtain perfect performance merely with HMM, an MFNN can be used as another generalizer to solve the question. All guesses generated by the HMMs in the level 0 learning set and in L_2 are used to train the MFNN. The MFNN is a level 1 generalizer and its guess (i.e., active output) is the correct guess. The guesses $g_1 g_2 \dots g_c$ of subset L_1 and L_2 can also be claimed priori knowledge that includes both right and error classification information. Thus, in our hybrid architecture, we use the HMMs to generate fixed dimensional feature vector $\{g_1 g_2 \dots g_c\}$ from temporal input sequence in the level 0 space and use the MFNN to classify these static time-alignment feature vector in the level 1 space. Apparently, these priori knowledge will give no benefit to the classifier if the distribution of knowledge about error is not correlated. On the contrary, the classifier will achieve powerful adaptive ability if it can learn from the priori knowledge whose distribution abides by somewhat correlative law. An HMM has the capability of accurately modeling statistical variation in spectral features which also represent speaker identity. When the HMM is trained insufficiently, the best matching model (i.e., the correct guess in level 0 space) may be a specific alternative model instead of a random one, e.g., the second-best

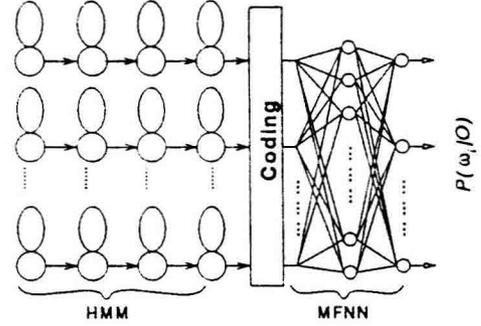


Figure 2: HMM/MFNN hybrid architecture

model related to the true model. In the recognition phase, the input unknown observation sequence is generated by the second-best HMM with great probability, and under the similar training condition, the second-best model keeps relatively stable. When the following unknown observation sequence generates the same incorrect score distribution, the MFNN then can rectify such kind of error.

3 HMM/MFNN Hybrid Architecture

The key idea of the proposed method is to use MFNN to learn from priori knowledge effectively. In the hybrid architecture, we train a full-connected MFNN with Levenberg-Marquardt algorithm for the purpose [7]. HMM/MFNN hybrid architecture is shown as Figure 2.

Given an HMM $\lambda_k = (A_k, B_k, \pi_k)$, $k = 1, 2, \dots, c$ and an unknown observation sequence $O = (O_1 O_2 \dots O_T)$, with forward-backward algorithm [2], we can obtain a set of $\hat{P}(O|\lambda_k) = \log[P(O|\lambda_k)]$. In order to serve for training MFNN, the logarithmic conditional probability is elaborately encoded first. Then the representation is used as the input of the MFNN. We denote the encoded feature vector as $X = \{p_1, p_2, \dots, p_c\}$, where p_k , $1 \leq k \leq c$, is the encoded value of the logarithmic conditional probability. Thus the output of MFNN approaches posterior probability

$$P(\omega_i|O) \cong f\left\{\sum_{j=1}^H w_{ji}^{h \rightarrow o} f\left[\sum_{k=1}^c w_{kj}^{i \rightarrow h} p_k\right]\right\} \quad (1)$$

where $w_{ji}^{h \rightarrow o}$ is a weight from the j th hidden node to the i th output node and $w_{kj}^{i \rightarrow h}$ is a weight from the k th input node to the j th hidden node. $f(\cdot)$ is a sigmoid

function. $P(\omega_i|O)$ is the active probability of class ω_i with input observation sequence O . Instead of simply selecting the model producing the maximum value of $P(O|\lambda_k)$, the proposed architecture makes an MFNN perform the exact classification according to the priori distribution of all $P(O|\lambda_k)$, $1 \leq k \leq c$. In this architecture, the HMM still elaborately represents the temporal variation while the MFNN yields a kind of static pattern of which the inherent temporal information has been carried. We select the output node label whose active value is maximum as the exact class, i.e.

$$i_{max} = \arg \max_i P(\omega_i|O), \quad 1 \leq i \leq c \quad (2)$$

During the reestimation procedure of HMM parameters, in order to avoid underflow of iteration computation, a *Scaling* technique was introduced [2],

$$\hat{P}(O|\lambda_k) = \log[P(O|\lambda_k)] = - \sum_{t=1}^T \log C_t \quad (3)$$

where C_t is the scaling factor as

$$C_t = \frac{1}{\sum_{i=1}^N \alpha_t(i)} \quad (4)$$

and generally the denominator of C_t is less than 1. Thus, for a long sequence, the value of $\hat{P}(O|\lambda_k)$ is very small. Directly input it to MFNN will give MFNN an extreme difficulty to adjust its connection weights to convergence. An encoding rule is introduced to the hybrid architecture,

$$\text{Encoding Rule: } p_k = \frac{P_{max} - \hat{P}(O|\lambda_k)}{P_{max}}, \quad k = 1, 2, \dots, c \quad (5)$$

where $P_{max} = \max\{\hat{P}(O|\lambda_k)\}$, $1 \leq k \leq c$. This rule considers the similar degree between every model and the best matching model as the input of MFNN.

Encoded by the rule, each observation sequence can finally obtain one vector as the input of MFNN. The amount of training data of MFNN is reduced extremely after the HMM processing while the temporal information of sequence is still represented accurately by HMM.

4 Experiments

The experiments were primarily conducted using a subset of the KING speech database which includes 5 wide-band sessions (labeled session 1 to session 5 as S-1, S-2, ..., S-5, respectively) per speaker. 1-2 sessions

are used as training data and the remaining sessions as testing data. First, silence and unvoiced parts are removed from the speech prior to feature extraction using an adaptive energy threshold detector, and then a sequence of feature vectors is produced with 256 frame-rate and 16 orders LPC cepstrum analysis [8]. The final average length of speech of each speaker in a session is about 15 to 20 seconds. To evaluate different test utterance lengths, the sequence of feature vectors is divided into overlapping segments of T feature vectors. The first two segments from a sequence would be,

$$\begin{array}{c} \text{Segment 1} \\ \overbrace{\bar{x}_1, \dots, \bar{x}_L, \bar{x}_{L+1}, \dots, \bar{x}_T, \bar{x}_{T+1}, \dots, \bar{x}_{T+L}} \\ \\ \text{Segment 2} \\ \bar{x}_1, \dots, \bar{x}_L, \overbrace{\bar{x}_{L+1}, \dots, \bar{x}_T, \bar{x}_{T+1}, \dots, \bar{x}_{T+L}} \end{array}$$

A test segment length of 3.2 seconds would correspond to $T = 100$ feature vectors for a 32 ms frame rate. L is the number of shifting frames between two adjacent segments. Here L is equal to 10, i.e. 320 ms speech. Each segment of T vectors is treated as a separate test utterance. The final recognition rate is then computed as the percentage of correctly identified T length segments over all test utterances.

The HMM for text-independent speaker identification is generally modeled as an ergodic continuous Gaussian mixture model. In our experiments, an ergodic Gaussian mixture CHMM with 8 states and 8 mixtures is adopted to provide the input vector for MFNN. The training phase of the ergodic CHMM is slightly different with that of left-right CHMM, as follows: The first is to produce an initial model through a *k-means* clustering algorithm by using all training data. The cluster number is identical to the number of states of CHMM, i.e. 8. Then the members of each state are clustered into 8 mixtures. The mean vector and variance matrix can be computed through these clusters and all training vectors. The initial transition probabilities are estimated by counting transitions between states in the training data. Although such an algorithm is not globally optimal, it is an effective way to generate an ergodic HMM. The second is the reestimation of the parameters. Standard *Segmental k-mean* algorithm is used for training the initial model until convergence [8]. In order to evaluate the performance according to our modeling method for different test utterance length, we carried out some experiments for 1.6s, 3.2s, 4.8s, 6.4s, 8.0s, 9.6s test length. Where S-1 and S-2 are used for training ergodic CHMM, S-3, S-4 and S-5 for testing. In the hybrid architecture, The training sets for MFNN are selected as follows: 30 segments are extracted from S-1 and S-2 randomly as well as 10 segments from S-3 and 5 segments from S-4. Thus, total 75 segments are used for training MFNN. S-5 and the remainder

of S-3 and S-4 are used for testing the hybrid architecture performance. A comparison of average identifying rates of S-4 and S-5 using ergodic CHMM and ergodic CHMM/MFNN is shown in Table 1.

Table 1: Average identifying accuracies(%) of S-4 and S-5 for different test utterance lengths with ergodic CHMM and ergodic CHMM/MFNN

Test length(s)	1.6	3.2	4.8	6.4	8.0	9.6
CHMM	74.55	81.98	85.03	85.75	86.83	86.55
CHMM/MFNN	76.64	86.44	90.86	94.29	97.51	94.11

As expected, with increase test utterance length, identification performance increases. Identification rate for the shortest test utterance length shows the greatest improvement. It is also evident that the necessary test utterance length is no less than 5 seconds in order to achieve a better performance. Furthermore, from the classification results provided by ergodic CHMMs, we find a general confusion error, i.e. in this session, some speaker is always recognized as another speaker incorrectly, and in another session, the same error occurs repeatedly. Such kind of regular priori knowledge about error classification benefits for a HMM/MFNN hybrid.

From the experimental results, merely a little better performance can be yielded for short test utterances, e.g. 1.6 and 3.2 seconds. By analyzing the confusion matrix of recognition results to different speakers, we found that the distribution of priori knowledge about error classification is very relaxed for short test utterances. It is briefly owing to the fact that short test utterance is not enough for elaborately representing speaker identity by CHMM, the matching score generated by CHMM naturally cannot provide benefit priori knowledge for MFNN. For 8.0 seconds test utterance length, using our hybrid architecture, we achieved the best identifying accuracy compared with conventional CHMM method. With these identification methods, it is evident that without enough training data, HMM/MFNN hybrid architecture method is far superior to a conventional CHMM method.

5 Conclusions

In this paper, we have presented a hybrid architecture based upon HMM and MFNN for speaker identification. The idea, that is, multiparts of a learning set are used to train a set of level 0 generalizers to obtain priori knowledge and such priori knowledge is used to train a level 1 generalizer to get the correct classification, can be generalized other classification problems. In particular, training HMM and MFNN individually reduces

the computational burdens of the whole identification system while encoding the output parameters of HMM does not change the accurate representation to temporal sequence but speed up training of MFNN. Experimental results show that this hybrid architecture is a promising architecture and yields better performance than HMM-based method.

6 Acknowledgments

This work was jointly supported by the Climbing Program-National Key Project for Fundamental Research in China with Grant NSC 92097 and NNSF Grant 69475007, 69635020. The authors gratefully acknowledge the colleague Z.Wu for his benefit discussion about this work, and also thank L.Yang for providing us his Levenberg-Marquardt algorithm program.

References

- [1] S. Furui, An Overview of Speaker Recognition Technology, Proc. ESCA Workshop on Automatic Speaker Recognition, Martigny, Switzerland, April 4-7, 1994, pp.1-9.
- [2] L.R.Rabiner, A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition. Proc. IEEE, Vol.77, No.2, 1989, pp.257-286.
- [3] H.Bourland and C.J.Wellekens, Links Between Markov Models and Multilayer Perceptrons, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 12, No. 12, Dec. 1990, pp.1167-1178.
- [4] R.P.Lippmann, Pattern Classification Using Neural Networks, IEEE Commun. Mag. Vol.27, No.11, Nov. 1989, pp.47-64.
- [5] Sung-Bae Cho and J.H.Kim, An HMM/MLP Architecture for Sequence Recognition, Neural Computation, 7,1995, pp.358-369.
- [6] D.H.Wolpert, Stacked Generalization, Neural Networks, 5, 1992, pp.241-259.
- [7] Li, Ker-Chan, From Stein's Unbiased Risk Estimates to The Method of Generalized Cross-Validation, The Annals of Statistics, 13, 1985, pp.1352-1377.
- [8] Weiquan Bao, An HMM/MFNN Hybrid Architecture for Speaker Identification, M.E. Thesis, Center for Information Science, Peking University, 1997
- [9] J.M.Naik and D.M.Lubensky, A Hybrid HMM-MLP Speaker Verification Algorithm for Telephone Speech, ICASSP'94, pp.1-153-156.
- [10] P. L. Cerf, W. Ma and D. V. Compernelle, Multi-layer Perceptrons as Labelers for Hidden Markov Models, IEEE Trans. Speech and Audio Processing, Vol.2, No.1, 1994, pp.185-193.

Neural Networks for Gesture-based Remote Control of a Mobile Robot

Hans-Joachim Boehme, Anja Brakensiek, Ulf-Dietrich Braumann, Markus Krabbes,
Horst-Michael Gross

Department of Neuroinformatics, Technical University of Ilmenau, 98684 Ilmenau, Germany.
E-mail: hans@informatik.tu-ilmenau.de

Abstract— We present a neural network architecture for gesture-based interaction between a mobile robot and its user, thereby spanning a bridge from the localisation of the user over the recognition of its gestural instruction to the generation of the appropriate robot behavior. Since this system is applied under real-world conditions, especially concerning the localisation of a human user, some proper techniques are needed which have an adequate robustness. Hence, the combination of several components of saliency towards a multi-cue approach, integrating structure- and color-based features, is proposed. At the moment, the gestures itself are very simple and can be described by the spatial relation between face and hands of the person. The organisation of the appropriate robot behavior is realised by means of a mixture of neural agents, responsible for certain aspects of the navigation task. Due to the complexity of the whole system, above all we use "standard neural network models", which are modified or extended according to the task at hand. Preliminary results show the reliability of the overall approach as well as the sufficient functionality of the already realised sub-modules.

I. INTRODUCTION AND SCENARIO



Fig. 1. The mobile robot MILVA. Provided with the necessary on-board equipment (68040-VME-system, 2 PC-systems, CNAPS-board, framegrabber) and different sensors (3 cameras, laserscanner, ultrasound and infrared distance measures, bumpers) MILVA serves as the testbed for the human-machine-interaction.

Figure 1 shows our robot platform MILVA (Multisensory Intelligent Learning Vehicle in neural Architecture). A two-camera-system with 7 degrees of freedom (for each

camera pan, tilt and zoom, additional pan for both cameras) serves for the interaction with a possible user and actively observes its operational environment. An additional camera, mounted at the front of the robot, provides the visual information for navigation.

The use of our system as an intelligent luggage carrier, for instance at a railway station or an airport, was chosen as a hypothetic scenario for the following reasons: First, we must take into account the capabilities of our robot which does not have manipulators and can only move itself. Second, the scenario is to naturally motivate a gesture-based dialogue between the user and the serving system. At a railway station with a lot of people and a high amount of surrounding noise a gesture-based dialogue seems to be the only possible way for interaction.

Recently there is an increasing interest in video based interface techniques, allowing more natural interaction between users and systems than common interface devices do. A considerable number of approaches for the design of intelligent and adaptive human-machine-interfaces have been proposed (see for instance [8], [15], [7]).

The superior goals of our research concerning the proposed architecture (GESTIK-project¹) are the highest possible robustness of the intelligent visual interface under highly varying environmental conditions as well as the sufficient organisation of the appropriate robot behavior, achieved by continuous interaction between robot and human user.

II. NEURAL ARCHITECTURE FOR USER LOCALISATION AND GESTURE RECOGNITION

Figure 2 provides a coarse sketch of the neural architecture for user localisation and gesture recognition. The several components of the architecture will be described in the following subsections.

¹Supported by the Thuringian Ministry of Science, Research and Culture (TMWFK, GESTIK-Project)

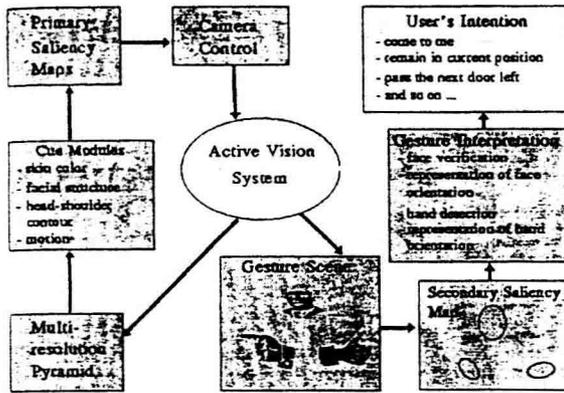


Fig. 2. Building blocks of the neural architecture for user localisation and gesture recognition

A. Multi-cue approach for user localisation

Initially both cameras of the two-camera-system operate in wide-angle-mode in order to cover the greatest possible area of the environment. Multiresolution pyramids transform the images into a multiscale representation. Four cue modules which are sensitive to *skin color*, *facial structure*, *structure of a head-shoulder contour* and *motion*, respectively, operate at all levels of the two pyramids. The utility of the different, parallel processing cue modules is to make the whole system robust and more or less independent of the presence of one certain information source in the images. Hence, we can handle varying environmental circumstances much easier, which, for instance, make the skin color detection difficult or almost impossible. Furthermore, high expense for the development of the cue modules can be avoided (see [4], [3], [11], too).

a) Skin color

For the generation of a skin color training data set, portrait images of different persons (of our lab) were segmented manually. The images were acquired under appropriate lighting conditions (typical for our lab environment).

In order to obtain almost constant color sensation, first we map the RGB color space into a fundamental color space and employ a color adaptation method (see [21]). Then, we return into the RGB color space and define a 2-dimensional Gaussian function via calculation of the mean and the covariance of that skin color data set to model the obtained skin color distribution roughly. Furthermore, if a face region could be verified, a new Gaussian model is created, more specific for the illumination and the skin type at hand. Via this model the

detection of skin colored regions, especially hands, can be improved. This is very important because the hand regions cannot be segmented by structural information (see [13], too).

A more detailed description of our skin color investigations can be found in [5] and [6].

b) Facial structure

In our scenario we assume that a person is an intended user if its face is oriented towards the robot. The detection of facial structure uses the gray value image and employs eigenfaces generated by a principal component analysis (PCA) of the images contained in the ORL data set (<http://www.cam-orl.co.uk/facedatabase.html>; see [19], too). The image regions (15 x 15 pixels) used for the PCA were extracted manually and were normalised by their mean and standard deviation (see also [24], [23]). Then, the input image is processed with 3 eigenfaces (largest eigenvalues). Besides the preprocessing steps, the classification of the obtained fit values remains a difficult problem. The best results we achieved with a supervised Growing-Neural-Gas-Network (GNG, [10]), performing a mapping from the fit values to 2 classes (face, no face). For the training of the GNG a data set of 174 positive (face) and 174 negative (no face) examples was created. To improve the generalisation ability of the network, we implemented a bootstrap algorithm [23] which encloses false classified image regions into the set of the negative examples automatically.

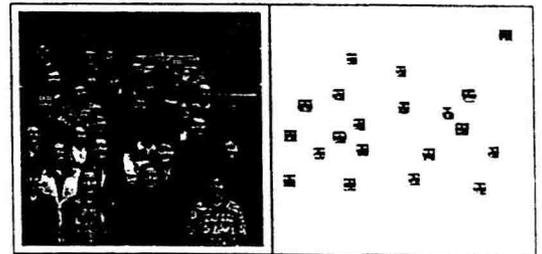


Fig. 3. Detection of frontally aligned faces. The detected faces are marked in the right image (likelihood higher than 0.7).

The performance of the face detection is demonstrated in figure 3, where an image taken from [23] was processed. False positive detected regions cannot be avoided entirely (top right), but this region very likely covers no skin color, and therefore, by combining skin color and facial structure such mislocalisations can be rejected.

c) Head-shoulder-contour

Similar to the detection of facial structure, the localisation of a head-shoulder-contour operates on the gray

level image of each level of the multiresolution pyramids. The basic idea is to use an appropriate spatial configuration of Gabor filters (filter arrangement, see figure 4) and to classify the obtained filter outputs by a specially tuned distance measure (Hamming distance) between the actual filter outputs and a prototype.

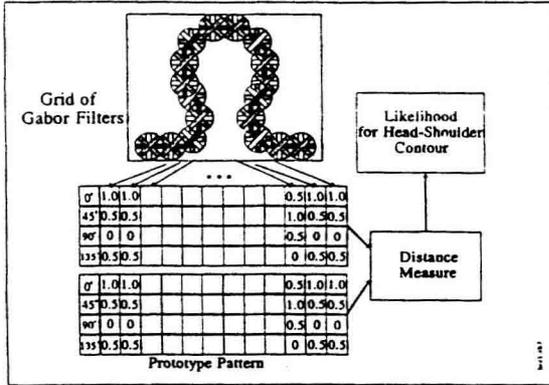


Fig. 4. Processing scheme for detection of a head-shoulder-contour.

d) Motion

Our favoured approach was proposed in [2] and [9]. Based on image differentiation motion is detected in the first step, leading to a binary motion energy image. The second step accumulates this motion energy over a certain period of time resulting in a motion history image. This approach is reliable especially for the following reason: The detection as well as the accumulation of motion could be realised via dynamic neural fields, and by means of different sets of parameters of such fields, different task specific aspects of motion information can be obtained.

e) Dynamic neural fields for generation of primary saliency maps

All cue modules supply input for the primary saliency maps at each level of the multiresolution pyramid, as shown in figure 5.

To achieve a good localisation a selection mechanism is needed to make a definite choice. This is not limited to a two-dimensional position. Since we use five resolutions (fine to coarse) we actually can localise persons even in different distances. Therefore, a neural field (array) for selection of the most salient region should be three-dimensional.

Those fields can be described as recurrent nonlinear dynamic systems (cf. [1], [14]). Regarding to the selection task we need a dynamic behavior which leads to

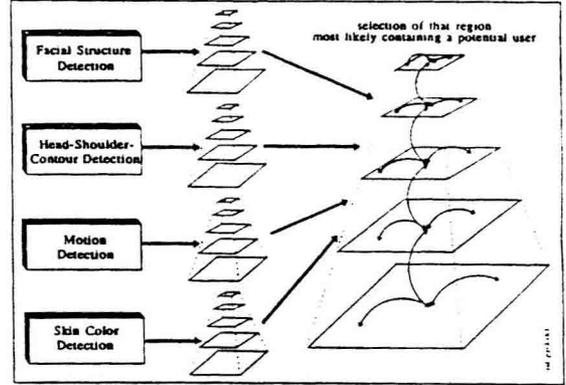


Fig. 5. Generation of a scale space pyramid of primary saliency maps

one local region of active neurons successfully competing against the others, i.e. the formation of one single blob of active neurons as an equilibrium state of the field. The following equations describe the system:

$$\tau \frac{d}{dt} z(\vec{r}, t) = -z(\vec{r}, t) - c_h h(t) + c_l \int_R w(\vec{r} - \vec{r}') y(\vec{r}', t) d^2 \vec{r}' + c_i x(\vec{r}, t) \quad (1)$$

$$\text{with } w(\vec{r} - \vec{r}') = 2 \exp\left(\frac{-3|\vec{r} - \vec{r}'|^2}{2\sigma^2}\right) - \exp\left(\frac{-|\vec{r} - \vec{r}'|^2}{\sigma^2}\right) \quad (2)$$

$$y(\vec{r}, t) = \frac{1}{1 + \exp(-z(\vec{r}, t))} \quad \text{and} \quad (3)$$

$$h(t) = \int_R y(\vec{r}'', t) d\vec{r}'' \quad (4)$$

Herein $\vec{r} = (x, y, z)^T$ denotes the coordinate of a neuron, $z(\vec{r}, t)$ is the activation of a neuron \vec{r} at time t , $y(\vec{r}, t)$ is the activity of this neuron, $x(\vec{r}, t)$ denotes the external input, $h(t)$ is the activity of a global inhibitory interneuron, $w(\vec{r} - \vec{r}')$ denotes the function of lateral activation of neuron \vec{r} from the surrounding neighbourhood R . Further, τ is the time constant of the dynamical system and σ is the deviance of the gaussians determining the function of lateral activation. For the computation we used the following values for the constants: $c_h = 0.025$, $c_l = 0.1$, $c_i = 0.1$, $\sigma = 2$ (halved z-direction), $\tau = 10$ with $\Delta T = 1$ (ΔT : sampling rate). The range R of the function of lateral activation reaches over 5 pixels and 3 pixels in z-direction, respectively (anisotropic neighbourhood).

The results of the systems shall be qualitatively illustrated in figure 6. The presented results are exem-

plary, the usage of the shape of contour provides one solution for the person localisation problem, even under quite different conditions. In our ongoing work, the same principle is extended to the whole saliency pyramid, integrating all cue modules. The novel approach with a three-dimensional dynamic neural field can be assessed as robust method for the selection process, very reliable for the task at hand.

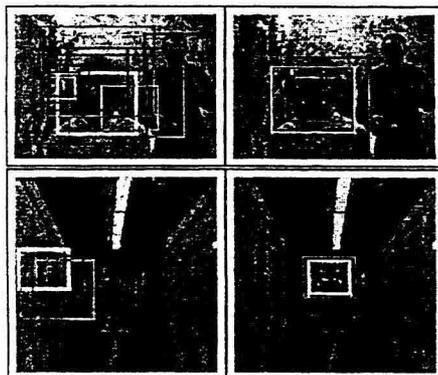


Fig. 6. Input images with marked head-shoulder-contours, obtained at the different levels of the multiresolution pyramid by the proposed method; *left*: without dynamic selection; *right*: by means of dynamic selection (white rectangles mark the highest likelihood).

B. Control of the two-camera-system

A camera control module, based on a neural approach proposed in [22], was extended for the control of the two-camera system. The basic idea is that a definite configuration of the cameras is assumed. Therefore, after a possible user (face region) was located in either camera image, the second camera is directed towards this user. This is realised by means of controlling the pan/tilt of this camera as well as the additional pan for both cameras. Therefore, the initial camera configuration (especially the base distance) remains stable.

As soon as a possible user (face region) is detected in one of the camera images, this camera serves as *general-view-camera*, whereas the second camera becomes the *gesture-camera*. The necessary distance estimation is provided by the cue modules detecting structural information (face and head-shoulder-contour). The resulting *gesture-scene* should contain all gesture-relevant parts of the intended user. Furthermore, the *gesture-camera* is controlled such that the expected face region will appear on a predefined position in the image with a predefined scale, too. Hence, we do not have to ensure scale invariance by the further processing steps.



Fig. 7. Possible intuitive gestures (poses); from left to right they could carry the following meanings for the robot: hello, stop, move right

C. Detection and interpretation of gestures

a) Definition of a gesture set

For complexity reasons, we have predefined a gesture alphabet and have assumed only static gestures (poses), which are stable for a certain period of time (see fig. 7). The mapping between the gestures to be recognized and the associated actions of the robot is predefined, too (see also [15]).

b) Generation of the secondary saliency map

A secondary saliency map is created for the gesture-scene, which determines the sequential processing of this scene. Similar to the primary saliency map we utilise topographically organised neural fields.

To simplify the task, we mainly employ the skin color information as the input for this field, thereby assuming that the skin color segmentation is robust enough.

Because of the camera control, the prominent position and size of a hypothetic face region is known. So, by means of specially tuned field parameters (coupling width and strength) the emergence of an activity blob covering the face region is highly supported. Therefore, the face region will be the first area to be analysed in detail (see the following section). The hand regions become salient, too.

c) Face verification and representation of gestures

The next processing step must provide a face verification, that means we have to decide if there is a face at all, and if it is oriented towards the robot.

To obtain this information, a very simple method for direct mapping of grey value image parts to the corresponding object orientation (both, faces and hands), based on a MLP network, was tested. Preliminary results show the sufficient functionality of such an approach under certain constraints (unstructured background). At present, the approach is examined under real world conditions. Furthermore, the detailed analysis of faces and hands via a regular grid of Gabor filters and a following classification of the Gabor filter outputs with a neural classifier (see also [17], [18]), will be taken