Measuring motion expressiveness in wheeled mobile robots

João Sequeira

Instituto Superior Técnico / Institute for Systems and Robotics
jseq@isr.ist.utl.pt

Abstract. This paper addresses the measurement of motion expressiveness in wheeled mobile robots.
A neural network based supervised learning strategy is proposed as a method to fuse information obtained from the measurement of selected features. The choice of these features is made to reflect the visual quality of the trajectory and hence carries semantic ambiguities that are filtered out through the ability to generalize knowledge by the neural network.
The paper presents results with two features that might be significant in what concerns motion expressiveness, namely, how confident/hesitant is the motion and whether or not contains local loops that might indicate, for example, a call for attention by the robot towards a group of humans.

1 Introduction

Motion expressiveness (ME) is a concept for which most humans can provide acceptable definitions, all of them with close semantics. Roughly, an expressive motion triggers some kind of emotion in people observing it.

In everyday life, human and animal societies use individual motion to express multiple behaviors and emotions, e.g., aggressiveness, anxiety, attention, autistic, curiosity, dominance, egotistic, love, neutral, submissive, etc. In mobile robotics contexts a natural ME definition can be extrapolated directly from the human locomotion context, that is, ME is an index that expresses the ability of a trajectory executed by a robot to convey a meaning to a human observer. Some locomotion behaviors by humans have clear and socially accepted meanings and hence are readily classified by people as expressive.

Expressiveness is a common concept in Information Sciences and software engineering, namely in interface analysis and design. Usability metrics often include expressiveness measures along with measures of other concepts such as concision, simplicity, transparency, and scriptability (see for instance [1]). Some authors also argue that there is a tradeoff between usability and expressiveness, [2,3], which could then provide alternative approaches to ME measuring.

Image processing techniques have been used to identify dynamic models of humans and animals using motion capture from video images (see for instance [4] for an application related to the generation of realistic movements by animation creatures). Sequences of expressive motions, chosen according to some classification criteria, can thus be processed to identify the relevant features of the models and the corresponding control requirements such that expressive motion can be replicated by robotic models. As an example, motion capture combined with frequency decomposition techniques and discriminant analysis is used in [5] to obtain Laban parameters for motion and classify the intensity of dance movements.

Nevertheless, most people can identify an expressive motion when seeing it just by observing and estimating a number of features, often unconsciously. Roughly, a motion qualifies as expressive when it conveys a meaning to an external observer. Some locomotion behaviors by humans have clear, socially accepted, meanings and hence qualify as expressive motion.

Assessing the ME of an anthropomorphic robot tends naturally to be easier as the motion is immediately compared with the corresponding human capabilities. In robotic heads the combined motion of neck, eyes, eyebrows and mouth easily induces an emotional response in external observers (see for instance, [6,7]). For legged locomotion the current state of the art in biped robotics has still to improve to reach the same degree of expressiveness of human walking. Similarly for manipulation through anthropomorphic robot arms.

Psychology identifies motion related features and the emotions they trigger in external observers. The Arouse-Valence-Stance space is used to represent EM in face movements, [7], also can also be used to map the basic emotions into real motion of robotic heads.

In wheeled mobile robots, ME has been mostly related to behavioral control of teams of robots, e.g., flocking and foraging (see for instance [8]). Expressive motion by single robots has been addressed also with the help of anthropomorphic features, [9,10], hence masking the effect of pure motion with the motion of the body of the robot itself.

Some results refer that people tend to prefer "machine-like appearance, serious personality and round shape", [6]. This argument supports ME analysis directly based in the estimation of relevant features in the trajectories performed by the robots.

In information sciences well defined models for expressiveness have been proposed. The weighted linear combination,

$$E = c_1 e_1 + c_2 e_2 + c_3 e_3 + c_4 e_4 \tag{1}$$

has been proposed in [11]. The c_i are constant weights, e_1 is estimated as the number of data elements in an information system, e_2 supports high-resolution concepts by allowing the user to distinguish between entities when the differences are very small and it is estimated from the fan-out of the entities at various levels in a data model, e_3 is estimated as the count of synonyms (multiple synonyms for the same entity increase the probability that the system can support users from different backgrounds where different terminology is used to express the same concept), and e_4 is estimated as the number of query types supported by the information system.

The e_i components in (1) are of course specific to information systems but the linear structure suggests a linear independence/decoupling assumption between them that is worth to explore in the robotics context.

A meaningful ME measure, that is a measure that can provide coherent information on ME, can be used to influence motion control. For instance, if the goal behavior is to socialize with a group of people, approaching them with a confident motion, and, due to obstacles in the way, the real time evolution of the ME measure shows a hesitant motion, then the motion parameters might need an adjustment to better reflect the goal behavior. This is the ultimate goal for this study.

2 A neural network based classifier

Within the human activities motion expressiveness is an ill-defined concept. Each person has a personal idea on what an expressive motion is, relying on the observation of a number of features chosen according some personal, often unconscious, criteria. The estimates for the values of these features are combined, often using ill-defined rules, to yield an estimate of a personal expressiveness index. Still, most people can identify an expressive motion when seeing it just by observing and estimating those features and associating this to a meaning.

Given the ambiguous nature of an ME index it seems appropriate to perform its estimation using supervised learning. This means that the values obtained are correct up to the human reference knowledge and training data used during the learning stage.

The architecture of the proposed ME estimator is shown in Figure 1. Feature extractors map the observations into a real number that expresses the strength of the feature. Features can be grouped according specific contexts to yield ME measures valid in these contexts Each context might be interpreted as corresponding to some specific perspective to estimate ME. For example, in an active surveillance application, using robots, it may be useful to distinguish between ME in normal and threat situations. The ME measures at the output of each context can then be further combined to obtain a more general index.

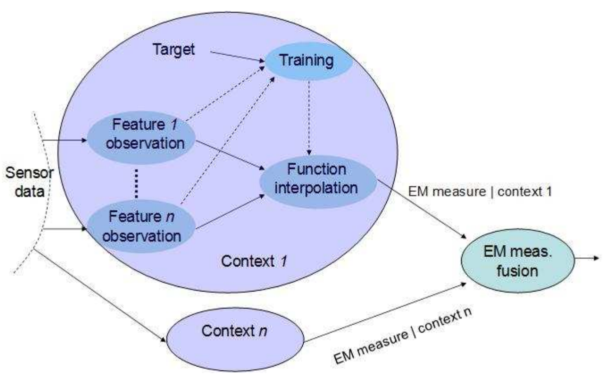


Fig. 1. The classifier architecture

The supervised learning approach in Figure 1 is a classical architecture. The function interpolation combines the feature values at the input such that the output yields a priori known target values for the specific input values that form the training set. Features can be grouped such that they are identified with particular situations, or contexts. ME indexes obtained for each of these contexts can then be fused to yield a global index.

As aforementioned, the selection of adequate features is quite an arbitrary process as it depends on personal feelings that induce subjective decision criteria. In what concerns wheeled mobile robot trajectories, there are a number of features that humans most likely use to qualify a trajectory as expressive. Among these, one can identify,

- the low and high frequency components of the trajectory, expressing the amount of smoothness or sharp movements (low frequency components in angular velocity measures a tendency for a trajectory to contain loops);
- the distance to objects or people, expressing the comfort or distress in the interaction;
- the maximum and minimum linear and angular velocities, expressing anxiety;
- the number of changes in the signs of the linear and angular velocities, expressing hesitation or confidence in the movement;
- the direction of the motion, expressing an intention;
- the space spanned by each of the coordinates, expressing how active a robot is (the larger the span the bigger the activity of the robot; it may be interpreted as having a robot deeply committed to complete survey the space);
- the entropy, expressing a measure of the organization in the trajectory (an entropic trajectory might induce feelings of low confidence or high anxiety).

The problem of combining the values of all the features in a context can be set using multiple approaches. If each feature observed yields a random variable with some probability distribution then the fusion of multiple features can be seen as the solution of combining these variables to minimize the variance of the estimation error between pre-specified target outputs and the outputs obtained for a priori known inputs. Under assumptions of gaussian input disturbances and knowledge on the feature dynamics, the Extended Kalman Filter (EKF) can be used to fuse all the feature measurements. However, features will often behave according nonlinear and nonsmooth models, difficult to identify, and hence such techniques might not be directly applicable to this problem.

A supervised learning technique based on neural networks is used in this paper. The motivation for this approach is drawn mainly from the fact that (i) standard data fusion structures, namely the EKF, also contain a linear combination of error inputs, and (ii) the model (1) also suggests that a linear structure be used to combine the values of the features.

The typical structure of a neural network includes a linear combination stage and, depending on the specific type of network, some additional elements that can shape the dynamics to arbitrary nonlinear models. Elman neural networks, [12], are characterized by the inclusion in the linear combination stage of past output values. Figure 2 shows the building blocks that form one cell of an Elman network. At each cell, the inputs x are weighted by W_i and added a bias, b, and the output, y, weighted by W_c. At the output stage , f, multiple functions can be used, e.g., linear and tansig. When using linear

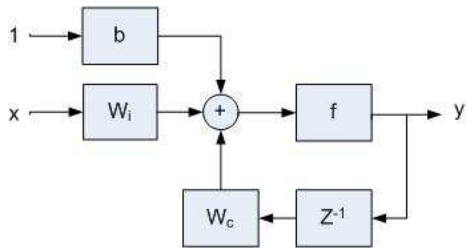


Fig. 2. An Elman network cell

output f blocks, each cell is a first order dynamic system and hence an Elman network is a mesh of first order systems. By establishing adequate interconnections among the cells and shaping the weights W_i and W_c it is possible to create arbitrary linear systems and approximations of nonlinear systems. Elman networks have been used in speech recognition, [13], plant monitoring, [14], and nonlinear dynamics identification, [15]. Multiple learning algorithms can be used to compute the W weights. In this work a standard backpropagation rule is used.

Defining the proper network topology for a given problem is in general a difficult issue. Some theoretical results point to two hidden layers when the network is to model a generic boolean function, [16]. Evolutionary programming has been used to determine the number of hidden cells in neural networks, [17,18].

3 Experiments

In this section two experiments are presented. The features used in these experiments were chosen aiming at recognizing hesitating-confident motion, in the first experiment. In the second one, the features chosen aim at measuring the curiosity by the robot towards a person or a group of persons. In both experiments the a Pioneer robot is teleoperated in a number of independent runs. A subset of these runs is used to train the neural network. The remaining runs are used to assess the performance of the classifier and its ability to generalize the knowledge acquired during the training stage.

During the training stage, an ME value is assigned to each of the trajectories by a human based on its visual qualities. An adequate visual quality might trigger the desired emotion.

In the first experiment the network has size $[30, 50, 10, 1]$. The second experiment was carried out in two versions, (i) the network size is $[5, 10, 10, 1]$ in the first version and $[70, 150, 100, 1]$ in the second one. In both cases the number of cells in each layer was empirically chosen (relatively easy to find in the first case, it is a single variable monotonic function; more difficult in the second case).

The robot used is of unicycle type. The input data for the ME measure is formed by the linear and angular velocities used to control the robot. For the first experiment the feature chosen was the number of changes in the sign of the linear velocity. In the second experiment the feature to measure the ME is the existence of localized loops in the trajectory. In the first version of the second experiment, the feature is measured by the low frequency components in the angular velocity that generates the trajectory. If a number of the low frequency components present in the angular velocity are higher than those corresponding to high frequencies then the trajectory tends to loop. The relative weight among these components can be used to separate between localized loops and loops spanning large areas. In this case, a 512 point FFT is used to get the components corresponding to the three lowest frequencies, followed by a normalizing transformation. For the second version, the low frequency components of the linear velocity are also used. Figures 3 and 4 show the complete data set (of trajectories) used in the experiments.

Table 1 shows the data for both experiments, the sets used to train the network, the target output for each training set, and the obtained output for each test set. The ME index values, i.e., the output of the network, during the training stage are chosen from the visual observation of the trajectories exclusively.

	Experiment 1			Experiment 2		
Set	Feature f_v	Target	Output	Feature f_ω	Target	Output
(a)	8	0.2	0.1946	8.38419, 9.18992, 10	0.1	0.0976
(b)	17	0.15	0.2029	1.26535, 3.95569, 10	0.1	0.0753
(c)	55	0.9	0.8732	2.35201, 7.09803, 10	0.1	0.1207
(d)	20	0.35	0.2873	10, 0.52747, 0.739253	0.1	0.2212
(e)	54	0.85	0.8619	10, 0.874886, 0.197938	-	1.0610
(f)	54	0.85	0.8619	10, 2.75835, 0.40712	-	1.0634
(g)	35	0.6	0.6016	10, 0.0973174, 0.217039	1	1.0209
(h)	30	0.5	0.5155	10, 0.241414, 0.37905	1	0.8405
(i)	6	-	0.2744	10, 8.12918, 4.15008	-	-0.1817
(j)	13	-	0.1181	10, 3.7375, 0.0230103	1	1.1500
(k)	16	-	0.1762	10, 7.51246, 2.19928	-	-0.2317
(l)	15	-	0.1521	10, 7.51036, 2.24441	-	-0.2342
(m)	34	-	0.5853	10, 1.64281, 0.0744162	-	1.1585
(n)	52	-	0.8386	0.9487, 10, 0.263107	-	0.0853
(o)	15	-	0.1521	10, 8.077, 3.38752	-	-0.2012
(p)	10	-	0.1315	10, 0.0152748, 0.168055	1	0.8770

Table 1. Data and results for experiments 1 and 2

In experiment 1, the network output (the ME index) increases with the "hesitation" of the robot. The target ME values were assigned based uniquely on the visual observation of the smoothness of the trajectories. This means that the plots hide those changes that are not associated to changes in motion direction. Given the target ME index values, this experiment amounts to identify an almost monotonic function of a single variable. This is a simple task for an Elman network. The results show that the network is able to properly identify such function despite the disturbance introduced by sets (a) and (b).

In what concerns experiment 2, the network is able to identify a function that returns values close to 1 when there are localized looping characteristics in the trajectory. Lo-

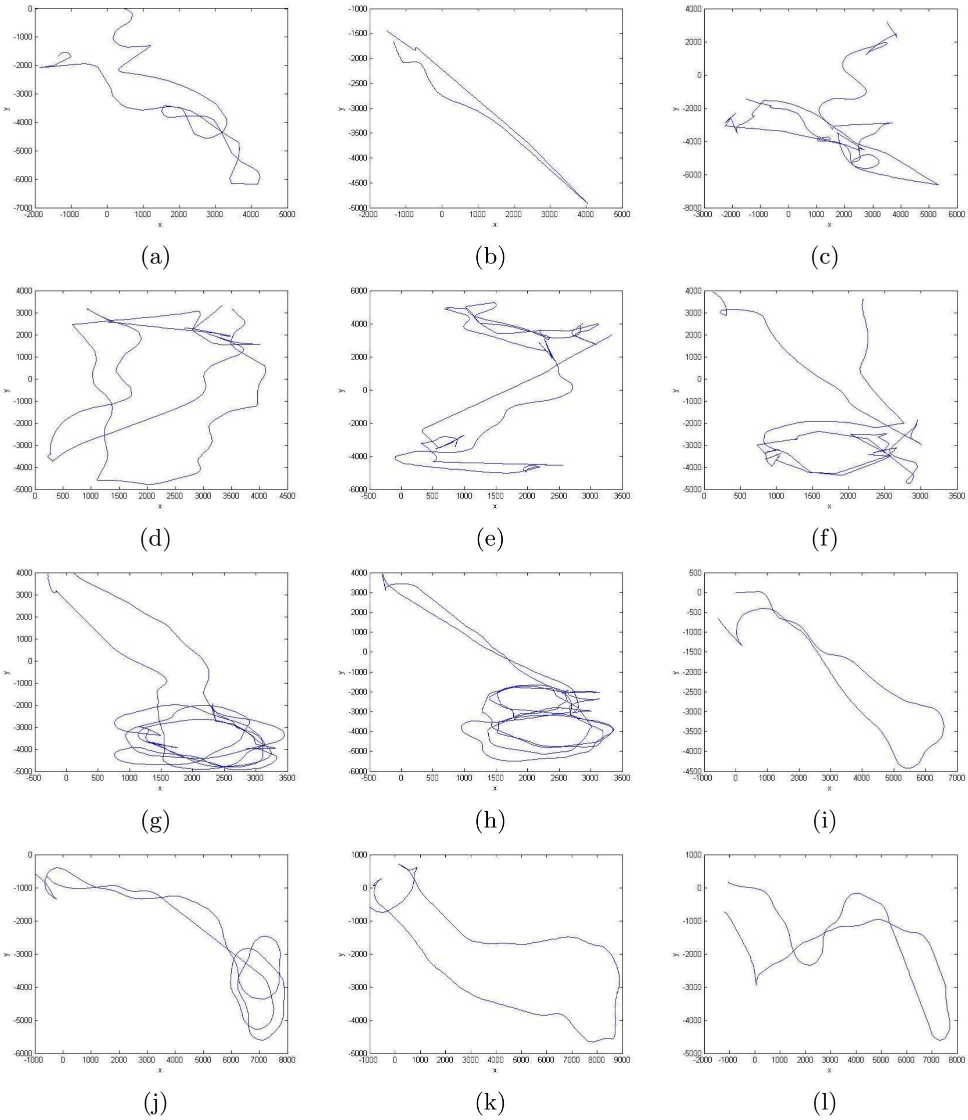
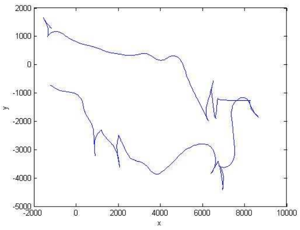
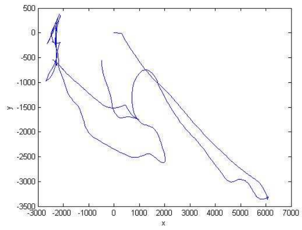


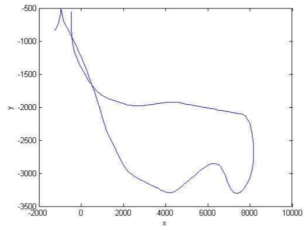
Fig. 3. Complete set of trajectories generated by the experiments



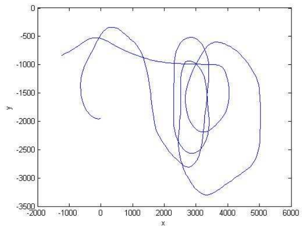
(m)



(n)



(o)



(p)

Fig. 4. Complete set of trajectories generated by the experiments (cont.)

calized loops such as those shown in Figure 4f-g might have been generated by a robot moving around a group of people trying to catch their attention. Loops covering large areas, such as those shown in Figure 4d, are more likely to be related to exploratory other than curiosity behaviors.

The net output indicates that sets (e), (f), (g), (h), (j), (m), and (p) contain localized loops. However, the features selected are not enough to define a one-to-one map between the observations space and the ME index space. The result obtained with the set (m) clearly does not correspond to the expectations created by the visual observation of the trajectory. This happens because the corresponding feature values are close to those of sets (f) and (j) which were taught as returning a high ME value. Also, the network fails on set (o) as its feature values match some of those used for training that do not contain any local loops, such as (k) and (l). Set (e) clearly contains localized loops and, though it is not clear from the visual observation what was the intention of the robot, the network classified it has having loops. Eventually, for some data sets these problems can be solved by carefully tuning the network topology and/or using additional frequency components of the angular velocity at the input of the network.

More generally, components from both the linear and angular velocity are required (as they define completely the characteristics of a trajectory for a given robot model). Table 2 shows results using the 3 lowest frequency components of both the linear and angular velocity with network topology $[70, 150, 100, 1]$. Clearly, there is a better discrimination in the sense that higher values are associated with localized loops better defined. Set (m) was now dropped and sets (o) and (p) are classified as close (as they are given close ME index values) which makes some sense as they contain loops that span wider areas than those shown in the taught sets.

Set	Feature f_ω	Feature f_v	Target	Output
(a)	8.38419, 9.18992, 10	10, 8.14684, 3.71772	0.1	0.1063
(b)	1.26535, 3.95569, 10	0.101767, 2.24074, 10	0.1	0.0895
(c)	2.35201, 7.09803, 10	10, 5.58878, 4.30554	0.1	0.1061
(d)	10, 0.52747, 0.739253	10, 1.24861, 0.105222	0.1	0.2138
(e)	10, 0.874886, 0.197938	10, 6.32206, 1.29682	-	0.4349
(f)	10, 2.75835, 0.40712	6.00649, 10, 1.35242	-	1.0180
(g)	10, 0.0973174, 0.217039	0.0598811, 10, 1.63062	1	0.9394
(h)	10, 0.241414, 0.37905	2.04612, 10, 3.77873	1	1.0432
(i)	10, 8.12918, 4.15008	0.770519, 4.64428, 10	-	0.3285
(j)	10, 3.7375, 0.0230103	2.04532, 10, 0.499195	1	1.1237
(k)	10, 7.51246, 2.19928	10, 5.39222, 0.405971	-	0.3682
(l)	10, 7.51036, 2.24441	10, 7.44558, 0.939264	-	0.5930
(m)	10, 1.64281, 0.0744162	2.74215, 3.30769, 10	-	0.4964
(n)	0.9487, 10, 0.263107	8.11451, 0.0382724, 10	-	0.4113
(o)	10, 8.077, 3.38752	7.89761, 10, 3.39284	-	0.7089
(p)	10, 0.0152748, 0.168055	10, 1.96887, 4.92446	1	0.7861

Table 2. Experiment 2 with additional inputs

4 Conclusions

The paper addressed the measurement of the expressiveness of trajectories described by wheeled mobile robots. Encouraging results on mapping robot movements into a space that can be identified with human idea of motion expressiveness were obtained with the supervised learning approach described in the paper.

The paper addressed exclusively the direct problem, that is, computing a ME index from the observation of trajectories. Still to address is the inverse problem of generating a trajectory with a given ME index and accounting for the robot and environment motion constraints. Also, including ME measures in standard motion control strategies will require the selection of the adequate temporal portions of a signal from where to extract features.

This study addressed only the direct problem of measuring a ME index. Future work will address the inverse problem, that is generating motion that corresponds to some ME index. Also, different ME values obtained from different features will exhibit some overlapping, i.e., comparable ME values will be obtained for identical data sets, whereas very different ME values can obtained for close data sets. By combining multiple such ME measures it might be possible to further improve the discrimination among data sets. The technique can be applied to all sorts of robots and hence naturally leads to challenging problems related to the fusion of ME measures taken from multiple robots, e.g., mobile manipulators and teams of robots.

4.1 Acknowledgments

This work was supported by the European Project FP6-2005-IST-6-045062-URUS, and ISR/IST plurianual funding through the POS_Conhecimento Program that includes FEDER funds.

References

1. Raymond, E.: The Art of Unix Programming. Addison-Wesley (2003)
2. Buchanan, G.: Expression and Interpretation in Spatial Hypertext. In: Procs of the 4th Workshop on Spatial Hypertext, ACM Hypertext Conference. (2004) Santa Cruz, USA.
3. Repenning, A., Ambach, J.: Tactile Programming: A Unified Manipulation Paradigm Supporting Program Comprehension, Composition and Sharing. In: Procs. 1996 IEEE Symp. of Visual Languages. (1996) 102–109 Boulder, CO, Computer Society.
4. Cheng-Yun Karen Liu: Towards a Generative Model of Natural Motion. PhD thesis, University of Washington (2005)
5. Alyssa Lees: Expressive Motion. PhD thesis, New York University (2006)
6. Fong, T., Nourbakhsh, I., Dauntenhahn, K.: A survey of socially interactive robots. Journal of Robotics and Autonomous Systems **42** (2003) 143–166
7. Mayor, L., Jansen, B., Lorotte, A., Siegwart, R.: Improving the Expressiveness of Mobile Robots. In: Procs. of ROMAN2002. (2002)
8. Ngo, T., Schioler, H.: An Approach to Sociable Robots through Self-distributed Energy. In: Procs. of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems. (October 2006) 2192–2199
9. Scheef, M., Pinto, J., Rahardja, K., Snibbe, S., Tow, R.: Experiences with Sparky, a social robot. In: Procs. of the Workshop on Interactive Robot Entertainment. (2000)
10. Mizoguchi, H., Sato, T., Takagi, K., Nakao, M., Hatamura, Y.: Realization of Expressive Mobile Robot. In: Procs. IEEE Int. Conf. Robotics and Automation. (1997)
11. Ceruti, M., Robin, S.: Infodynamics: Analogical analysis of states of matter and information. Int. Journal on Information Sciences **177** (2007) 969–987
12. Elman, J.: Finding structure in time. Cognitive Science **14** (1990) 179–211
13. Rothkrantz, L., Nollen, D.: Speech Recognition Using Elman Neural Networks. In: Procs. of Text, Speech and Dialogue: Second International Workshop, TSD'99. Volume 1692/1999 of Lecture Notes in Computer Science., Springer Berlin / Heidelberg (1999) Plzen, Czech Republic, September.
14. Seker, S., Ayaz, E., Türcan, E.: Elman's recurrent neural network applications to condition monitoring in nuclear power plant and rotating machinery. In: Engineering Applications of Artificial Intelligence. Volume 16. Elsevier (October-December 2003) 647–656
15. Gao, X., Gao, X., Ovaska, S.: A Modified Elman Neural Network Model With Application to Dynamical Systems Identification. In: Procs. IEEE Int. Conf. on Systems, Man, and Cybernetics. Volume 2. (1996) 1376–1381 Beijing, China.
16. Brightwell, G., Kenyon, C., Paugam-Moisy, H.: Multilayer neural networks: one or two hidden layers? In: Procs. NIPS '96. (1996)
17. McDonnell, J., Waagen, D.: Determining Neural Network Hidden Layer Size Using Evolutionary Programming. In: Proc. World Congress on Neural Networks. Volume 3. (1998) 564–567
18. Pengt, K., Ge, S., Went, C.: An Algorithm to determine the neural network hidden layer size and weight coefficients. In: Procs. 15th IEEE Int. Symp. on Intelligent Control, ISIC 2000. (2000) Rio, Patras, Greece, July 17-19.