Combining Neural Networks for Gait Classification

Nigar Sen Koktas[1], Nese Yalabik[2], and Gunes Yavuzer[3]

[1] Informatics Institute, METU, Ankara, Turkey
nigar@ii.metu.edu.tr
[2] Computer Engineering Department, METU, Ankara, Turkey
yalabik@ceng.metu.edu.tr
[3] Department of Physical Medicine and Rehabilitation, Ankara University Medical School, Ankara, Turkey
gyavuzer@ankara.medicine.edu.tr

Abstract. Gait analysis can be defined as the numerical and graphical representation of the mechanical measurements of human walking patterns and is used for two main purposes: human identification, where it is usually applied to security issues, and clinical applications, where it is used for the non-automated and automated diagnosis of various abnormalities and diseases. Automated or semi-automated systems are important in assisting physicians for diagnosis of various diseases. In this study, a semi-automated gait classification system is designed and implemented by using joint angle and time-distance data as features. Multilayer Perceptrons (MLPs) Combination classifiers are used to categorize gait data into two categories; healthy and patient with knee osteoarthritis. Two popular approaches of combining neural networks are experimented and the results are compared according to different output combining rules. In the first one, same set is used to train all networks and afterwards the features are decomposed into five different sets. These two experiments show that using entire data set produces more accurate results than using decomposed data sets, but complexity becomes an important drawback. However, when a proper combining rule is applied to decomposed sets, results are more accurate than entire set. In this experiment sum rule produces better results than majority vote and max rules as an output combining rule.

1 Introduction

Gait analysis is the process of collecting and analyzing quantitative information about walking patterns of people and it is important for developing treatment plans or tracking the improvement of persons having gait problems (i.e. Parkinson, cerebral palsy, arthritis). This process is facilitated by the use of computer-interfaced video cameras to measure patient motion, by the use of electrodes placed on the surface of the skin to appreciate muscle activity, and by the use of force platforms imbedded in a walkway to monitor the forces and torques produced between the patient and the ground. After collecting data the essential part of the process is the interpretation of these by experts and related software. Gait analysis, when considered as an automated system, is used for two main purposes: human identification and clinical applications.

J.F. Martínez-Trinidad et al. (Eds.): CIARP 2006, LNCS 4225, pp. 381 – 388, 2006.

Human identification is an important security issue. In most cases it is not so easy to determine the identity of the person but many applications work well for some special cases, such as gender classification [1], age classification [2] etc. In most of these studies data sources are image sequences, but it is inappropriate to measure joint and segment gait kinematics directly from the videotape or monitor. They do not give an indication of the cause of the gait abnormality and so have limited value in clinical decision-making.

The application of automatic gait analysis in medicine is also a well-studied subject. There are studies that have shown that the number of surgical procedures is reduced after a three-dimensional (3-D) gait analysis [3]. In medical applications measurements are obtained more sensitively. Kinetic and kinematical temporal changes are obtained from the subject. In addition to temporal changes of joint angles and force data, time-distance parameters of the gait such as velocity, cadence, stride length, step length are recorded.

The outcome of musculoskeletal diseases can be followed in two ways, as shown in Figure 1. The first one is the traditional method, and the second one is the gait analysis method. Traditional diagnosis starts with the examination of patients according to their complaints. Two mostly used and expensive traditional methods are based on determining cartilage damages. The first one is MR technique which determines the degree of damage. The second one is the determination of cartilage damage by blood and urine analysis. Since these traditional methods are expensive and harmful to subjects to some degree, they are not suitable for frequent long term follow-up of the patients.

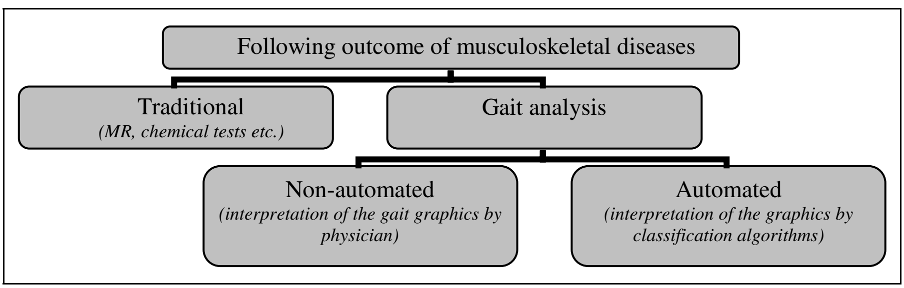

Fig. 1. The role of gait analysis in clinical context

Gait analysis helps the physician in both the diagnosis and follow-up non-automatically or automatically. If the physician him/herself interprets the obtained gait data, then this is a non-automated diagnosis. But if this data is interpreted by software, then it is called automated diagnosis. Since non automated diagnosis requires high level of expertise, only specifically trained orthopedists or physiatrist can use gait data. Automated system is expected to decrease this requirement which may help to increase the number of physicians and patients, making use of the laboratory. In addition, automated systems save experts' time and decrease the possibility of human made errors.

Here the word *automated* does not mean that gait systems are tools to replace physicians. Rather, it is an assisting tool to search the diseases that physicians suspected, to determine the level of the already diagnosed diseases, or to examine invisible changes of patients.

For the design of automated diagnoses systems some well known pattern recognition algorithms are used. These are neural networks (NNs) [2, 4, 5, 6, 11], support vector machines (SVMs) [2, 8], and radial basis functions (RBFs) [7]. The use of NNs for experimental gait classification is not new. There are studies in which NNs are trained by force platform data to distinguish 'healthy' from 'pathological' gait [4, 5, 7, 9]. In addition to these, there are studies to recognize walking people among a few subjects (less than 10) by using joint angles as features [6, 10, 12]. These studies produce reasonable results (about %51-%83 on testing set and about %76-%98 on training set) for NNs use in gait classification. Since these studies differ from each other in description of gait variables (such as subject type, measurement tools, type of variables, anatomical levels), and in construction of classifiers, comparing the performances of them with the current one may not be reasonable.

As the dimension of the features and the size of the data increase same accuracies may not be guaranteed. In similar pattern recognition studies this problem is tried to be solved by combining classifiers. Combination of NNs are widely used today especially in speech recognition [14] and character recognition [13] studies and they have showed an increase in the performance of the classifiers. In [16] Sharkey made a comprehensive experiment to compare two different NNs combining methods; modular and ensemble ones. She concluded that using an entire set for training produces more accurate results than decomposing it. In this study comparison of these two approaches are done in the context of gait classification. There are also different approaches on combining outputs of classifiers. Kittler et. al. [17] has comparative studies on efficiency of output combination rules such as majority voting, sum, product, max., and min. rules. In [17], they concluded that sum rule is superior to others in most of the cases.

The objective of this study is to design a software system for physicians for supplying accurate and practical ways to diagnose and further classify a musculoskeletal disease using only gait data. The accuracy of the proposed system will be safeguarded by using all features used for diagnoses. To be able to combine all features in one classification system, combination methods are expected to be most suitable. As our previous studies [11] and similar studies proved MLP usage for gait classification produces reasonable results. A group of MLPs are used to classify the subjects as healthy or sick, using temporal changes of knee joint angle and time-distance parameters as features. Current study is one of the first studies in which classifier combination techniques have been applied to gait data. Two different NNs combination methods are tried. In the first experiment data set is decomposed into five different sets and five MLPs are trained and tested by these sets. Then test set results are combined by sum, majority vote and max rules to produce final class label. In the second experiment, entire data set is used to train three different architectural MLPs and again outputs are combined by three different rules and accuracy rates on test set are compared.

The remainder of the report is organized as follows. Section 2 introduces data collection process and the characteristics of data. This is followed by details of our experiments and results. Finally, in last section conclusions are presented.

2 Characteristics of Data

There are many data collection methods in gait analysis literature. However, some have disadvantages over others because of harmful effects to the subjects. The stereo metric method is the most popular one currently used. It employs visible markers attached to the skin on rigid segments of the body structure and tracks their motion using imaging equipment. This technique is implemented using charge coupled device (CCD) cameras and frame-grabber electronics to allow digital images to be captured as the subject moves within the field of view. Digital image analysis allows the physical location of each marker to be computed, using triangulation of the views from an array of camera systems. This technique has minimal impact on the natural motion of the subject and allows data capture without the need to tether the subject to the data acquisition hardware. But, it is not feasible to measure gait patterns or variability with only one traversal of the instrument walkway. Thus, multiple walking trials need to be collected, which may fatigue the subject.

While data collection techniques for gait analysis have continually evolved over the past 40 years, the method of data presentation has not changed much. The data is still reported in 2-D charts with the abscissa usually defined as the percentage of the gate cycle and the ordinate displaying the gait parameter. Figure 2 shows the graphical representation of the gait data used in this study for both a healthy and with knee osteoarthritis person.

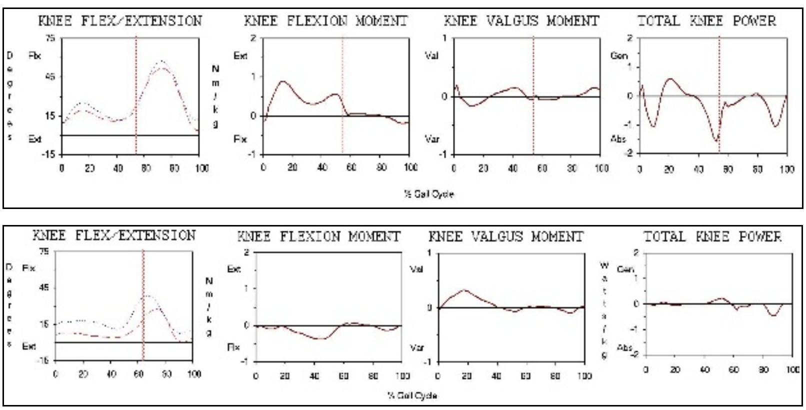

Fig. 2. Graphs of the gait data (*healthy (a) and knee osteoarthritis (b)*)

In this study, data are collected in Ankara University Faculty of Medicine Department of Physical Medicine and Rehabilitation Gait Laboratory by the gait analysis experts. In this laboratory, there are standard gait laboratory equipments which are supported by "VICON" a commercial system for gait analysis. Subject is walked on the platform and in one cycle of gait, temporal changes of joint angles, joint moments, joint powers, force ratios and time-distance parameters are gathered and recorded to database. Decision of which features to use is done according to inspected illness. In this study, Osteoarthritis, a disease that affects knee joints is selected as an example;

therefore, by the advice of the medical expert, knee flexion, knee flexion moment, knee valgus moment and total knee power are selected as the features of the knee joint angle. In addition, walking velocity, single support and step length are selected as the time-distance parameters of the gait.

Each of joint angle related features are represented by a graph that contains 51 samples taken in equally spaced intervals in the time for gait cycle, which is the time spent for one step. These points composed feature vectors which are used as inputs of the related MLP. On the other hand, time-distance parameters are static numerical values which are also used to train a MLP.

Before passing to classification phase data is cleaned by eliminating rows having missing values. Finally, 91 healthy and 110 sick subjects' data is prepared for classification purpose and shared for training and testing purposes as shown in Table 1 (H: healthy, S: Sick, SMP: Samples).

Table 1. Dataset characteristics

<table>
<tr><th rowspan="2">FEATURE VECTOR (FV)</th><th rowspan="2">DATASET</th><th rowspan="2">#SMP.</th><th colspan="2">#TRAIN</th><th colspan="2">#TEST</th></tr>
<tr><th>H</th><th>S</th><th>H</th><th>S</th></tr>
<tr><td>FV1</td><td>KFlex: Knee flexion/extension</td><td>51</td><td rowspan="6">61</td><td rowspan="6">77</td><td rowspan="6">30</td><td rowspan="6">33</td></tr>
<tr><td>FV2</td><td>KMFlex: Knee flexion/ extension moment</td><td>51</td></tr>
<tr><td>FV3</td><td>KMVal: Knee Valgus Moment</td><td>51</td></tr>
<tr><td>FV4</td><td>KPTot: Total Knee Power</td><td>51</td></tr>
<tr><td>FV5</td><td>Time-dist: Velocity, single support, step length</td><td>3</td></tr>
<tr><td>FV6</td><td>Entire set (all of above)</td><td>207</td></tr>
</table>

3 Experiments and Results

The basic classifier structure, used in this study is MLPs combination. Weaknesses of each classifier are diminished by combining classifiers, and more accurate results are expected.

In [15], two methods are described for combining multiple networks. The first one is the modular approach, in which the task is first decomposed into several subtasks and a specialist network is then trained using the inputs pertaining to the corresponding subtask. The second approach is the ensemble one, in which each network is trained using the same inputs and provides a different solution to the same task. Outputs from these networks are combined to reach an integrated result. Complexity is an important issue to be considered in this case. Differentiation among classifiers may be done by using initial random weights, different topologies, and varying the input data.

As stated previously the final data that is used here has five feature vectors; four for temporal changes of knee joint angle (KFlex, KMFlex, KMVal, KPTot) and one for time-distance parameters. Before training all data sets are scaled to interval [-1, 1]. Totally eight MLPs are trained using Matlab neural network toolbox. These MLPs are combined in different schemas for experiment 1 and experiment 2 as shown in

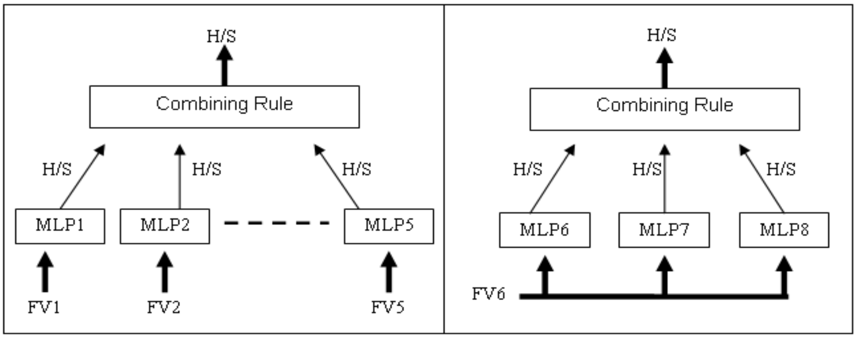

Fig. 3. MLP combination schemas for experiment 1 (a), and experiment 2 (b) H/S: Healthy or sick, FV: Feature vector

Figure3. Table 2 and Table 3 show the topology of each network and their individual success rates on test data. For the first five networks number of hidden nodes and hidden layers are determined experimentally.

Experiment 1: Input data is decomposed in five sets composed of different feature vectors. Five MLPs are trained by these input sets and then outputs of test set are combined by three different combining rules to reach a final result. So, accuracy of different combining rules is compared.

Table 2. Properties of MLPs used in experiment 1

NETWORK	#NODE			#MIS-CLASSIFIED	SUCCESS RATE (%)
	input	*hidden1*	*hidden2*		
MLP1	51	35	10	10	84
MLP2	51	35	10	8	87
MLP3	51	35	10	15	76
MLP4	51	35	10	18	71
MLP5	3	2	-	13	79

Experiment 2: Three different MLPs are trained by using the same composite input set without any decomposition. Here, differentiation of each network is done by different number of hidden layers and hidden nodes.

In both experiments different combination approaches are used, but in both cases combining outputs of classifiers became and important issue. In this study three of these rules, sum, majority vote and max rules, are experimented and results are compared by success rates on test data set.

After training each network with corresponding input set, test data are presented and the outputs are normalized to use them as posterior probabilities. Since *tansig* function is used as the activation function in all layers of networks, outputs are in interval [-1, 1]. To normalize an output, its absolute value is taken as posterior

Table 3. Properties of MLPs used in experiment 2

NETWORK	#NODE			#MIS-CLASSIFIED	SUCCESS RATE (%)
	input	*hidden1*	*hidden2*		
MLP6	207	50	-	6	90
MLP7	207	150	40	6	90
MLP8	207	207	50	7	89

probability, and its sign is taken as class label (i.e minus sign is for normal and plus sign is for sick subject). Then, its 1-complement is recorded as posterior probability of the other class. Thus, sum and max rules for combining outputs can be applied.

For sum rule, created posterior probabilities are added up for two classes and higher value determined the class label. In max rule, the network, producing the maximum of posterior probabilities determined the class label and the others are ignored. To find the majority vote, each networks' output is converted to class labels by applying a threshold and three agreeing classifiers determine the class label of the test datum. Table 4 shows the obtained success rates on test set by applying these combining rules.

Table 4. Success rates (number and percentage) for combining rules

Combining rule	Combined networks			
	MLP1-MLP5		MLP6-MLP8	
	#misclassified	*success rate (%)*	*#misclassified*	*success rate (%)*
Sum	4	94	6	90
majority vote	5	92	6	90
Max	5	92	6	90

4 Conclusion

According to these results, it can be concluded that the best individual performance is produced by MLP6 and MLP7 in which entire data set is used for training and testing purpose. However, as the dimension of the data and relatively network size increase, complexity becomes an important drawback. Since it is difficult to process a large set of data training time increases. However, smaller MLPs which use only one feature vector produce less accurate results and combining their outputs increase the accuracy reasonably.

In addition, combining outputs do not increase the accuracy in experiment 2 as much as in the first one. Increasing the number of networks does not cause any improvement after an optimum number, which is "three" in our experiment.

The combining rules show equal performance in experiment 2, but in experiment 1 sum rule is superior to others. Then, as complexities are considered combining many small networks may be preferred when dealing with large dimensional data.

In further stages of the study, to improve classification accuracy, some feature reduction and/or selection techniques can be tried to reduce the dimension of data and so more features can be included in classification process.

References

1. Lee L., Grimson W.: Gait analysis for recognition and classification, Proceedings of the IEEE Conference on Face and Gesture Recognition, (2002)
2. Beggy R., Kamruzzaman J.: A Comparison of Neural Networks and Support Vector Machines for Recognizing Young-Old Gait Patterns, Proceeding of IEEE TENCON Conference, (2003)
3. Davis R. B., DeLuca P. A., Romness M. J.: Clinical Gait Analysis and Its Role in Treatment Decision-Making", Medscape General Medicine 1-1, (1999)
4. Kohle M., Merkl D., Kastner J.: Clinical gait analysis by neural networks: issues and experiences, Proceedings of the 10th IEEE Symposium on Computer-Based Medical Systems, pp.138, (1997)
5. Wu W., Su F., Cheng Y., Chou Y.: Potential of the Genetic Algorithm Neural Network in the Assessment of Gait Patterns in Ankle Arthrodesis, Annals of Biomedical Engineering, 29, pp. 83–91, (2001)
6. Barton J. G., Lees A.: An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams, Gait & Posture, 5, pp. 28-33, (1997)
7. Kohle M., Merkl D.: Analyzing human gait patterns for malfunction detection, Proceedings of the 2000 ACM symposium on Applied computing, 1, (2000)
8. Begg R. K., Palaniswami M., Owen B.: Support Vector Machines for Automated Gait Classification, IEEE Transactions On Biomedical Engineering, 52-5, (2005)
9. Bertani B., Cappello A., Benedetti M.G., Simoncini L., Catani F.: Flat foot functional evaluation using pattern recognition of ground reaction data, Clinical Biomechanics 14, pp. 484-493, (1999)
10. Lakany H., Hayes G.: An algorithm for recognizing walkers, Proceedings of the International Conference on AVBPA, pp. 112-118, (1997)
11. Anonymous
12. Tanawongsuwan R., Bobick A.: Gait Recognition from Time-Normalized Joint-Angle Trajectories in the Walking Plane, CVPR, 02-2, pp. 726, (2001)
13. Mao J., A case study on bagging, boosting and basic ensembles of neural networks for OCR, Proceeding of IJCNN-98, 3, pp.1828-1833, (1998)
14. Kim J., Ahn J., Cho S.: Ensemble competitive learning neural networks with reduced input dimension, Int J Neural Syst., pp 133-42, (1995)
15. Sharkey A. J. C: On combining artificial neural networks, Connection Science, 8, pp 299-314, (1996)
16. Sharkey A. J. C: Types of multinet system, Lecture Notes in Computer Science, Vol. 2364. Springer-Verlag, Berlin Heidelberg, (2002)
17. Kittler J., Hatef M., Duin R. P.W., Matas J.: On Combining Classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20-3, (1998)
18. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd. ed., Wiley, New York (2001)