Analyzing Music Effects on Brain Waves using Minimal Data based on Functional Measures and Genetic Algorithm

Qingzheng Xu

Research School of Computer Science,
Australian National University
Canberra, Australia
u6174243@anu.edu.au

Abstract. Researchers have found that music can affect brain activity. For example, Rahman et. al. can identify music type by analyzing Electroencephalogram (EEG) and the accuracy can reach up to 97.5% [1]. However, for commercial use, electronic devices (such as virtual reality headsets) cannot collect the same amount of data as electrodes placed along the scalp and it is necessary to find a method to analyze music effects with less data. To achieve this goal, in this paper, a neural network is built to predict the music type among 3 different music genres. Those networks only use data from one electrode placed at the frontal lobe area (F7), which is 1/14 compared to Rahman et. al.'s dataset and the testing accuracy is around 38%. Based on Functional Measures and Genetic Algorithm, five kinds of models are created to find the least and most significant inputs. By pruning the 5 most significant inputs, the testing accuracy can drop from around 38% to 33%; by pruning the 5 least significant inputs, the model performance will not be influenced. By further improving, networks in this paper can predict music types with 38% accuracy using only 6% of Rahman et. al.'s dataset, which also shows the potential of neural network technology in the virtual reality field.

Keywords: Brain Activity, Electroencephalogram, Neural Network, Classification, Data Mining, Functional Measures, Genetic Algorithm

1 Introduction

The relationship between music and brain activity is attractive to researchers in the field of affective neuroscience. Music is supposed to be helpful in the treatment of tension and negative emotions [2]. Moreover, music seems can improve children's reading abilities and mathematical task performances [3]. Rahman et. al. [1] built a classification model which can classify the music type according to brain wave patterns and the accuracy is 97.5%. By analyzing the effects of different kinds of music on brain activity, researchers can identify which music type will have positive effects on human brains and use them in childhood education and psychotherapy. However, to make full use of this technique and generate business value, there are still some gaps. For example, Rahman et. al. was using an electroencephalograph system, but it is not realistic or profitable to use this system in a customer-level product, such as virtual reality headsets. A method of implementing Rahman et. al.'s technique using minimal data can solve this gap.

In this paper, the task is using Neural Networks to classify the music type from 3 music genres based on less data. Features are extracted from 24 participants' frontal lobe EGG provided by Rahman et. al., and the dataset used is only 1/14 of Rahman et. al.'s dataset. Those Neural Networks were trained using error-backpropagation [4] and the initial network topology is 26-40-3. Simple weighted links are used to connect each neuron in a layer to each neuron in the next layer, without backwards connections and multi-layer connections. The activation function is the sigmoid function, and the optimizer is Adam. The initial dataset is spitted into the train, validation, and test set in the ratio of 8:1:1. There are only 576 data points, so Full Batch Learning is applied. With a learning rate of 0.01, the net with the highest validation accuracy will be selected from the first 1000 epochs while the average testing accuracy, precision, recall and F1 score for the chosen net will be calculated as an evaluation method.

To further remove redundant data, models based on Functional measures and the Genetic Algorithm are applied to prune inputs. More information is provided in the Method section.

This paper facilitates the following:

- Automatic analysis of electroencephalograms and identify the behaviours of brain waves under different types of music. This would assist the research in finding the suitable music genre for psychological treatment and the development of children's learning skills.
- The ranking of inputs' importance and suitable inputs selecting. This paper shows and verifies an effective and efficient method for dataset mining. Its limitations will be discussed, too.
- Rahman et. al.'s technique is modified and can be used in customer-level products, which can make business value and promote the development of virtual reality techniques.

2 Method

Pre-processing steps including standardization and transcoding are applied to make the dataset more suitable for training. Because the goal of this paper is to classify music types using less data, feature selection is important. Methods used should be able to remove irrelevant information and do not decrease the model performance.

The Functional measure is a powerful method, which can sort inputs based on their importance according to their similarities to the other input [5][6]. Four kinds of models (I, C, W, U) are created to find the 5 least significant inputs as well as the 5 most significant inputs for further analyse. Modifications have been made on these models and thus this method is well suited for the dataset and it can help achieve the goal of this paper.

Furthermore, in this paper, the dataset only has 26 kinds of inputs and 5 inputs are chosen. If there are more input types, such as n kinds of inputs and choosing k inputs, the time complexity will be O(n ^k), which will cost much time. To solve this problem, the fifth model is created based on the Genetic Algorithm, which further improves the feature selecting ability of created neuron networks.

2.1 Functional Measures

Functional measures can determine the similarity between two hidden neurons over a training set by calculating the angles between activation results vectors of those neurons [6]. Here is its formula.

$$angle\ (i,j) = tan^{-1}\left(\sqrt{\frac{\sum_p^{pats} sact\ (p,i)^2 * \sum_p^{pats} sact\ (p,j)^2}{\sum_p^{pats}(sact\ (p,i) * sact\ (p,j))^2} - 1}\right) \qquad \textbf{(1)}$$

$$\text{where}\ \ sact\ (p,h) = activation\ (p,h) - 0.5 \qquad \textbf{(2)}$$

Gedeon extended this technique, and it can now determine the similarity between two hidden neurons based on the weight matrix [7]. To evaluate the inputs, the technique should be modified to calculating vectors of the weight matrix belongs to different input fields [5]. Here is the new formula for equation (2).

$$\text{where}\ \ sact\ (p,h) = norm\ (weight\ (h)) - 0.5 \qquad \textbf{(3)}$$

In equation (3), the weight matrix is normalized. By subtracting 0.5, about half of the values in the matrix will be positive and the others will be negative, which will lead to better output angles. In this paper, according to the dataset characteristic and model performance, this equation is modified to use standardization instead. Here is the new formula for equation (3), which is used in model W.

$$\text{where}\ \ sact\ (p,h) = stand\ (weight\ (h)) \qquad \textbf{(4)}$$

Besides, this technique can be adjusted to analyse the input data itself as well. In this case, each feature column will be considered as a vector for calculating angles [5]. Here is the new formula for equation (2).

$$\text{where}\ \ sact\ (p,h) = pattern\ (h) - 0.5 \qquad \textbf{(5)}$$

As mentioned before, instead of using normalization, standardization is used for the dataset in this paper. Thus, equation (5) should be modified as well. Here is the new formula for equation (5), which is used in model I.

$$\text{where}\ \ sact\ (p,h) = pattern\ (h) \qquad \textbf{(6)}$$

After some testing on this technique, cosine works better than tangent. When applying tangent, all angles are between 30 to 90 degrees, which is crowded and hard to judge which pair has the largest similarity. Cosine is a common way of finding the degree between vectors. When using cosine, all degrees are distributed between 0 to 180 degree. The difference between pairs is larger and can find similar pairs with higher confidence. Thus, equation (1) is modified:

$$angle\ (i,j) = cos^{-1}\left(\frac{\langle sact\ (i), sact\ (j)\rangle}{\|sact\ (i)\| \times \|sact\ (j)\|}\right) \tag{7}$$

For those angles, if one angle between two input fields is close to 90 degrees, it indicates that these two inputs are less similar to each other [6]. Input pairs are sorted based on their angle distances to 90 degrees in ascending order. In that sorted list, extracting the last 5 unique inputs as the most significant inputs and the first 5 unique inputs as the least significant inputs. When choosing the least important inputs, different from choosing the most significant inputs, only choose one for each pair, because input pair with a small degree indicate they are similar, only one of them should be removed to reduce the dataset duplication. Selected features will be the outputs of model W and I.

For model C (aggregate of I) and U (aggregate of W), instead of sorting input pairs, they are created by sorting the average angle of each input to all the other inputs. This paper will compare the performances of initial networks with the networks that removed the most/least significant inputs.

2.3 Genetic Algorithm

As mentioned previously, data is collected from one electrode and thus the data amount is reduced to 1/14 compared to Rahman et. al.'s dataset. The dataset only has 26 features, but for other problems, it may have more features. Functional measures need to calculate the relationship between every two pair, so it will be much more difficult and time-consumed when the feature number is increasing.

To deal with this problem, the Genetic Algorithm is applied, which is an algorithm inspired by Darwin's theory of evolution and it simulates the process of natural selection. In each iteration, individuals which adapt to the environment the most (get the highest evaluation mark) will survive and produce offspring. Those offspring will inherit their parent's genes and characters. By setting a suitable population size, mutation probability and crossover probability, after sufficient iterations, the best offspring survive, and it can be considered as the optimal solution.

In this paper, when using GA to find the 5 least significant inputs, the DNA size is 5 (5 bits in DNA) and the value bound is from 0 to 25 (26 features). The target function is the sum of degrees between each input pair and the model will try to minimize the output of this target function. The population size is 100, the mutation probability is 0.1, the crossover probability is 0.5 and this model is using Uniform Crossover. The parent portion is 0.3, so after each iteration, there will be 30% parents and 70% offspring. Moreover, the elite ratio is 0.01, so 1% of the best individuals will be retained at each generation. The max iteration without improvement is set to 100 and there is no max iteration number, which means that GA will stop when it cannot make any improvement after 100 iterations.

Because function "geneticalgorithm" can only minimize the target function, to find the most significant inputs, the target function is adjusted the sum of the opposite number of the degree between each input pair.

Moreover, GA is applied to choose 5 different inputs, so in the DNA, each digit should be different. To solve this problem, when there are duplicated digits in the DNA, the output of the target function is set to a large number (2000), so those DNAs will be filtered during evolution.

2.3 Data Inspection

The size of this dataset is 576*27, which means the dataset contains 576 data points and each data point has 27 attributes. The first attribute is the participant number P1, P2 … P24, as mentioned in part 1, there are 24 participants. The last attribute is the labels of music type, 1 for classical, 2 for instrumental and 3 for pop.

All the other attributes are features extracted from the frontal lobe (represented by F7) electroencephalograms, which are the Mean, Maximum, Minimum, Standard Deviation, Interquartile Range, Variance, Sum, Skewness, Kurtosis, Means of the first differences, Means of the second differences, Root Mean Square, Sum of Absolute Values, Simple Square Integral, Variance of Absolute Values, Means of the absolute values of the first and second differences, Log Detector, Average Amplitude Change, Difference Absolute Standard Deviation Value, Detrended Fluctuation Analysis, Fuzzy Entropy, Shannon's Entropy, Permutation Entropy, Hjorth Parameters and Hurst Exponent of electroencephalograms data. Thus, the first and last attribute are both nominal data represented by integers, while the other attributes are floats.

	subject no	mean	max	min	std	iqr	var	sum	skw
count	576.000000	576.000000	576.000000	576.000000	576.000000	576.000000	576.000000	576.000000	576.000000
mean	12.500000	1.819086	6.354124	0.243775	1.900527	1.902208	12.188467	24.834008	0.857294
std	6.928203	2.570039	9.783731	0.763617	2.931105	2.934445	33.388805	37.050789	0.772770
min	1.000000	0.177153	0.331130	0.000036	0.058098	0.022917	0.003375	1.311274	-1.368202
25%	6.750000	0.489684	1.227658	0.005698	0.363696	0.514053	0.132275	6.050474	0.320250
50%	12.500000	0.814839	2.014539	0.015316	0.580049	0.807612	0.336458	9.727701	0.773900
75%	18.250000	1.632414	6.132998	0.230895	1.885264	1.514906	3.554224	23.708503	1.421059
max	24.000000	21.310083	49.653301	14.591949	16.243698	22.097017	263.857719	230.562023	2.390834

Fig. 1. A part of summary statistics of the data, generated by function "describe()"

As shown in Fig. 1, by comparing the maximum value of mean and var, the range of different inputs varies a lot, which indicates that some pre-processes, such as normalization might be applied to make it easier for model training.

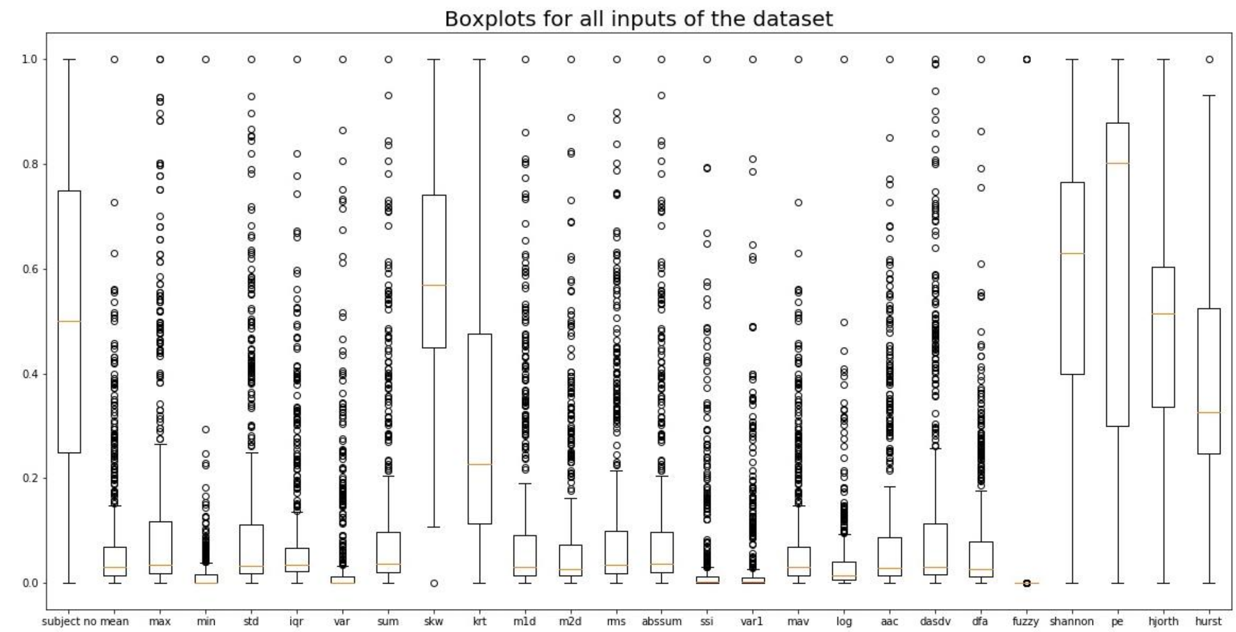

Fig. 2. Boxplots for showing the distributions of each input (first 26 attributes) of this dataset after basic pre-processing steps (normalizing numeric variables), exported as a jpg file.

As shown in Fig. 2, most of the inputs have many outliers, which indicates that it is not suitable to squash data to range 0 – 1. Moreover, for the model I and C, when applying Functional measures on inputs themselves, normalized inputs with many outliers are more likely to have different vectors. In this case, angles between inputs will all be similar, and it will be hard to determine the significant inputs. Thus, to apply Functional measures, specifically tailored pre-processing steps to the dataset should be applied to get a model with better performances.

2.4 Data Preparation

First, for the target attribute "label", its value is replaced from "1, 2, 3" to "0, 1, 2" to enable using the Cross-Entropy Loss function in the network for classification. Each music genre has the same number of data points, which is 192.

Then, all feature attributes (all attributes expect "label") are renamed ("F7" in the feature name is removed) and standardized by subtracting the mean value and divided by the standard deviation of that attribute. As shown in Fig.3, this pre-process ensures that around half of the data value will be positive and around half of the data value will be negative, which can directly apply functional measures now (for the model I and C).

Moreover, the attribute "subject no." (participant number) is nominal data, but in this paper, no further pre-processes are applied to modify this input. The reason is that to deal with nominal data, a common way is to convert it into several columns using one-hot representation. In that case, there will be more input fields and those new input vectors will be different from each other absolutely, which is not good for applying Functional measures.

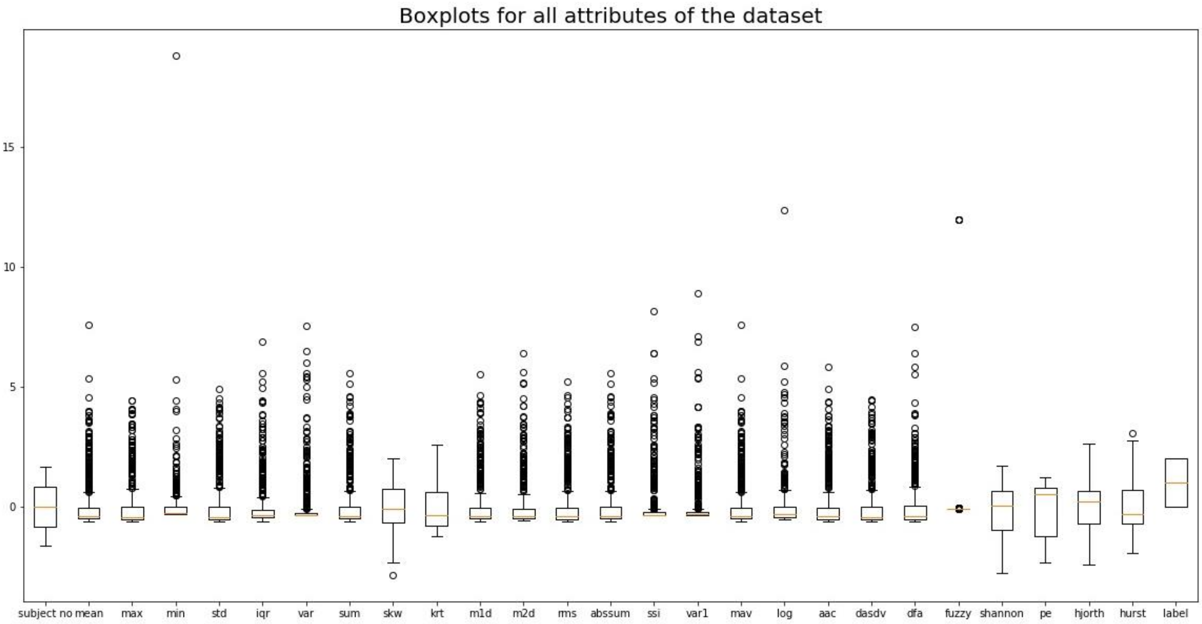

Fig. 3. Boxplots for showing the distributions of each attribute of this dataset after specifically tailored pre-processing, the order of attributes is the same as the original dataset, exported as a jpg file.

As shown in Fig. 3, after specifically tailored pre-processes, all attributes are prepared and ready for building networks.

3 Results and Discussion

As mentioned in part 1 (introduction), the initial dataset is spitted into the training, validation, and testing set in the ratio of 8:1:1 using two train_test_split functions. Neural Networks will be trained on the training set only. By scanning through Networks in the first 1000 epochs, the Network with the highest validation accuracy is chosen and the Network is judged based on the corresponding average testing accuracy, macro precision, macro recall and F1 score. All these steps ensure that the evaluation is fair, valid and all of the available data has been used for training and evaluation.

To compare the created models, the program will run 10 times and the result of each run will be recorded to generate an average result. Here is a table showing the average metric of each model mentioned in this paper compared with the results provided by Rahman et. al. [1].

Table 1. Average metrics (%, rounded to 2 decimals) of models for 10 runs compared with the results provided by Rahman et. al. [1]. For the Network Type, (least) means that it is a model after pruning the 5 least significant inputs, while (most) means that it is a model after pruning the 5 most significant inputs.

Network Type	Validation Accuracy	Testing Accuracy	Testing Precision	Testing Recall	Testing F1 score
Initial Network	47.07	38.41	36.84	37.99	37.38
Model I (least)	46.90	39.66	38.81	39.58	39.18
Model C (least)	46.21	38.28	37.40	38.13	37.75
Model W (least)	48.62	40.21	39.47	40.05	39.75
Model U (least)	49.82	37.36	36.82	37.04	36.93
GA (least)	46.90	37.83	36.33	37.57	36.91
Model I (most)	48.28	40.79	40.79	40.10	40.44
Model C (most)	**41.03**	**33.38**	**30.93**	**33.80**	**32.21**
Model W (most)	49.48	38.17	37.43	37.94	37.67
Model U (most)	**44.69**	**35.12**	**34.08**	**35.04**	**34.51**
GA (most)	**40.69**	**34.03**	**33.17**	**34.33**	**33.70**
The result from Rahman et. al.	97.50		(Precision, recall and F1 score are only shown in the plot and that paper does not provide a specific number)		

By comparing the metrics of each model, for models removed the 5 least significant inputs, the model performance does not drop, and each metric of testing model stays at 37% to 39%, which indicates that Functional measures [5] is a suitable technique for removing redundant information, and it conforms to this paper's stated goals.

When checking the least significant inputs found by each model, it shows that input 'sum' and 'abssum' are usually selected (excepted for model C and GA). The reason is that 'abssum' is the absolute value of 'sum' before pre-processing, which is the same value for this dataset so they are easier to be considered as redundant inputs. However, for model C and GA, they are choosing the inputs based on the average degrees with other inputs, although 'sum' and 'abssum' are the same, they may be completely different from the others, so they can be significant in some situations as well.

While for models removed the 5 most significant inputs, model C and GA shows a clear bad effect on the model performance (5%, from 38% to 33%), while model U has a less bad effect on the model performance (3%, from 38% to 35%). The reason might be that model C and GA directly select inputs based on the dataset while model U is generated based on the weight matrix, the randomness of neuron networks may decrease the performance of model U. The other models do not have a bad effect on model performance, which indicates that model I and C have some limitations. The result shows that, to find the most significant inputs, models based on the average degree (model C, U and GA) have better performances than models based on signal pair (model I and W).

By analyzing the most significant inputs found by model C, U and GA, 'subject no' and 'hjorth' are usually selected, which indicates that these two inputs are important. Feature 'subject no' is the participant number and different 'subject no' means different participants; while 'hjorth' is the Hjorth Parameters, which is normalized slope descriptors (NSDs) in EEG. The reason might be that the brain activities of the same participant should have similar behaviours and Hjorth Parameters can show the activity and complexity of EEG data when the 'subject no' is fixed, these two inputs together will be useful when classifying the music types.

4 Conclusion and Future Work

In this paper, Neural Networks are built to classify the music types using minimal data extracted from EEG, while the Functional measures technique and the Genetic Algorithm have been applied to remove irrelevant inputs. The accuracy of Rahman et. al.'s model is 97.5%; by using methods in this paper, networks can have a testing accuracy of 38% based on only 6% of Rahman et. al.'s dataset, which will be helpful for the development of virtual reality techniques.

The goal of this paper is to identify music types with less data and two methods are introduced for pruning duplicated inputs effectively. To remove redundant information, the Functional measures technique is a powerful method. After pruning 5 least significant inputs among 26 inputs using 4 different models, the testing accuracy, precision, recall and F1 score will not change and stay at around 37% to 39%. However, Functional measure still has some limitations. When the feature number (n) is large and need to select more features (k), it will be time-consuming, and the time complexity is O(n ^k). To solve this problem, the Genetic Algorithm is applied to create the fifth model, which can find the possible optimal solution by simulating natural selection and it can work efficiently when the feature number is large.

To find the most important inputs, models based on the average degree (model C, U and GA) have better performances than models based on signal pair (model I and W). Thus, for future work, instead of dropping a fixed number of redundant inputs, researchers can use model C, U and GA to find the most suitable number of extracting how many most important inputs. Moreover, for model W and U, they will sort inputs based on the existing weight matrix, so if the reliability of the current weight matrix is low and cannot provide models with a high validation/testing score, the effect of Functional measures might be small as well. More efforts need to be taken to minimize the negative effect of models providing unsuitable weight matrix when using the Functional measures technique. Furthermore, those models should be run on a larger dataset and compare the time spent with each other to check to performance of the Genetic Algorithm.

References

1. Rahman, J. R., Gedeon, T., Caldwell, S., Jones, R.: Brain Melody Informatics: Analysing Effects of Music on Brainwave Patterns. In 2020 International Joint Conference on Neural Networks (IJCNN). 1--8 (2020)
2. McCraty, R., Barrios-Choplin, B., Atkinson, M., Tomasino, D.: The effects of different types of music on mood, tension, and mental clarity. Institute of HeartMath, Boulder Creek, Calif., USA (1998)
3. Das, P., Gupta, S., Neogi, B.: Measurement of effect of music on human brain and consequent impact on attentiveness and concentration during reading. Procedia computer science. 172, 1033--1038 (2020)
4. Rumelhart, D. E., Hinton, G. E., Williams, R. J.: Learning internal representations by error propagation. Parallel distributed processing: explorations in the microstructure of cognition. 1, 318--362 (1986)

5. Gedeon, T. D.: DATA MINING OF INPUTS: ANALYSING MAGNITUDE AND FUNCTIONAL MEASURES. International Journal of Neural Systems. 8(2), 209--218 (1997)
6. Gedeon, T. D., Harris, D.: Network Reduction Techniques. Proceedings International Conference on Neural Networks Methodologies and Applications, AMSE. 1, 119--126 (1991)
7. Gedeon, T. D.: Indicators of Hidden Neuron Functionality: Static versus Dynamic Assessment. Australasian Journal of Intelligent Information Systems, invited paper. 1--10 (1996)