# AUTOMATIC EVALUATION OF PSYCHOLOGICAL QUIZ TESTS

## PINTEA, F[lorentina] A[nica]; LACRAMA, L[aurentiu] D[an] & GUI, V[asile]

*Abstract: This paper is concerned with the automatic evaluation of psychological quiz tests using pattern recognition techniques. The subjects' answers are coded and used to build a vector $V_x$. Then $V_x$ is fed to a MLP (MultiLayer Perceptron) neural network which performs the classification in one of the predefined classes of evaluated subjects. A special attention is given to the way the answers are mapped in binary space and to the methods used in order to implement cross-correlations between questions and check points.*

*Key words: pattern recognition, neural networks, automatic evaluation*

## 1. INTRODUCTION

Automatic processing of psychological quiz tests is a field of research where pattern recognition methods can be employed outside of theirs typical purpose - image processing. An adequate numeric encoding of the answers and a high-quality classifier (neural net or statistic), can provide a useful tool able to give a quick evaluation of the subjects' answers. The main problem of this strategy resides in its efficiency. It is not straightforward to numerically encode the responses and it is not simple to build and train a good classifier, thus spending so much work is not efficient if it is used for ever changing set of questions and answers.

Thus the method is useful only in the case of repetitive quizzes with great number of subjects. It is the case of standard psychological tests were large quantity of individual responses must be processed and interpret.

The computer aided evaluation of tests is not a new idea, but previous solutions were focused on the school examination tests. These quizzes contain a set of queries with usually three to five alternative responses - some correct, some incorrect ones. Therefore the program implements a "hit or miss" strategy and determines if the student did select the good answers. Finally the evaluation software gives the student's final score (Karnyanszky & Lacrama, 2008).
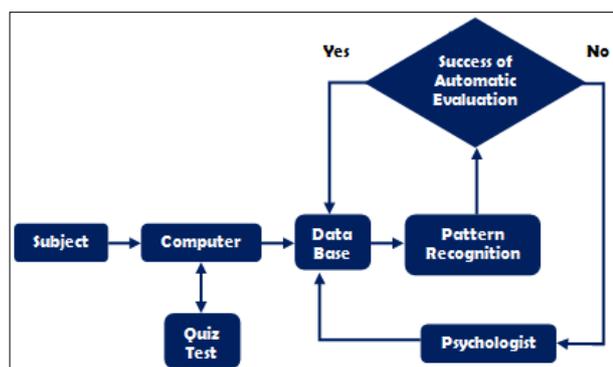


Fig.1. The procedure for the automatic evaluation of psychological quiz tests

The psychological interviews do not mach this blueprint. The subject's responses represent his opinions or his feelings over a certain matter and consequently cannot be classified as correct or incorrect. The estimation of subject's psychological abilities or troubles must be done using a complex procedure as the one described in Fig.1.

As shown, the computer added evaluation have to include a step were the unsuccessful classifications (difficult or atypical cases) must be reevaluated by a psychologist. The authors consider that, while the psychological quizzes are a complex and sensible matter, it is much better to accept a bigger rejection rate in order to avoid errors (Vlasceanu et al., 2009).

It is also very important that, the automatic evaluation system must adapt to some important features of this quizzes:

a. The psychological tests have usually many questions because they have to give both an overall image of the subject's personality and on his reaction on some particular matters. Each such separate topic corresponds to separate quiz's sections and must first be evaluated individually. Finally all particular decisions are assembled in a general conclusion.

b. Many people do not respond entirely sincere to all the queries. Some others do not pay enough attention to the quiz and give quick responses without understanding the questions. Consequently, cross-correlations between queries or check points must be inserted in order to detect some of the dishonest or superficial answers.

c. The neutral response has to be among the alternative responses of each question. It must be taken into account even if it sometimes corresponds to situations described above.

d. The individual outcome of this kind of questions can be put in a standard format conclusion like: "yes, no, not sure or don't want to answer to …".

## 2. THE ENCODING OF THE ANSWERS

Taking into account the above aspects the authors thought that it is helpful to use the classification techniques employed in the Pattern recognition field in order to propose a method and to build a computer expert system able to evaluate each subject's answers and determine his problems if any.

The first step in building such a method is to put up a set of rules for encoding into digital descriptors each of the following features: the weight of each query, the value of each answer in the quiz, the cross-correlations between different questions, the check points, the classes of evaluation and the supplemental parameters (e.g. speed of the response, facial expression, hand movement, subject's temperature etc.).The proposed solution for the value of the answers is given in Tab.1.

| No. | Answer | Descriptor | Code |
|-----|--------|------------|------|
| 1. | Yes | Y | 11 |
| 3. | Neutral/Not sure | 0 | 10 |
| 4. | Don't want to answer | R | 01 |
| 2. | No | N | 00 |

Tab.1. Answers encoding

For example, an item as "Are you happy with your job? " usually have four alternative answers: yes, neutral/not sure, don't want to answer and dissatisfied. It was considered that the rejection answer 'R' is closer to negative attitude 'N' than the neutral one 'O'. Thus the responses where encoded as shown in the Tab.1.

The quiz can have queries with different weight $w_p$ as established by the psychologist when the test was created. This are implemented as constant transfer weights from first to second layer.

The cross-correlations are implemented as logic operations between answers. If, for example, questions $Q_k$ and $Q_m$ are searching the same aspect, but the text logic is reversed in order to discover non-sincere or superficial answers, the computer will execute the operation described in equation (1):

$$if \left( R(Q_K) == -R(Q_K) \right) \, validate \, R(Q_K) \qquad (1)$$

where $R(Q_k)$ and $R(Q_m)$ are the response to the questions $Q_k$ and $Q_m$ respectively and $(-R(Q_m))$ performs the compensation of the reversed logic for the second question.

Similarly, the check points do the same logic operations between two or more different questions, but their result is included in $V_{Xi}$ as a separate item independent from questions answer. This fact gives them a bigger importance in the evaluation. Such verifications can imply more than two questions and are clever traps able to detect non-valid answers.

In the specific case of psychological tests, the authors' experience shows that is better to perform a two step evaluation:
- Stage I: separate evaluation of each test section researching an individual topic because habitually one subject has different opinions on different aspects,
- Stage II: detection of the subject's global situation.

The best way to do this is to construct the learning and the testing sets using already evaluated answers from the previous individuals. If the two sets are carefully selected the convergence will be quick and the outcome will meet the requirements (Hastie & Tibishirani, 1996).

The descriptor vector $V_{Xj}$ encoding the subject's $X_j$ test responses has the form given in equation (2).

$$V_{Xj} = \begin{bmatrix} R(Q_1) \\ R(Q_2) \\ R(Q_3) \\ ... \\ 0 \quad if \quad unvalidated \\ R(Q_k) \quad if \quad validated \\ ... \\ ChP_x \\ ... \\ R(Q_p) \\ ... \\ R(Q_n) \end{bmatrix} \qquad (2)$$

where $1 \leq k,x,p \leq N$, $R(Q_p)$ are the answers to ordinary questions, $R(Q_k)$ are the answers to questions which have cross-correlations for validation and $ChP_x$ is a check point.

Additional parameters $a_p$ (e.q. speed of the response, facial expression, hand movement, subject's temperature etc.) can be added to some or all the questions as multiplicative factors to the corresponding $w_p$ if the test creator do need them and if the computer where the subject type his responses has the devices able to measure them. (Slaney & Covell, 2001)

The initial idea of the authors was to use a standard statistic classifier, but it proved to be difficult to establish each class of responses ideal model. Thus we tried to employ a tree of MLP nets as shown in the Fig.2. For each first step MLP has three-layer architecture and an additional entrance layer with transfer constant weights equal with $w_p a_p$. This first layer is not affected by the changes performed during the training procedure.
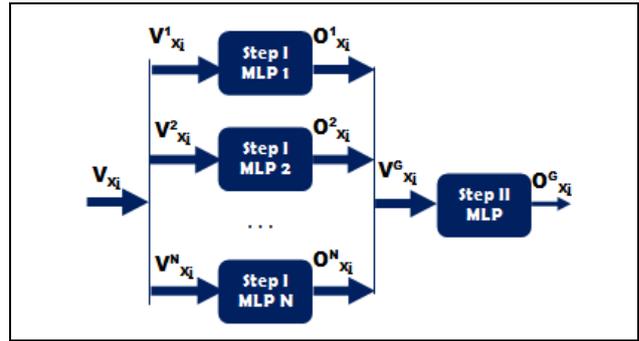


Fig.2. The tree of MLP networks

The N Stage I MLPs are fed with one of the quiz section encoded answers $V^P_{Xj}$ and give the $X_j$'s assessment $O^P_{Xj}$ ($1<P<N$) for the its corresponding problem. For the whole Stage I MLP's results $O^P_{Xj}$ are used to construct the vector $V^G_{Xj}$ feeding the Stage II MLP. This one finally gives the global decision $O^G_{Xj}$. (Duda et al., 2001)

As stated before, the outcomes are also stored in the database in order to use them in the subsequent statistical procedures.

## 3. CONCLUDING REMARKS

The proposed system is still in the testing phase, but the preliminary experimental results over a set of test's answers show that the classifications are accurate enough to be used as a faster and suitable tool to reduce the redundancy in the psychologists' work.

The experiments show that if the neural network training was done for each quiz section separately, the results can arrive up to 96% accuracy. Still we encounter problems with the overall decision evaluation because it seems that the link between the partial evaluations and the global opinion is highly variable between individuals.

## 4. ACKNOWLEDGEMENTS

## 5. REFERENCES

Duda, R.O.; Hart, P.E. & Stork, D.G. (2001), *Pattern Classification,* 2nd edition, New York: John Wiley & Suns

Hastie, T.R. & Tibishirani, R. (1996). Discriminant Adaptive Nearest Neighbor Classification, *IEEE,* Trans. on PAMI, Vol. 18(6), pp.126-133

Karnyanszky, T.M. & Lacrama, D. L. (2008). Teacher's Evaluation – a Component of Quality Assessment System, *Anale. Seria Informatica,* volume 6, pp. 36-41

Schalkoff, R.J. (2009*). Intelligent Systems: Principles, Paradigms and Pragmatics,* Sudbury, MA

Slaney M. & Covell M. (2001). FaceSync: A linear operator for measuring synchronization of video facial images and audio tracks, in: *Neural Information Processing Systems*, MIT Press

Vlasceanu, L.; Hâncean, M.G.; Voicu, B. & Tufis, C. (2009). Statistical distribution, interpretation and options regarding higher education quality status, Quality Barometer, *Internal Report*, Babes Bolyai University, Cluj-Napoca