Acta Universitatis Matthiae Belii ser. Mathematics, 16 (2010), 5–16.
Received: 23 June 2009, Accepted: 2 February 2010.

INTERPRETATION OF THE MMPI-2 TEST BASED ON FUZZY SET TECHNIQUES

IVETA BEBČÁKOVÁ, JANA TALAŠOVÁ, AND PAVEL ŠKOBRTAL

ABSTRACT. MMPI-2 (Minnesota Multiphasic Personality Inventory) is a psychological test for detecting pathological features of personality. After answering all the items of the test each patient is assigned a codetype describing his/her mental health. The diagnosed profile of a patient is verified by the comparison of his/her data to the prototypic profile of the given codetype. This paper introduces a mathematical model for codetype determination and codetype verification. The model has two parts. The first solves the problem of codetype determination by using a fuzzy expert system to formally express the linguistic description of the original method. In the second part, each prototypic profile is described by an n-tuple of fuzzy numbers. This allows us to effectively find the degree of agreement between the profiles and data obtained from the patient. The proposed mathematical model is realized in the MATLAB Fuzzy Logic toolbox.

1. INTRODUCTION

MMPI-2 (Minnesota Multiphasic Personality Inventory) is one of the most frequently used tests for characterization of personality features and psychic disorders. The first version of the test, MMPI, was developed by psychologist S. R. Hathaway and psychiatrist J. C. McKinley [6] of the Minnesota University. Their goal was to develop an instrument to describe patient's personality more effectively than what was allowed by the psychiatric interview with the patient, [1]. At the same time it was desirable to replace a great number of tests, focusing on single features, by a single test capable of full characterization. The fruit of their labor was an extensive testing method with applications far beyond the clinical practice. Today, a revised version of the test, MMPI-2 [3], is used. MMPI-2 is an important screening method for detecting pathological personality features, which is used in clinical practice, as well as in entrance interviews for universities, military, police, or leading positions [7].

2000 *Mathematics Subject Classification.* 03E72, 91E45, 93A30.
Key words and phrases. MMPI-2, Psychometry, Fuzzy expert systems, Base of rules, Linguistic variable.

Use of the MMPI-2 is very demanding. The examiner needs to possess knowledge of theory and use of psychological tests; he/she should have a Master degree in personal psychology and psychopathology, [7]. Furthermore, correct interpretation of the test requires experience with the MMPI-2 and a special training. For this reason, a software with transparent results providing solid basis for the clinic deliberation would be an enormous asset.

1.1. **Quantitative interpretation of MMPI-2.** An important part of the testing process is quantitative interpretation, [5]. Answers to questionnaire questions are used to saturate a large number of scales. Their rough point values are then transformed into linear T-scores. Based on values of these, a codetype of the patient is determined.

The basis for the MMPI-2 interpretation is a determination of codetype, if possible. Each codetype is defined by T-scores of ten clinical scales (1-Hypochondriasis, 2-Depression, 3-Hysteria, 4-Psychopathic deviate, 5-Masculinity-Feminity, 6-Paranoia, 7-Psychasthenia, 8-Schizophrenia, 9-Hypomania, 0-Social introversion). Value of each T-score comes from the interval $[0, 120]$. Values higher than 65 are considered significantly elevated. According to number and type of increased clinical scales we define 55 different codetypes. Codetypes with one significantly elevated clinical scale are designated "Spike" (ten possible types), while two significantly elevated scales represent a "Two Point" (45 possible types). For a codetype to be well defined, there has to be at least five point difference between the T-scores of the highest scales and remaining T-scores. If this is not satisfied, there is a possibility of triad, for example, and it is not possible to use codetypes.

After finding the codetype, the agreement between patient's data and the respective prototypic profile is checked. In this testing, T-scores of all scales need to be considered. Each of 55 prototypic profiles is defined by specific values of all scales. To have a perfect match between the patient and a given prototypic profile, T-scores of patient's scales must not differ from T-scores of the profile by more than ten points.

For finding the T-scores and determining the codetype, the MMPI-2 software was developed [7]. This software finds the codetype only from the two highest T-scores and rest of the data is not involved in the process. This leads to loss of information and it is wasteful of the full MMPI-2 potential. Furthermore, the software does not strictly adhere to the five-point-difference condition and therefore may return an erroneous result.

In this paper, we present a mathematical model, which can help to find several codetypes best fitting the patient. The codetypes are determined in two steps. In the first step, the model searches for codetypes using the MMPI methodology with slightly modified conditions. In the second step, the additional suitable codetypes are found by comparing the patient's data to the prototypic profiles. Similar approach has already been employed in [2], where the second part of the model employed a base of rules, which caused several problems. For example, prototypic profiles were not detected with unit overlap and the method was was not universal and tended to prefer certain profiles. In this paper we introduce a different treatment of the second part of the model, which addresses the aforementioned issues.

The Czech version of the MMPI-2 does not work with all of the scales. It uses and saves values of only 79 of them. The mathematical model will consider this simplified version of the MMPI-2.

2. Preliminaries

The codetype determination requiring full satisfaction of all 79 conditions of a prototypic profile is problematic. Classification based on such a crisp mathematical model may not work, because only rarely a patient satisfies fully a prototypic profile. It will be shown that in a situation like this, as well as in many areas of social sciences and psychology, it is effective to use the so called fuzzy approach.

Fuzzy set theory [4, 8] gives us a tool to model the vagueness phenomenon. It allows us to describe mathematically linguistic values and linguistically defined rules. This is the reason why in this special case, where we look for a mathematical model of a linguistically described methodology, description by fuzzy sets is very helpful.

Let U be a nonempty set. A fuzzy set A on U is defined by the mapping $A : U \to [0, 1]$. For each $x \in U$ the value $A(x)$ is called a membership degree of the element x in the fuzzy set A, $A(\cdot)$ is a membership function of the fuzzy set A.

A height of a fuzzy set A on U is a real number $\mathrm{hgt}(A) = \sup_{x\in U}\{A(x)\}$. An intersection of fuzzy sets A, B on U is a fuzzy set $A \bigcap B$ on U with a membership function $(A \bigcap B)(x) = \min\{A(x), B(x)\}$ for any $x \in U$.

A fuzzy number A is a fuzzy set on $\mathbb{R}$ which fulfills the following conditions: the kernel of the fuzzy set A, $\mathrm{Ker}A = \{x \in \mathbb{R}|A(x) = 1\}$, is a non-empty set, the α-cuts of the fuzzy set A, $A_\alpha = \{x \in \mathbb{R}|A(x) \geq \alpha\}$, are closed intervals for all $\alpha \in (0, 1]$, the support of A, $\mathrm{Supp}A = \{x \in \mathbb{R}|A(x) > 0\}$, is bounded.

The family of all fuzzy numbers on $\mathbb{R}$ is denoted by $F_N(\mathbb{R})$. If $\mathrm{Supp}A \subseteq [a, b]$ then A is referred to as a fuzzy number on the interval $[a, b]$. The family of all fuzzy numbers on the interval $[a, b]$ is denoted by $F_N([a, b])$. A linear fuzzy number on the interval $[a, b]$ that is determined by four points $(x_1, 0), (x_2, 1), (x_3, 1), (x_4, 0)$, $a \leq x_1 \leq x_2 \leq x_3 \leq x_4 \leq b$, is a fuzzy number A with the membership function depending on parameters x_1, x_2, x_3, x_4, as follows

$$\forall x \in [a,b] : A(x, x_1, x_2, x_3, x_4) = \begin{cases} 0, & \text{for } x < x_1; \\ \frac{x-x_1}{x_2-x_1}, & \text{for } x_1 \leq x < x_2; \\ 1, & \text{for } x_2 \leq x \leq x_3; \\ \frac{x_4-x}{x_4-x_3}, & \text{for } x_3 < x \leq x_4; \\ 0, & \text{for } x_4 < x. \end{cases}$$

This linear fuzzy number A will be denoted by $A \sim (x_1, x_2, x_3, x_4)$.

A linguistic variable is a quintuple $(X, T(X), U, G, M)$, where X is a name of the variable, $T(X)$ is a set of its linguistic values (linguistic terms), U is an universe, which the mathematical meanings of the linguistic terms are modelled on, G is a syntactical rule for generating the linguistic terms, and M is a semantic rule, which to every linguistic term $\mathcal{A}$ assigns its meaning $M(\mathcal{A})$ as a fuzzy set on U. If the set of linguistic terms is given explicitly, then the linguistic variable is denoted by $(X, T(X), U, M)$.

Let $(X_j, T(X_j), U_j, M_j)$, $j = 1, 2, \ldots, m$, and $(Y, T(Y), V, N)$ be linguistic variables. Let $\mathcal{A}_{ij} \in T(X_j)$ and $M(\mathcal{A}_{ij}) \in F_N(U_j)$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. Let $\mathcal{B}_i \in T(Y)$ and $M(\mathcal{B}_i) \in F_N(V)$, $i = 1, 2, \ldots, n$. Then the following scheme F

If X_1 is $\mathcal{A}_{11}$ and ... and X_m is $\mathcal{A}_{1m}$, then Y is $\mathcal{B}_1$
If X_1 is $\mathcal{A}_{21}$ and ... and X_m is $\mathcal{A}_{2m}$, then Y is $\mathcal{B}_2$
....
If X_1 is $\mathcal{A}_{n1}$ and ... and X_m is $\mathcal{A}_{nm}$, then Y is $\mathcal{B}_n$

is called a linguistically defined function (base of rules).

The process of calculating linguistic values of an output variable for the given linguistic values of input variables by means of such a rule base is called an approximate reasoning. There are several methods of approximate reasoning. The most popular and the most widely used one is the Mamdani algorithm.

Let F be the base of rules defined above and let us assume the observed values to be

$$X_1 \text{ is } \mathcal{A}'_1 \text{ and } X_2 \text{ is } \mathcal{A}'_2 \text{ and } \ldots \text{ and } X_m \text{ is } \mathcal{A}'_m,$$

then by entering the observed values into the base of rules F, according to the Mamdani algorithm, we obtain the output value $Y = \mathcal{B}'$, where the $\mathcal{B}'$ is the linguistic approximation [4] of a fuzzy set B^M. The membership function of the fuzzy set B^M is defined for all $y \in V$ as follows $B^M(y) = \max\{B_1^M(y), \ldots, B_n^M(y)\}$, where $B_i^M(y) = \min\{h_i, B_i(y)\}$, $h_i = \min\{\mathrm{hgt}(A_{i1} \bigcap A'_1), \ldots, \mathrm{hgt}(A_{im} \bigcap A'_m)\}$, for $i = 1, \ldots, n$.

3. Designed mathematical model

The quantitative interpretation of the MMPI-2 is performed in two steps. First, based on values of clinical scales, a patient's codetype is determined. This is followed by the verification, where the relevant prototypic profile is compared with the patient's data.

The proposed mathematical model respects this structure of MMPI-2. In the first step, the model finds the three clinical scales with the highest T-scores, and with help of the linguistically described function decides on a codetype. In the second step, the model works with values of all 79 scales and calculates the overlap between the linear T-scores of the patient and the prototypic profile of the codetype found in the previous step. Simultaneously the model searches for other prototypic profiles, which agree well with patient's data.

3.1. **Codetype determination.** Two conditions are important for correct determination of the codetype. First, T-scores of significantly elevated scales must be higher than 65. Second, values of the highest scales must be at least five points higher than values of all remaining scales. In practice, it is often difficult to strictly fulfill this conditions. It has shown to be more effective to use the fuzzy approach and define these conditions linguistically. Furthermore, use of the fuzzy set theory was instrumental in finding more variants of the codetype, which can be presented to the evaluator.

Prior to further processing, the scales need to be ordered from the highest T-score to the lowest. Based on the above mentioned requirements, we then define linguistic variables as:

(1) $\langle$The First Scale Elevation,
$\{$Insignificant, Significant$\}, [0, 120], M_1\rangle$,
(2) $\langle$The Second Scale Elevation,
$\{$Insignificant, Significant$\}, [0, 120], M_1\rangle$,
(3) $\langle$The Third Scale Elevation,
$\{$Insignificant, Significant$\}, [0, 120], M_1\rangle$,
(4) $\langle$The Difference between the First Two Scales,
$\{$Small, Big Enough$\}, [0, 120], M_2\rangle$,
(5) $\langle$The Difference between the 2^{nd} and the 3^{rd} Scale,
$\{$Small, Big Enough$\}, [0, 120], M_2\rangle$,
(6) $\langle$Codetype Shape,
$\{$Spike, Two Point, Potential Triad, Within-Normal-Limits$\}$,
$\{1, 2, 3, 4\}, N\rangle$,

where
M_1(Insignificant) $= IE \sim (0, 0, 63, 65)$, M_1(Significant) $= SE \sim (63, 65, 120, 120)$, M_2(Small) $= SM \sim (0, 0, 0, 5)$, M_2(Big Enough) $= BE \sim (0, 5, 120, 120)$, N(Spike) $= S \sim (1, 1, 1, 1)$, N(Two Point) $= 2P \sim (2, 2, 2, 2)$, N(Potential Triad)$= PT \sim (3, 3, 3, 3)$, N(Within-Normal-Limits) $= WNL \sim (4, 4, 4, 4)$. Some of defined variables are illustrated in Fig. 1 and 2.

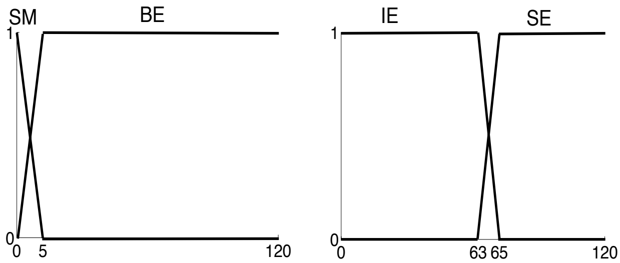

FIGURE 1. Input linguistic variables
Left: *The Difference between the First Two Scales* and its two linguistic values *Small* and *Big Enough* modelled by fuzzy numbers SM and BE.
Right: *The First Scale Elevation* and its two linguistic values *Insignificant* and *Significant* modelled by fuzzy numbers IE and SE.

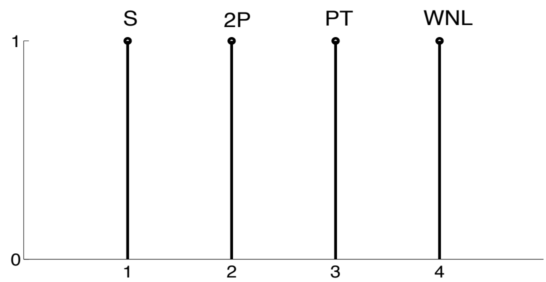

FIGURE 2. Output linguistic variable *Codetype Shape* and its four linguistic values *Spike, Two Point, Potential Triad* and *Within-normal-limits* modelled by fuzzy numbers *S,2P,PT* and *WNL.*

With help of these six linguistic variables and four rules we construct a base of rules F:

rule 1: If The First Scale Elevation is Significant and The Second Scale Elevation is Insignificant and The Difference between the First Two Scales is Big Enough, then the Codetype Shape is a Spike.

rule 2: If The First Scale Elevation is Significant and The Second Scale Elevation is Significant and The Difference between the 2^{nd} and the 3^{rd} Scale is Big Enough, then the Codetype Shape is Two Point.

rule 3: If The First Scale Elevation is Significant and The Second Scale Elevation is Significant and The Third Scale Elevation is Significant and The Difference between the 2^{nd} and the 3^{rd} Scale is Small, then the Codetype Shape is Potential Triad.

rule 4: If The First Scale Elevation is Insignificant, then the Codetype Shape is Within-Normal-Limits.

The base of rules F has five input linguistic variables - the three highest T-scores of clinical scales and the two differences between them - and one output linguistic variable, which determines the shape of the codetype.

Together with the Mamdani approximate reasoning algorithm, the linguistic function F forms an expert system for determination of the codetype shape. With values of clinical scales as an input, the model produces a fuzzy set B^M that helps the evaluator to determine possible codetype shapes. The membership degree of an element of the set $\{1, 2, 3, 4\}$ in fuzzy set B^M, representing a particular codetype shape, equals to the degree of satisfaction of the respective rule. See, for example, Fig. 3. To determine the complete codetype of the patient, we need to combine the information about the codetype shape with knowledge of the initial ordering of clinical scales. For example, if the codetype shape is Spike and the designation of the highest scale is 8-Schizophrenia, then the codetype is Spike 8.

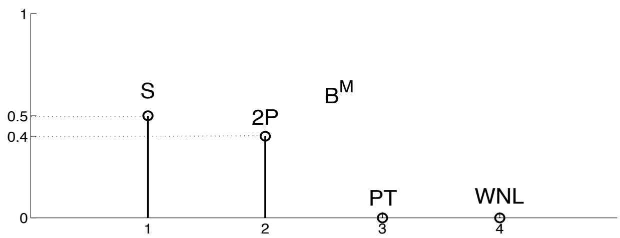

FIGURE 3. The fuzzy set B^M as obtained by entering input values [67 64 62 3 2] into the designed expert system. The degrees of satisfaction express the possibility that the corresponding codetype shape is a Spike (possibility 50%) or a Two Point (possibility 40%).

3.2. **Codetype verification.** Each of all 55 codetypes is described in detail by a so called prototypic profile. Codetype verification is based on the calculation of the degree of agreement between the patient's data and the prototypic profiles corresponding to the codetypes, which were determined in the first part of the model. Besides the verification, the model also searches for other prototypic profiles with a good overlap. Each profile is described by a vector of 79 real numbers representing values of the 79 scales with the T-scores ranging from 0 to 120. For a patient's profile to match a prototypic profile, all the patient's T-scores must be within 10 point distance from the prototypic values.

In the second part of the mathematical model we replaced all crisp numbers t_{ij} describing the prototypic profiles by linear fuzzy numbers $T_{ij} \sim (t_{ij}-20, t_{ij}-10, t_{ij}+10, t_{ij}+20)$, $i = 1, 2, \ldots, 55$, $j = 1, 2, \ldots, 79$. The example is illustrated in the Fig. 4. The kernel of the designed fuzzy number corresponds to the requirements of the methodic, i.e. if the patient's T-score is within 10 point distance from the prescribed value, there is a perfect match and the membership degree is equal to 1. The support of the fuzzy number was set at twice the length of the kernel, i.e. if the distance of the patient's T-score and the prototypic value is bigger than 20 points, then there is no match at all and the membership degree is zero.

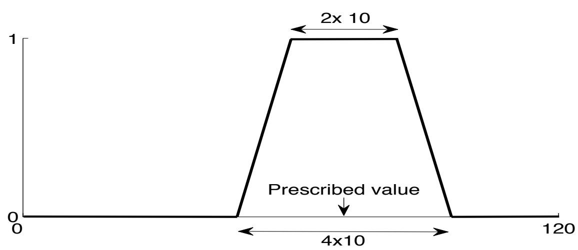

FIGURE 4. Fuzzy number replacing the crisp prototypic value of a scale. The membership degree corresponds to the degree of agreement between the patient's T-score and the prescribed value.

Each i-th, $i = 1, 2, \ldots, 55$, prototypic profile is then described by 79 of these fuzzy numbers. Entering the patient's T-scores $t'_1, t'_2, \ldots, t'_{79}$ into the designed fuzzy numbers, we obtain 79 membership degrees $T_{i1}(t'_1), T_{i2}(t'_2), \ldots, T_{i,79}(t'_{79})$. The degree of agreement between the patient's data and the i-th prototypic profile is calculated as an arithmetic mean of these membership degrees:

$$h_i = \frac{1}{79} \sum_{j=1}^{79} T_{ij}(t'_j), \; i = 1, 2, \ldots, 55. \tag{1}$$

During the development of the model we tried various aggregation operators. However, the common aggregation operators used for modelling the operation of logical conjunction, such as minimum, proved unfeasible, because a patient rarely satisfies the full range of conditions. On the other hand, the arithmetic mean proved itself to be the most convenient in this case. The degree of agreement between the given data and the prototypic profile here represents the average satisfaction of all 79 prescribed conditions. Compared to minimum, for example, the arithmetic mean provides better information about the satisfaction of given conditions. In the future, the aggregation operator can be readjusted to the requirements of the examiner and the arithmetic mean can be replaced by an other aggregation operator, for example weighted mean or OWA, [9].

Applying the aforementioned approach we are able to test all the 55 prototypic profiles. The result can be modelled by a fuzzy set H on the set U, $U = \{1, 2, \ldots, 55\}$, where each integer between 1 and 55 corresponds to one prototypic profile and the membership degrees $H(i)$, $i = 1, 2, \ldots, 55$, represent the overlap of the respective prototypic profiles with the profile of the patient. The example is illustrated in Fig. 5.

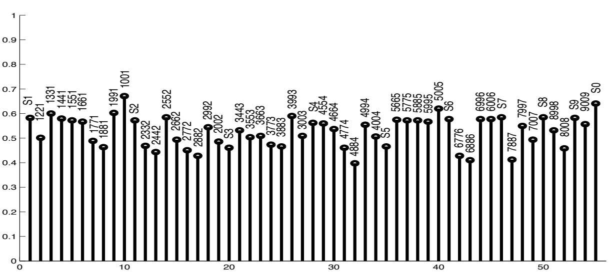

FIGURE 5. The fuzzy set H as obtained by comparing 79 given values with the prescribed prototypic profiles. The degrees of satisfaction represent the overlap between the prototypic profiles and the patient's profile.

4. The implementation of the mathematical model in MATLAB

Both parts of the proposed mathematical model were realized in MATLAB. At first, we have used the Fuzzy Logic Toolbox to create the base of rules and to set the proper approximate reasoning algorithm. Then to each one of the 55 prototypic profiles we have assigned a 79-tuple of fuzzy numbers, as was described in the previous section.

An example of the output can be seen in Fig. 6. The output of the utility is in the form of three figures and linguistic description of the situation. The first figure presents values of clinical scales as obtained from the patient - the patient's profile. The second figure presents possible codetypes, together with their respective degrees of satisfaction. The third figure shows all prototypic profiles and their overlap with the patient's profile. The evaluator can therefore decide, whether the found codetypes are in good agreement with all available patient's data. The linguistic output presents possible codetypes and three prototypic profiles with the best agreement. In addition it comments on a possibility of a triad or scales within normal limits.

In Fig. 6 we demonstrate performance of the implementation. According to clinical scales values, codetype 6-9 was determined. The result is in agreement with the original software. However, during the prototypic profile analysis, the codetype 6-9 didn't show sufficient agreement, as the degree of overlap was only 0.48. The three most faithful profiles were those of codetypes 6-8/8-6, 8-9/9-8, and 7-8/8-7, with 6-8/8-6 showing the best overlap. This suggests that for further deliberation, codetypes 6-8/8-6 should be considered in addition to 6-9.

5. Conclusion

In the paper we have created a mathematical model which can help an examiner with quantitative interpretation of the results given by the MMPI-2 test. For determination of a MMPI codetype we have employed a fuzzy expert system to formally express the linguistically described method of data analysis. To verify the overlap between the prototypic profiles of the found codetypes and the patient's data, all the prototypic profiles were described by 79-tuples of special fuzzy numbers. This allowed us to effectively find the degree of agreement between the respective prototypic profiles and data obtained from the patient. The model contains several free parameters which allow for further fine tuning needed before a practical application

The created fuzzy model was implemented in MATLAB. The created utility, employing the fuzzy approach, can analyze the data while avoiding the shortcomings of the existing software "MMPI-2". In this way the utility can serve as a valuable tool for a human psychiatrist or psychologist in the tuning process.

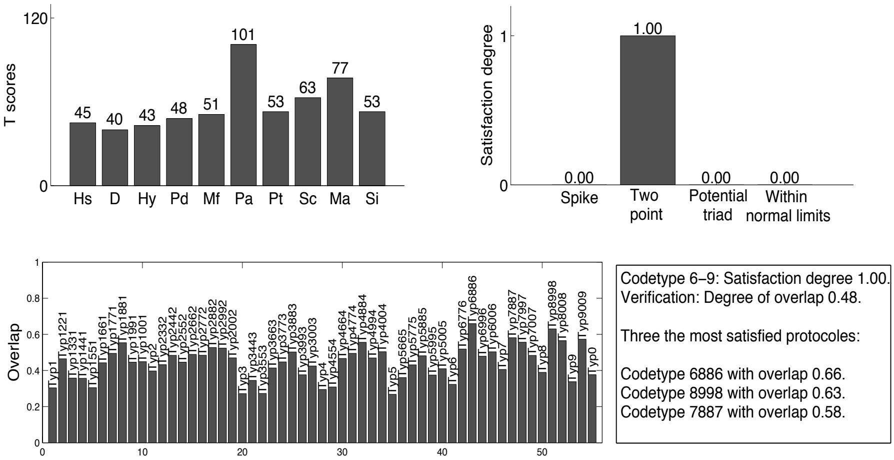

FIGURE 6. Three figures and linguistic description as returned by the MATLAB implementation of the model.

REFERENCES

[1] Archer, R. P., MMPI-A: Assessing adolescent psychopathology, Third edition, Lawrence Erlbaum Associates Publishers, 2005.

[2] Bebčáková, I., Talašová, J., Škobrtal, P., *Fuzzy Approach to Quantitative Interpretation of MMPI-2.* Journal of Applied Mathematics, Vol. 2, Number 1, 2009, 65-75, ISSN 1337-6365

[3] Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., and Kaemmer, B., MMPI-2: Manual for administration and scoring. University of Minnesota Press, Minneapolis, 1989.

[4] Dubois, D., Prade, H.(Eds.), Fundamentals of fuzzy sets. Kluwer Academic Publishers, Boston/London/Dordrecht, 2000.

[5] Greene, R. L., The MMPI-2 An interpretive manual, Second edition, Allyn & Bacon, 1999.

[6] Hathaway, S. R. and Mckinley, J. C., *A multiphasic personality schedule (Minnesota): I. Construction of the schedule*, Journal of Psychology, 10, 249-254, 1940.

[7] Hathaway, S. R., Mckinley, J. C. and Netík, K., Minnesota Multiphasic Personality Inventory - 2, First czech edition, Testcentrum, Praha, 2002.

[8] Talašová, J., Fuzzy metody vícektriteriálního hodnocení a rozhodování, Univerzita Palackého v Olomouci, Olomouc, 2003.

[9] Torra, V. and Narukawa, Y., Modeling decisions: Information fusion and aggregation operators, Springer-Verlag, Berlin Heidelberg, 2007.

(I. Bebčáková) Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University Olomouc, tř. 17. listopadu 12, 771 46 Olomouc, Czech Republic

E-mail address: bebcakova.iveta@post.cz

(J. Talašová) Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University Olomouc, tř. 17. listopadu 12, 771 46 Olomouc, Czech Republic

E-mail address: talasova@inf.upol.cz

(P. Škobrtal) Department of Psychology and Pathopsychology, Faculty of Education, Palacký University Olomouc, Žiškovo nám. 5, 771 40 Olomouc, Czech Republic

E-mail address: pavel.skobrtal@email.cz