Supporting a Medical Diagnostic Process by Selected AI Methods: an Asperger Syndrome Case Study

Tamar Kakiashvili[1], Waldemar W. Koczkodaj[1,5], Duncan Matheson[1], Phyllis Montgomery[1], Kalpdrun Passi[1], Filip Rybakowski[2], Ryszard Tadeusiewicz[3], Marc Woodbury-Smith[4]

[1] Laurentian University, Sudbury, Ontario, Canada
[2] Poznan University of Medical Sciences, Poland
[3] AGH University of Science and Technology, Poland
[4] Autism Research Centre & Section of Developmental Psychiatry, University of Cambridge, UK
[5] the corresponding author: wkoczkodaj@cs.laurentian.ca

Abstract. A medical knowledge-driven diagnostic process can be supported by AI methods as presented here by an Asperger Syndrome case study. Two methods: consistency-driven pairwise comparisons (CDPC) and automatic understanding (AU) are presented in this study. Deficiencies of a data-driven model for the medical diagnostic process and clinical reasoning are also discussed.

Keywords. Medical diagnostic process, consistency-driven pairwise comparisons, automatic understanding, Asperger syndrome.

1 Introduction

Medicine is not an exact science. It is also applicable even in a greater extent, to the medical diagnostic process. For example, it took two years and 126 medical appointments for a woman of letter to have her illness diagnosed as MS (multiple sclerosis). The TV program has stressed the woman's education as well as her high level of organization (how many of us keep track of the number of medical appointments?). As a patient, she was doing her best to help her physicians in getting the proper diagnosis. Unfortunately, most patients in mental disorders are so helpful. A brain, or data processing capabilities of a mentally disabled patient can be compared to a runaway car with no brakes, broken steering mechanism, and all four wheels attached to axis by one lose screw only. No wonder that it is reflected by the four editions of the Diagnostic and Statistical Manual of Mental Disorders (DSM, a Roman literal is often used for the edition) published by the American Psychiatric Association. DSM is a handbook for mental health professionals that lists different categories of mental disorder and the criteria for diagnosing them. It is used worldwide by clinicians and researchers as well as insurance companies, pharmaceutical companies and policy makers. It has attracted controversy and criticism as well as praise. The first edition of 1952 was 134 pages long and listed 182 disorders. The most current IV edition of 1994 is 886 pages long and lists 297 disorders.

One telling example is the declassification of homosexuality as a mental disorder. Homosexuality was listed as a mental disorder in the DSM until 1974, when gay activists demonstrated in front of the American Psychiatric Association Convention. The DSM-IV is a categorical classification system. The categories are prototypes, and a patient with a close approximation to the prototype is said to have that disorder. DSM-IV states that "there is no assumption that each category of mental disorder is a completely discrete entity with absolute boundaries..." but isolated, low-grade and non-criterion (unlisted for a given disorder) symptoms are not given importance.

A medical diagnostic process consists of correlating known *patterns of disease* with the various classes of clinical data elicited from the history, physical examination, and tests which are usually resulted by diagnostic decisions. It has highly iterative nature and is knowledge-driven. Any improvement to medical diagnostic process is of utmost importance. The use of expert systems in medicine is limited. In the medical diagnostic process, they turned to be of less use that it has been anticipated. We can observe a graduate shift from expert systems to knowledge-based and decision support systems. We will attempt to show that the consistency-driven pairwise comparisons (CDPC) and automatic understanding (AU) approaches seems to be particularly useful for improving the medical diagnostic process.

A Monte Carlo statistical study demonstrated that the error of assessing lengths of randomly generated bars has decreased from 15% to 5% when bars were compared in pairs. This is rather remarkable improvement. Considering unreliability of the psychiatric assessments, much can be achieved if a better technology is applied. It needs to be stressed that it is a supplementary method fully respecting DSM-IV classifications and its procedures.

This paper will only focus on the Asperger syndrome. However, it is easy to see that the presented methodology is applicable not only to other mental disorders but other physical illnesses making this approach universal.

2 The Asperger syndrome characteristics from the psychiatry point-of-view

Characteristics of Asperger syndrome are reduced to bare minimum and include:

A. Limited social relationships – social isolation

1. Few/no sustained relationships; relationships that vary from too distant to too intense.
2. Awkward interaction with peers.
3. Unusual egocentricity, with little concern for others or awareness of their viewpoint; little empathy or sensitivity.
4. Lack of awareness of social rules; social blunders.

B. Problems in communication

1. An odd voice, monotonous, perhaps at an unusual volume.
2. Talking ‘at’ (rather than ‘to’) others, with little concern about their response.
3. Superficially good language but too formal/stilted/pedantic; difficulty in catching any meaning other than the literal.
4. Lack of non-verbal communicative behavior: a wooden, impassive appearance with few gestures; a poorly coordinated gaze that may avoid the other’s eyes or look through them.
5. An awkward or odd posture and body language.

C. Absorbing and narrow interests

1. Obsessively pursued interests.
2. Very circumscribed interests that contribute little to a wider life, e.g. collecting facts and figures of little practical or social value.
3. Unusual routines or rituals; change is often upsetting.

(From Berney, 2004 after Gillberg et al, 2001)

The above questionnaire does not resemble the questionnaire published by Gillberg et al in 2001 although Barney claims it in [9] by including “after Gillberg et al, 2001” below the questionnaire. However, it does not much matter from point of view of the presentation of the proposed method. It works not only for the above questionnaire but also for any other questionnaire

Asperger Syndrome or (Asperger's Disorder) is a neurobiological disorder named for a Viennese physician, Hans Asperger. In spite of the publication of his paper in the 1940s, it was not until 1994 that Asperger Syndrome was added to the DSM IV and only in the past few years has AS been recognized by professionals and parents. More recent research review on Asperger’s Syndrome can be found in [10].

3 The pairwise companions preliminaries and Asperger syndrome model

From the mathematical point of view, the pairwise comparisons method creates a matrix (say A) of values (aij) of the i-th candidate (or alternative) compared head-to-head (one-on-one) with the j-th candidate. A scale [1/c,c] is used for i to j comparisons where c>1 is a not-too-large real number (5 to 9 in most practical applications).

It is usually assumed that all the values (aij) on the main diagonal are 1 (the case of i compared with i and that A is reciprocal: (aij) =1/(aij) since i to j is (or at least, is expected to be) the reciprocal of j to i . (As explained below, the reciprocity condition is not automatic in certain scenarios of comparisons.) It is fair to assume that we are powerless, or almost powerless, as far as inconsistency is concerned. All we can do is to locate it and reconsider our own comparisons to reduce the inconsistency in the next round.

Using the scale 1:5 (1 stands for equal of unknown importance and 5 for the highest preference, we have three group and we compare them against each other receiving results in Table 1.

Table 1. Comparisons for the first group level

1	2	4
1/2	**1**	2.5
1/3	1/2	**1**

Clearly, the above matrix is not consistent since a_{13}=3 but it is not equal to $a_{12}*a_{23}$. However, the inconsistency *ii* can be computed from the following formula:
$ii=\max(|1-a_{ij}/(a_{ik}*a_{kj})|,|1-a_{ik}*a_{kj}/a_{ij}|)$ for i=1, j=2, and k=3 (as introduced in [3] and presented in Appendix B).

The value of the inconsistency indicator *ii* is 0.20 from the above formula and it is lower than the assumed threshold 1/3. The computed weights (as normalized geometric means of rows) are presented in Table 2. By changing, for example, 4 to 5, we could receive the new inconsistency index ii=0 but it should only take place if there is new evidence collected to support such a change.

Table 2. Weights for the first level (groups)

Group	weight
A	0.5648
B	0.3042
C	0.1311

As explained in [3], the above values are computed as normalized geometric means of the matrix row. The above method is applied to subgroups receiving overall results for all criteria as presented in Table 3.

Table 3. Weights for individual questions sorted by weights and question id

Item	weight	item	weight
A1	0.2179	C1	0.0524
A2	0.1566	C2	0.0524
A3	0.1107	B4	0.0511
B1	0.0966	B3	0.0511
A4	0.0796	B2	0.0511
B5	0.0542	C3	0.0262

4 About the automatic understanding support

The "automatic understanding of the images" is well established (for example, [4]). It helps with the recognition of such type images as:

- morphology of healthy organs is different for every human being, so we not have any kind of template of "*proper view*" of the analyzed object,
- deformations of the organ shapes and sizes (caused by the illness) can differ in the general form, or even in the number (e.g., kidneys), and in their localization even if diseases are in fact identical.

One of the ways leading to the improvement of analysis of medical images is by using the computers as the advisors for the doctors. Computers cannot replace doctors in analyzing medical images but can be very helpful in their interpretation, especially for general practitioners or psychiatrists who are usually not excessively well trained medical imaging interpretation. Computer vision methods, traditionally used for helping with the medical image interpretation, perform all kind of operations as previously discussed but it is definitely not sufficient in practice.

Surprisingly, the automatic understanding (AU) approach is also applicable, with very little of modifications, to medical interviews. All traditional steps of automatic processing and interpretation of medical images are applicable to it. Psychiatrists are needed to do more than just processing, analyzing and recognizing symptoms. The full interpretation of a complex problem must be supported by the semantic interpretation of the "image" content. In fact, the psychiatrist's activity during the interview interpretation is not exclusively devoted to "measuring" some parameters or doing some classification.

The understanding process is always based on the medical knowledge. This is the main difference between every method of the interview processing, analysis and recognition, which is a data-driven procedure while the task performed by medical doctors is knowledge-driven.

Taking into account all the facts mentioned above, we can build mathematical models and construct practical algorithms for automatic understanding of medical data gathering by interviews, observations, images, or tests. The method presented here is based on the linguistic description of the interviews, which must be prepared for every kind of situation under consideration (e.g., uncontrolled outburst of anger or a total lack of response) on the base of specially designed artificial *interview content describing language*. After designing the structure of artificial language devoted to description the merit index of an interview under consideration, we must define special kind of graph-grammar, describing the rules of the proposed language and prepare automatic procedures for extracting necessary elements (so called graphical primitives and graphical relations) playing role "nouns" and "verbs" of defined grammar.

When we have the proper knowledge-based grammar, we can convert every medical interview to the merit based linguistic description. For it, we can perform automatic parsing. The parsing process transforms this description of the form of the interview to the description of the medical By creating a knowledge base system, we improve our knowledge and interpretation of the medical interviews in time.

5 Conclusions

A major undertaking is required to computerize the mental disorder diagnostic process. This is just the beginning in establishing interest. However, the obtained results are solid and can be used for planning and training purposes. While the pairwise comparisons method has been used for 222 years (Condorcet made a reference to it in 1785, the time of the French Revolution, [5]), to our knowledge, it has not been used to improve the psychiatric disorders. The Asperger syndrome is associated with the learning disability and it is in our interest to help psychiatrists with its diagnosis.

6 Acknowledgment

This work has been supported by the Ministry of Science and Higher Education, Republic of Poland, under project number N516 007 32/0978 decision No. 0978/T02/2007/32 in Poland, the NSERC grant and LURF in Canada.

References

[1] Gillberg C., Gillberg C., Rastam M., et al (2001). *The Asperger Syndrome (and high-functioning autism) Diagnostic Interview (ASDI).* A preliminary study of a new structured clinical interview. Autism, 5, 57–66.
[2] Kakiashvili T., Kielan K., Koczkodaj W.W., Passi K., Tadeusiewicz R., *Supporting the Asperger Syndrome Diagnostic Process by Selected AI Methods*, Proceedings of artificial intelligence studies. Vol. 4, pp21-27, 2007.
[3] Koczkodaj W.W., *A new definition of consistency of pairwise comparisons*, Mathematical and computer modelling, (18)7, pp, 79-84, 1993.
[4] Tadeusiewicz R., Ogiela M.R.: *Automatic Understanding of Medical Images*, New Achievements in Syntactic Analysis of Selected Medical Images, Biocybernetics and Biomedical Engineering, vol. 22, nr 4, 2002, pp. 17-29
[5] Condorcet M., *The Essay on the Application of Analysis to the Probability of Majority Decisions*, Paris: Imprimerie Royale, 1785.
[6] Fechner G., *Elemente der Psychophysik* (1860, 2nd ed., 1889)
[7] Thurstone L.L. (1927). *A law of comparative judgement*. Psychological Review, 34, 278-286.
[8] Saaty T.L. (1977). *A scaling method for priorities in hierarchical structures*. Journal of Mathematical Psychology, 15, 234-281.
[9] Berney T., *Asperger syndrome from childhood into adulthood*, Advan. Psychiatr. Treat., 2004; 10: 341-351.
[10] Woodbury-Smith M.R., Volkmar F.R. (2008). *Asperger Syndrome: A Review*. European Child and Adolescent Psychiatry, 56: 1-11.

Appendix A

Basic Concepts of Pairwise Comparisons

A n by n pairwise comparisons matrix is defined as a square matrix $A=[a_{ij}]$ such that $a_{ij}>0$ for every $i,j=1,\ldots,n$. Each a_{ij} expresses a relative preference of criterion (or stimulus) s_i over criterion s_j for $i,j=1,\ldots,n$ represented by numerical weights (positive real numbers) and w_i and w_j respectively. The quotients $a_{ij}= w_i / w_j$ form a pairwise comparisons matrix:

$$\boldsymbol{A} = \begin{vmatrix} 1 & a_{13} & \ldots & a_{1n} \\ 1/a_{13} & 1 & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ 1/a_{1n} & 1/a_{2n} & \ldots & 1 \end{vmatrix}$$

A pairwise comparisons matrix $\boldsymbol{A}$ is called *reciprocal* if $a_{ij}=1/a_{ji}$ for every $i,j=1,\ldots,n$ (then automatically $a_{ii}=1$ for every $i=1,\ldots,n$ because they represent the relative ratio of a criterion against itself). A pairwise comparisons matrix $\boldsymbol{A}$ is called *consistent* if $a_{ij} \cong a_{jk}=a_{ik}$ holds for every $i,j,k=1,\ldots,n$ since $w_i /w_j \cong w_j /w_k$ is expected to be equal to w_i /w_k. Although every consistent matrix is reciprocal, the converse is not generally true. In practice, comparing of s_i to s_j, s_j to s_k, and s_i to s_k often results in inconsistency amongst the assessments in addition to their inaccuracy; however, the inconsistency may be computed and used to improve the accuracy.

The first step in pairwise comparisons is to establish the relative preference of each combination of two criteria. A scale from 1 to 5 can be used to compare all criteria in pairs. Values from the interval $[1/5,1]$ reflect inverse relationships between criteria since $s_i/s_j=1/(s_j/s_i)$. The consistency-driven approach is based on the reasonable assumption that by finding the most inconsistent judgments, one can then reconsider one's own assessments. This in turn contributes to the improvement of judgmental accuracy. Consistency analysis is a dynamic process which is assisted by the software.

The central point of the inference theory of the pairwise comparisons is Saaty's Theorem, [8], which states that for every n by n consistent matrix $A=[a_{ij}]$ there exist positive real numbers $w_1,\ldots,w_n$ (weights corresponding to criteria $s_1,\ldots,s_n$) such that $a_{ij}=w_i/w_j$ for every $i,j=1,\ldots,n$. The weights w_i are unique up to a multiplicative constant. Saaty (1977) also discovered that the eigenvector corresponding to the largest eigenvalue of A provides weights w_i which we wish to obtain from the set of preferences a_{ij}. This is not the only possible solution to the weights problem. In the past, a least-squares-solution was known, but it was far more computationally demanding than finding an eigenvector of a matrix with positive elements. Later, a method of row geometric means was proposed (Jensen, 1984), which is is the simplest and the most effective method of finding weights. A statistical experiment demonstrated that the accuracy, that is, the distance from the original matrix A and the matrix AN reconstructed from weights with elements

$[a_{ij}]=[w_i/w_j]$, does not strongly depend on the method. There is, however, a strong relationship between the accuracy and consistency. Consistency analysis is the main focus of the consistency-driven approach.

An important problem is how to begin the analysis. Assigning weights to all criteria (e.g., $A=18$, $B=27$, $C=20$, $D=35$) seems more natural than the above process. In fact it is a recommended practice to start with some initial values. The above values yield the ratios: $A/B=0.67$, $A/C=0.9$, $A/D=0.51$, $B/C=1.35$, $B/D=0.77$, $C/D=0.57$. Upon analysis, these may look somewhat suspicious because all of them round to 1, which is of equal or unknown importance. This effect frequently arises in practice, and experts are tempted to change the ratios by increasing some of them and decreasing others (depending on knowledge of the case). The changes usually cause an increase of inconsistency which, in turn, can be handled by the analysis because it contributes to establishing more accurate and realistic weights. The pairwise comparisons method requires evaluation of all combinations of pairs of criteria, and can be more time-consuming because the number of comparisons depends on n^2 (the square of the number of criteria). The complexity problem has been addressed and partly solved by the introduction of hierarchical structures [8]. Dividing criteria into smaller groups is a practical solution in cases in which the number of criteria is large.

Appendix B

Consistency Analysis

Consistency analysis is critical to the approach presented here because the solution accuracy of *not-so-inconsistent* matrices strongly depends on the inconsistency. The consistency-driven approach is, in brief, the next step in the development of pairwise comparisons.

The challenge to the pairwise comparisons method comes from a lack of consistency in the pairwise comparisons matrices which arises in practice. Given an *n* by *n* matrix *A* that is not consistent, the theory attempts to provide a consistent *n* by *n* matrix *AN* that differs from matrix *A* Aas little as possible@. In particular, the geometric means method produces results similar to the eigenvector method (to high accuracy) for the ten million cases tested. There is, however, a strong relationship between accuracy and consistency

Unlike the old eigenvalue-based inconsistency, introduced in [8], the triad-based inconsistency locates the most inconsistent triads [3]. This allows the user to reconsider the assessments included in the most inconsistent triad.

Readers might be curious, if not suspicious, about how one could arrive at values such as *1.30* or *1.50* as relative ratio judgments. In fact the values were initially different, but have been refined and the final weights have been calculatedby the consistency analysis. It is fair to say that making comparative judgments of rather intangible criteria (e.g., overall alteration and/or mineralization) results not only in imprecise knowledge, but also in inconsistency in our own judgments. The improvement of knowledge by controlling inconsistencies in the judgments of experts, that is, the *consistency-driven approach*, is not only desirable but is essential.

In practice, inconsistent judgments are unavoidable when at least three factors are independently compared against each other. For example, let us look closely at the ratios of the four criteria *A*, *B*, *C*, and *D*. Suppose we estimate ratios *A/B* as *2*, *B/C* as *3*, and *A/C* as *5*. Evidently something does not add up as *(A/B)≅ (B/C)=2≅ 3=6* is not equal to *5* (that is *A/C*). With an inconsistency index of *0.17*, the above triad (with highlighted values of *2*, *5*, and 3) is the most inconsistent in the entire matrix. A rash judgment may lead us to believe that *A/C* should indeed be *6*, but we do not have any reason to reject the estimation of *B/C* as *2.5* or *A/B* as *5/3*. After correcting *B/C* from *3* to *2.5*, which is an arbitrary decision usually based on additional knowledge gathering, the next most inconsistent triad is *(5,4,0.7)* with an inconsistency index of *0.13*. An adjustment of *0.7* to *0.8* makes this triad fully consistent (*5≅ 0.8* is *4*), but another triad *(2.5,1.9,0.8)* has an inconsistency of *0.05*. By changing *1.9* to *2* the entire table becomes fully consistent. The corrections for real data are done on the basis of professional experience and case knowledge by examining all three criteria involved.

An acceptable threshold of inconsistency is *0.33* because it means that one judgment is not more than two grades of the scale 1 to 5 Adifferent@ from the remaining two judgments. There was no need to continue decreasing the inconsistency, as only its high value is harmful; a very small value may indicate that the artificial data were entered hastily without reconsideration of former assessments.