An uncertainty aware medical diagnosis support system

Krzysztof Dyczkowski, Anna Stachowiak, Andrzej Wójtowicz, Patryk Żywica

Department of Imprecise Information Processing Methods
Faculty of Mathematics and Computer Science, Adam Mickiewicz University
Umultowska 87, 61-614 Poznań, Poland
chris@amu.edu.pl

Abstract. In the paper we describe a system that store and process uncertain data in such a way as to be able to obtain information essential to make an effective diagnosis while also indicating the uncertainty level of that diagnosis. We consider the problem of incompleteness and imprecision of medical data and discuss some issues connected with such kind of information - like modeling, making decision that is aware of the imperfection of data, evaluating results in the context of uncertain medical data. As an example we describe a method of supporting medical decision that is based on interval-valued fuzzy cardinality and that was implemented in the OvaExpert system.

Keywords: diagnosis support, medical data, uncertainty, imperfect information

1 Introduction

Computer decision-making systems are highly effective in terms of prognosis when solving many diagnostic problems. This is true especially for common diseases for which there is access to large number of cases. The situation is less satisfactory for diseases which are less common and thus the access to large number of well-depicted cases is limited. Lack of centralized system for gathering uniform data from many medical institutions is also a problem. If such databases exist they are gathered in a specific medical center and are not accessible to others. Another problem is lack of access to full required diagnostics (e.g. due to unavailability of proper diagnostic equipment or high cost of diagnostic examinations), which contributes to ambiguities and omissions in patient's record. In addition, by their very nature, medical descriptions are often imprecise and ambiguous. In most cases, they are descriptive and terminology used in them is not standardized. Their quality often depends on the education of the doctor (including the center where he or she was educated) as well as the doctor's experience. The existing situation calls for the use of unconventional data modeling and reasoning methods. It requires methods factoring in both the imprecision and incompleteness of the data. Those methods must ensure high efficacy for disease entities for which there are no sufficiently large databases available.

In this paper we demonstrate some part of a bigger concept - the OvaExpert system - that was meant to deal with a forementioned situation. A set of concepts and methods cover the problem at every stage - collecting, modeling and processing of uncertain data. They combine theoretical knowledge with the capabilities of a computer system. We propose how to maximize the use of such a system and of computing power to solve efficiently the problem of uncertain data.

In section 2 we give a brief view on the OvaExpert system and two research path that we have taken. Section 3 is devoted to one of the implemented methods, among many others, that supports gynaecologists in a diagnosis of ovarian tumors. In Section 4 we present the results of the analysis of methods based on counting. Section 5 gives some conclusions and areas for further research.

2 OvaExpert system - two research tracks

OvaExpert, the intelligent system for ovarian tumor diagnosis, introduces a completely novel approach to the imprecision connected with data imperfection (see [1–4]), The aim of the system is to store and process uncertain data in such a way as to be able to obtain information essential to make an effective diagnosis while also indicating the uncertainty level with which the information is suggested.

Traditionally, gynaecologists are assisted by many prognostic models, ultrasonographic morphological scales, and other risk of malignancy calculators that are used for differential diagnosis of ovarian tumors. The most common diagnostic models are based on scoring systems [5, 6] and logistic regressions [7]. Another predictive models were proposed by IOTA group: the most recent one is ADNEX [8].

The starting point for presented research was finding out that some of those models in some specific cases are complementing each other, i.e. applying them simultaneously yields better diagnostic efficacy as opposed to applying them separately (see [9]).

Consequently, there were two research tracks. The first one concerned the design of a decision model while the other involved using the synergy of the existing diagnostic models. Both tracks used interval-valued fuzzy sets in an epistemic sense which allowed us to include imperfect input data.

The first research track resulted in the concept of interval valued classifier based on similarity measures allowing imperfect input data. The results of this part of the research have been published in [10–13]. The method based on this algorithm will be marked as IVFC (method in the Ovaexpert system based on similarity measures).

The other research track involved using the method of aggregation/synergy of imperfect knowledge from several decision models. Our previous research has shown that fuzzy aggregation methods prove to be very effective in improving the quality of diagnosis and minimizing the impact of lack of data and imprecision.

This is due to the variety of models and their different levels of efficacy across different patient groups. Many models, when used simultaneously, considerably improve the quality of the decision. As a part of this research path we applied the theory of interval-value fuzzy set cardinality, that will be described in more details in the next section. This approach allows to make a decision supported by majority of data sources (models) preserving the information about the level of uncertainty about this decision. The other research centers are also currently developing this approach using intuitionistic fuzzy preference relations (cf. [14–17].

3 Algorithm for decision support based on interval-valued fuzzy set cardinality

Interval-valued fuzzy sets are a special variant of type-2 fuzzy sets, also introduced by Zadeh (see [18]). The notion of interval-valued fuzzy sets is a generalization of the notion of a usual fuzzy set. Its significant role is to introduce uncertainty as an actual value of membership function (epistemic interpretation of interval-valued fuzzy set (see [19])) that can be anywhere between the given interval values. Two approaches to interval-value fuzzy set cardinality were used: scalar (sigma f-Count) and fuzzy (f-FECount). Both make use of the cardinality patterns – functions that help determine the influence of single elements of an interval-valued fuzzy set on the value of its cardinality. In the case of interval-valued fuzzy set its cardinality is an interval or an interval-valued fuzzy set, a notion of interval representative was introduced to compare cardinalities. It is a single real number belonging to this interval. The most obvious interval representatives include: interval center, right limit (minimum value) and left limit (maximum value).

The idea behind the algorithm conforms with a usual method of making decisions by counting crisp sets. We make a decision supported by majority of data sources, on condition that they are more numerous that the reverse option by a specified value. If both options have the same support of decision sources (or the difference is minimal), then we do not decide. The idea behind decision algorithm is to use bipolar perspective on IVFS. Because an IVFS contains information about uncertainty level, it carries both information supporting and rejecting the decision. This property of IVFS is used in decision algorithm.
The basic idea behind this algorithm consists of a couple of steps:

1. On the basis of input data, we define two IVFSs: P "Pro" modeling support level for a positive decision and C (“Contra”) mirroring support level for a negative decision.
2. We calculate cardinalities of these IVFSs with the selected calculation method.
3. We compare cardinalities to find out whether we can make a decision i.e. whether one of them significantly outweighs the other, and if so we select the decision supported by greater cardinality.

In order to make a decision we need to determine a method for comparing cardinality intervals. For this purpose, we defined two approaches (modes):

– interval approach consisting in comparing overlap of intervals of respective distances between their endpoints,
– numerical approach consisting in determining numerical interval representatives.

Depending on selection of calculation methods and comparison methods we obtain various decision algorithms based sigma f-count from specific groups:

– SC-cen – based on interval center representatives
– SC-int – based on interval comparison method
– SC-max – based on left limit representatives.

and based on $f - FEcount$:

– FE-cen – based on interval center representatives
– FE-int – based on interval comparison method
– FE-max – based on left limit representatives.

An outline of the solution is presented in Fig. 1.

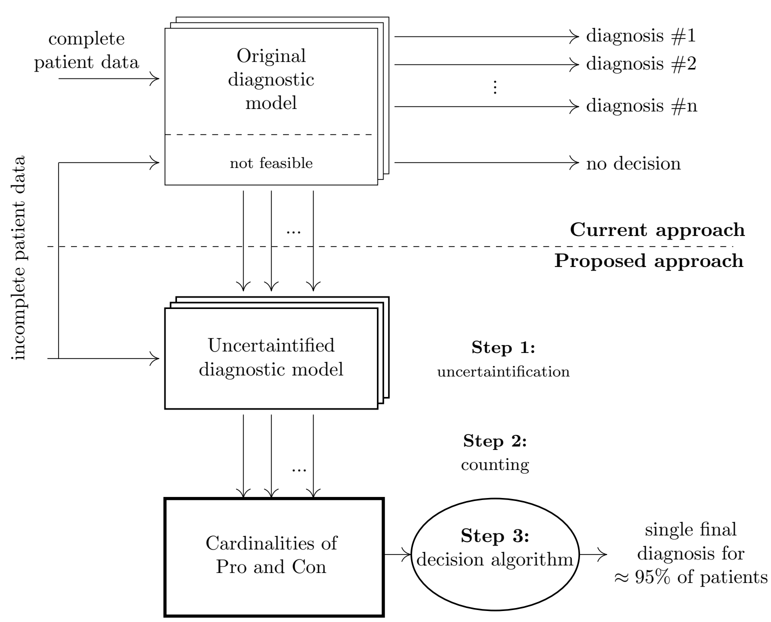

Fig. 1. A solution based on the cardinality of IVFSs.

The interval mode is much more restrictive and only efficient in situations with small amount of missing information (size of ignorance intervals). Definition of cardinality pattern is also of key importance. If using identity function as cardinality pattern (using sigma-count for calculating cardinality) cardinalities of both IVFSs P and C are symmetrical and decision is only made if both IVFSs are sufficiently similar to crisp sets. It is also important that thresholds (parameters defining cardinality pattern) are selected in such a way as to reduce the significance of input decisions close to 0.5.

4 Evaluation of efficiency - results

A very important aspect of construction and application of tumor malignancy classification (prediction) methods is to evaluate their efficiency (prediction quality). In a binary classification, we divide the decisions into two classes: positive (malignant tumors, which also include borderline malignant tumors which require the same treatment as ovarian cancer), and negative (benign tumors and non-neoplastic changes). In addition, in our research we allow a situation in which a classifier may not make a decision due to data being of too low quality. In medicine, numerous quality classification measures are applied, i.e. sensitivity, specificity, accuracy, f-measure etc. For classifiers that operate on data of poor quality (e.g. incomplete data), in some applications, it is necessary to consider a situation in which the classifier has insufficient information to make a sufficiently certain decision. This is often the case in medical applications when insufficiently certain decision can have serious consequences for the patient. This is why an additional measure has been introduced – decisiveness – which determines in how many cases the classifier was able to make a decision.

In many applications (often including medical ones) the above measures do not reflect the actual required quality of the classifier. This is the case when the significance of the individual classes of errors (actual effects of wrong decisions) are different. For example, in the medical diagnosis of ovarian tumors the situation when the system diagnoses a tumor as benign and, in fact, it was malignant causes much more significant effects for the patient as opposed to the situation when the benign tumor is diagnosed as cancer. In such models, the concept of cost matrix (cost function) is used where for each error type a weight (penalty) is assigned for a wrong decision. The quality value is the sum of costs (penalties) assigned to the classifier for making wrong decisions. Such a cost matrix will be used to evaluate the quality of classifications in our system.

The presented algorithms have been tested on real medical data. These data described 388 cases of patients diagnosed and treated in the Division of Gynecological Surgery, Poznan University of Medical Sciences, between 2005 and 2015. Out of them 61% have been diagnosed as suffering from benign tumors and 39% as suffering from malign tumors. Moreover, 56% of patients had full diagnostic (no test required by diagnostic scales was missing), 40% had significant amounts of missing data varying from 0% to 50%, and for the remaining ones 50% of data was missing. Detailed description of data used for evaluation can be found in [9]. More information on the data format and technical details can be found in [20].

Fig. 2 and 3 present classification results based on the proposed algorithms with the best versions obtained from optimization in specific groups. Efficiency

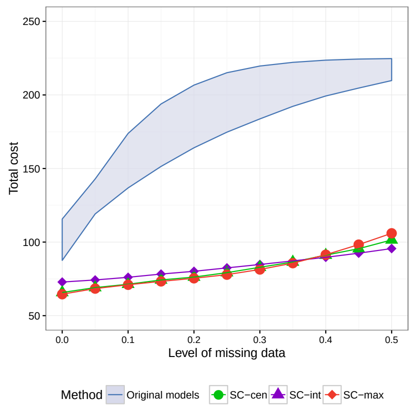

Fig. 2. Decision making efficacy of algorithms based on sigma f-count

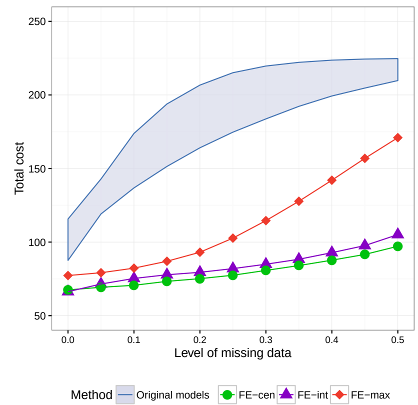

Fig. 3. Decision making efficacy of algorithm based on f-FEcount

area of original models has been marked in grey. The graphs show the total cost (the higher the cost the lower the classification quality) in relation to the level of missing data. The graph in Fig. 2 presents the best three algorithms based on sigma f-count: SC-cen, SC-int, SC-max, whereas, the graph presented in Fig. 3 presents three best algorithms based on $f - FEcount$: FE-cen, FE-int, FE-max.

As a result of analysis of the obtained decision efficiency, the algorithm FE-cen has been selected as the best for application in the OvaExpert system from amount the counting methods. A method based on this algorithm with the use of cardinality pattern is designated as FSC (the OvaExpert system method based on counting). The prognostic results of all three decision modules implemented in the OvaExpert system - OEA, IVFC and FSC - are presented in Table 1 and for the purpose of comparison the results for the original diagnostic models are also presented.

The original diagnostic models differ in their classification properties: some of them tend to make more conservative decisions (i.e. LR1, LR2, SM), and some of them are more liberal (i.e. RMI, Tim.). This can be observed in discernible differences in values between sensitivity and specificity. Only one of these models ensures the balance of both factors (Alc.). It should be noted that all original models have very low decisiveness (due to deficiencies in diagnostic data), which results in high total cost.

The new models implemented in the OvaExpert system have high sensitivity and specificity values. Two of them tend to be more conservative (OEA and IVFC), while FSC is more balanced. All three models provide a high level of decisiveness because they are able to deal with deficiencies in data. This is why their total cost is much lower than the original models.

It can be noted that the diagnostic models of the OvaExpert system differ significantly from the original models in terms of classification. Although diagnostic

		Total cost	Dec.	Sen.	Spec.	Acc.
Original models	Alc. [5]	189.0	20.6 %	88.2 %	89.5 %	88.9 %
	LR1 [21]	184.0	27.4 %	92.6 %	57.1 %	77.1 %
	LR2 [21]	164.0	33.1 %	94.3 %	65.2 %	82.8 %
	RMI [22]	156.0	56.6 %	75.9 %	87.1 %	83.8 %
	SM [23]	142.0	62.9 %	94.6 %	71.2 %	79.1 %
	Tim. [24]	159.0	47.4 %	66.7 %	97.1 %	91.6 %
New diag. modules	OEA	72.0	96.6 %	90.2 %	86.4 %	87.6 %
	IVFC	72.5	100.0 %	90.4 %	84.6 %	86.3 %
	FSC	67.0	93.7 %	90.0 %	90.2 %	89.4 %

Table 1. The results of the decision-making quality of the original models compared to the OvaExpert methods

modules differ in classification quality indicators, the differences in classification are not statistically significant.

In the light of these results, the OvaExpert system based on the presented modules is a promising tool for supporting the prognosis of ovarian cancers, especially in the case of partial gaps in diagnostic data that are common in the everyday medical practice.

5 Conclusions and Further Research

At the moment, the OvaExpert system is tested in several medical centers offering the diagnosis and treatment of gynecological tumors. Its further implementation depends on overcoming legal and organizational obstacles concerning medical systems in Poland.

The demo version of the system is available on the project website http://ovaexpert.pl/ where one can get acquainted with the functions and possibilities offered by the system.

Statistical evaluation and implementation of the proposed methods have been performed with R, version 3.1.2. Scripts, documentation and non-sensitive data are available at GitHub (see http://ovaexpert.github.io/ovarian-tumor-aggregation). Because of large amount of calculations needed to do the research, we did them using Microsoft Azure cloud service available to our team under Microsoft Azure Research Grant "Azure Machine Learning – Development of an Intelligent System for Ovarian Tumor Diagnosis".

The OvaExpert system was designed to take advantage of the synergy of many classic diagnostic models and those newly created based on the knowledge derived from the data. The system has implemented all well-known prediction models since medical specialists trust their results. Additionally, new diagnostic methods have been implemented - among others a method based on IVFSs cardinalities, described in this paper. This method, based on a solid theoretical foundation, is relatively easy to implement and interpret, and, most importantly, achieves very good effectiveness in real-life applications such as medical diagnostics. Our approach is meant to be adapted also to non-medical problems where data quality is a matter of concern. It could be applied when the information that comes from independent experts is imperfect and it is important to preserve information about this imperfection in the final result. By returning bipolar information – concerning the quantities of positive and negative premises – we are able to evaluate that imperfection and the quality of the information.

All of our effort may be summarized with the following achievements:

- Development of computational intelligence methods that help make decisions based on low quality data, in particular:
  - Development of representation and processing methods for low quality data using interval-valued fuzzy sets.
  - Development of selection and optimization methods for decision making algorithms based on interval-valued fuzzy sets.
  - Development of methods calculating the cardinalities of interval-valued fuzzy sets.
  - Development of decision making algorithms based on the cardinalities of interval-valued fuzzy sets.
- Application of the above-mentioned methods in designing the intelligent system OvaExpert supporting medical diagnosis.
- Pilot implementation of the OvaExpert system that supports gynecologists and helps gathering data for further research and development.

References

1. K. Dyczkowski, Intelligent Medical Decision Support System Based on Imperfect Information: The Case of Ovarian Tumor Diagnosis, Vol. 735, Springer, 2017.
2. K. Dyczkowski, A. Wójtowicz, P. Żywica, A. Stachowiak, R. Moszyński, S. Szubert, An Intelligent System for Computer-Aided Ovarian Tumor Diagnosis, in: Intelligent Systems'2014, Springer, 2015, pp. 335–343.
3. A. Wójtowicz, P. Żywica, et al., Dealing with Uncertainty in Ovarian Tumor Diagnosis, in: K. Atanassov, W. Homenda, et al. (Eds.), Modern Approaches in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Volume II: Applications, SRI PAS, Warsaw, 2014, pp. 151–158.
4. P. Żywica, K. Dyczkowski, A. Wójtowicz, A. Stachowiak, S. Szubert, R. Moszyński, Development of a fuzzy-driven system for ovarian tumor diagnosis, Biocybernetics and Biomedical Engineering 36 (4) (2016) 632–643.
5. J. L. Alcazar, L. T. Merce, et al., A new scoring system to differentiate benign from malignant adnexal masses, Obstetrical & Gynecological Survey 58 (7) (2003) 462–463.
6. D. Szpurek, R. Moszyński, et al., An ultrasonographic morphological index for prediction of ovarian tumor malignancy, European Journal of Gynaecological Oncology 26 (1) (2005) 51–54.
7. D. Timmerman, A. C. Testa, et al., Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group, Journal of Clinical Oncology 23 (34) (2005) 8794–8801.

8. B. Van Calster, K. Van Hoorde, et al., Evaluating the risk of ovarian cancer before surgery using the adnex model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study, BMJ 349 (2014) 5920.
9. R. Moszyński, P. Żywica, et al., Menopausal status strongly influences the utility of predictive models in differential diagnosis of ovarian tumors: An external validation of selected diagnostic tools, Ginekologia Polska 85 (12) (2014) 892–899.
10. A. Stachowiak, P. Żywica, K. Dyczkowski, A. Wójtowicz, An interval-valued fuzzy classifier based on an uncertainty-aware similarity measure, in: Advances in Intelligent Systems and Computing, Vol. 322, Springer, Berlin Heidelberg, 2015, pp. 741–751.
11. P. Żywica, Similarity Measures of Interval–Valued Fuzzy Sets in Classification of Uncertain Data. Applications in Ovarian Tumor Diagnosis, Ph.D. thesis, Faculty of Mathematics and Computer Science of Adam Mickiewicz University, in Polish (June 2016).
12. P. Żywica, A. Stachowiak, M. Wygralak, An algorithmic study of relative cardinalities for interval-valued fuzzy sets, Fuzzy Sets and Systems 294 (2016) 105 – 124.
13. P. Żywica, A. Stachowiak, A new algorithm for computing relative cardinality of intuitionistic fuzzy sets, in: K. Atanassov, et al. (Eds.), Modern Approaches in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Volume I: Foundations, IBS PAN – SRI PAS, Warsaw, 2014, pp. 181–189.
14. U. Bentkowska, A. Król, Preservation of fuzzy relation properties based on fuzzy conjunctions and disjunctions during aggregation process, Fuzzy Sets and Systems 291 (2016) 98–113.
15. U. Dudziak, B. Pękala, Intuitionistic fuzzy preference relations, in: Proceedings of the 7th conference of the European Society for Fuzzy Logic and Technology, Atlantis Press, 2011, pp. 529–536.
16. B. Pękala, Properties of interval-valued fuzzy relations, Atanassov's operators and decomposable operations, in: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Springer, 2010, pp. 647–655.
17. B. Pękala, Operations on interval matrices, Rough Sets and Intelligent Systems Paradigms (2007) 613–621.
18. L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning—I, Information sciences 8 (3) (1975) 199–249.
19. D. Dubois, W. Liu, J. Ma, H. Prade, The basic principles of uncertain information fusion. an organised review of merging rules in different representation frameworks, Information Fusion 32 (2016) 12–39.
20. A. Wójtowicz, P. Żywica, et al., Solving the problem of incomplete data in medical diagnosis via interval modeling, Applied Soft Computing 47 (2016) 424–437.
21. D. Timmerman, B. Van Calster, A. C. Testa, S. Guerriero, D. Fischerova, A. Lissoni, C. Van Holsbeke, R. Fruscio, A. Czekierdowski, D. Jurkovic, et al., Ovarian cancer prediction in adnexal masses using ultrasound-based logistic regression models: a temporal and external validation study by the iota group, Ultrasound in Obstetrics & Gynecology 36 (2) (2010) 226–234.
22. I. Jacobs, D. Oram, et al., A risk of malignancy index incorporating ca 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer, BJOG: An International Journal of Obstetrics & Gynaecology 97 (10) (1990) 922–929.

23. D. Szpurek, R. Moszynski, W. Zietkowiak, M. Spaczynski, S. Sajdak, An ultrasonographic morphological index for prediction of ovarian tumor malignancy, European Journal of Gynaecological Oncology 26 (1) (2005) 51–54.
24. D. Timmerman, T. H. Bourne, A. Tailor, W. P. Collins, H. Verrelst, K. Vandenberghe, I. Vergote, A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: the development of a new logistic regression model, American Journal of Obstetrics and Gynecology 181 (1) (1999) 57–65.