Study of Similarity Measures for Case-Based Reasoning in Transcatheter Aortic Valve Implantation

Hélène Feuillâtre[1,2], Vincent Auffret[1,2,3], Miguel Castro[1,2], Hervé Le Breton[1,2,3], Mireille Garreau[1,2], Pascal Haigron[1,2]

[1]INSERM U1099, Rennes, France
[2]Université de Rennes 1, LTSI, Rennes, France
[3]CHU Rennes, Service de Cardiologie et Maladies Vasculaires, Rennes, France

Abstract

Case-Based Reasoning (CBR) uses previous experiences to solve similar current problems. The basic hypothesis is that similar cases should have similar solutions. In the case of Transcatheter Aortic Valve Implantation (TAVI), the CBR could help practitioners to plan the procedure. Four steps compose a CBR: retrieve, reuse, revise and retain. Defining a convenient similarity measure (SM) is essential in the retrieve step. This study aims to analyze the performance of different similarity measures and attribute selections. Generally in the retrieve step, a standard weighted heterogeneous similarity measure (WHSM) is used, in association with the k-nearest neighbor algorithm. Based on WHSM, we considered new definitions of SMs dedicated to decision support for TAVI. They include attributes selection and weight determination through a clinical decision tree. The performance of SMs was evaluated on a set of 100 cases with a leave-one-out cross validation. Results show that the CBR retrieving process can be improved by using dedicated SMs.

1. Introduction

The experience can play an essential role in decision making when facing new problems. Case-Based Reasoning (CBR) makes the assumption that past experience can be useful to solve similar current problems. CBR systems are structured in four main steps (retrieve, reuse, revise and retain). They allow retrieving, from a case-base, a set of the most similar cases to a new case and to take the decision about the most suitable solution to the problem.

CBR have already been applied in different domain such as in statistical quality control [1], chemical engineering [2] or health science [3]. In our study the CBR is aimed at the management of valvular heart disease and especially at the planning of Transcatheter Aortic Valve Implantation (TAVI). The feasibility of designing CBR for TAVI has been previously reported in El-Fakdi et al. [4]. Their work concentrated on the overall framework and did not focus on investigating similarity functions. A classical definition of the similarity measure was used and only simple representations of cases were considered. Our work focuses on the most computational part of CBR and addresses the issue of defining a relevant similarity measure to retrieve similar past cases.

Wilson and Martinez [5], Lesot et al. [6] and Choi et al. [7] proposed different comparison studies about similarity measures used in various applications (data mining, data analysis or information retrieval, etc.). Other research works studied similarity measures specifically in CBR context, such as Liao et al. [8], Núñez et al. [9], Avramenko and Kraslawki [2] or more recently Gu et al. [10]. These different papers highlight that the type of the different attributes representing a case influences the performance of the similarity measure, as their degree of importance and the consideration of missing values.

In the retrieve step, most of the CBR systems used a generalized weighted similarity measure. Commonly, a distance measure is used to compute the dissimilarity between attributes of two cases. A diversity of distance measures is available such as the Minkowski, Camberra, Chebychev, Mahalanobis, Cosine or Jaccard metrics [5–7]. A large amount of CBR approaches used the weighted Euclidean distance. Even if most of the attributes are quantitative, the Euclidean distance and the other distance metrics are not suitable for all data types. The Euclidean distance is more appropriate for continuous quantitative values. However, to overcome this problem, some works [1,2] converted the ordinal attributes to discrete values. Another solution is to use a heterogeneous distance

 ISSN: 2325-887X DOI:10.22489/CinC.2017.134-299

measure [4,5,10]. Wilson and Martinez [5] proposed a distance function, the Heterogeneous Euclidean-Overlap Metric (HEOM), which uses the overlap metric for qualitative (i.e. nominal) attributes and the normalized Euclidean metric for quantitative attributes.

According to the decision to make, different issues have to be addressed in defining the similarity measure such as the choice of metrics, the selection of attributes used, their degree of importance and their modes of combination. This paper analyses the performance of different similarity measures and the influence of attributes selection through a clinical decision tree (CDT). In the following we first summarize the main features of the CBR. We then describe the implemented similarity measures and report the results obtained for decision-making about valve bioprosthesis in TAVI.

2. CBR for TAVI procedure

The aim of the CBR system is to provide the practitioner a selection of cases most relevant to the current candidate patient to plan the TAVI procedure (which vascular access, which type and size of prosthesis). The operating of the proposed CBR is based on the human-machine cooperation.

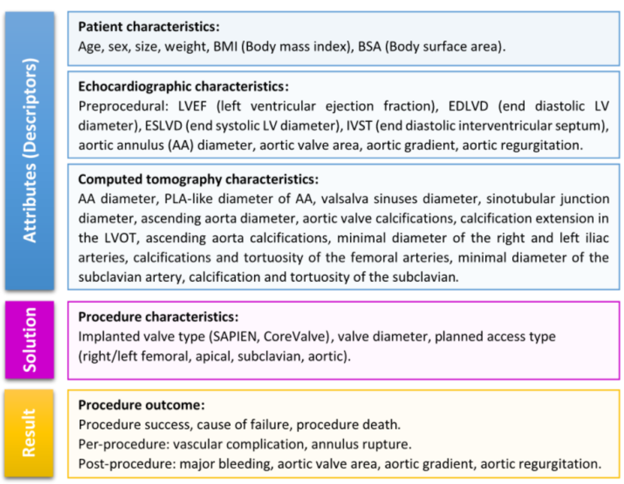

Figure 1: Description of a patient in the current case-base.

A case, i.e. a patient, which is the central notion in a CBR system, represents the medical experience of physicians. A set of past cases allow building a case-base. Each case is composed of three categories (Figure 1): the description of the problem (clinical attributes from patient characteristics and medical imaging), the solution (procedure characteristics) and the results (procedure outcome). Clinical attributes are used in different steps of the CBR process: their resemblance in different patients is exploited to propose relevant solution, i.e. suitable information concerning the TAVI procedure.

The CBR solving cycle is composed on four steps (Figure 2) [11]:

- RETRIEVE: to get a set of the most similar cases to the current candidate patient.
- REUSE: to take the decision about the best solution from the set of retrieved similar cases (i.e. to adapt similar cases to the current patient).
- REVISE: to get the result after application of the solution and to complete the case (procedural characteristics and outcomes).
- RETAIN: to update the case-base (useful experience is retained for future reuse).

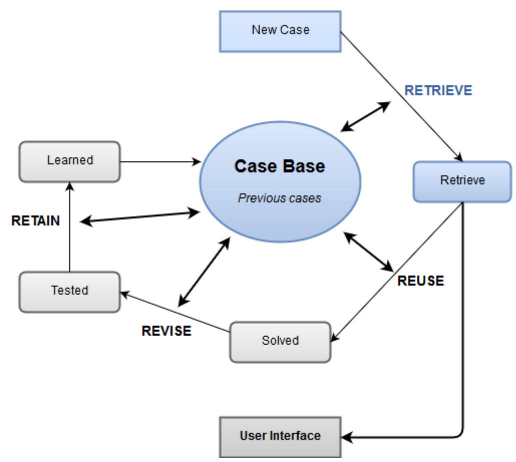

Figure 2: The four cycle of the CBR.

The retrieved cases represent the start point for finding the best solution to the problem. The retrieve step requires data processing to evaluate reliably the similarity between cases and to recover relevant past cases. To retrieve similar cases, a distance is computed between the current candidate patient and each patient present in the case-base through a similarity measure. After applying the similarity measure on the different cases present in the case-base, a k-nearest neighbors algorithm (k-NN) is used to obtain the k most similar cases. In our approach, the retrieve step represents the most computational part of the CBR. The other steps are realized by the practitioner through the user interface in order to leave her/him the final choice for decision making.

3. Similarity Measures

Since the practitioner choose the best solution for the current procedure among the past similar cases, defining a relevant and efficient similarity measure is essential.

In the case-base, the different attributes have different data types (Figure 1). For the definition of the similarity measure (SM), different metrics can be used according to the type of attributes. In the following, three different SMs dedicated to the planning of the TAVI procedure are defined. They are based on the weighted heterogeneous similarity measure (WHSM).

In *WHSM1*, no attribute's selection is performed and each attribute in the case-base (Figure 1) has the same importance in the decision making. *WHSM1* is defined by $1 - diss(C_i, C_j)$, where the distance metric (Eq. 1) is computed from the two cases C_i and C_j. w_p corresponds to the weight of the attribute p (fixed at 1) and q represents the number of attributes used in the measure. $d(C_{ip}, C_{jp})$ represents the difference between the values of the attribute p in the cases C_i and C_j. If p is a quantitative attribute, the Euclidean distance is used. The Hamming distance is computed for qualitative attributes corresponding to Boolean category. For each kind of qualitative attributes which are ordered (i.e. ordinal attributes), a distance matrix is built according to expert knowledge. The distance between attributes in the matrix are normalized in the range [0,1]. For example, in the similarity matrix corresponding to the tortuosity attribute, the distance between the two values *Mild* and *Moderate* is 0.2, i.e. $ord(C_{ip}, C_{jp}) = 0.2$ with $C_{ip} = Mild$ and $C_{jp} = Moderate$. When attribute is missing, the neutral approach is chosen to give directly the value 0.5.

$$diss(C_i, C_j) = \sqrt{\frac{\sum_p^q w_p \, d(C_{ip}, C_{jp})^2}{\sum_p^q w_p}} \quad (1)$$

Where $d(C_{ip}, C_{jp})$

$$= \begin{cases} d_e(C_{ip}, C_{jp}) & if\ C_{ip} and\ C_{jp}\ are\ quantitative \\ d_H(C_{ip}, C_{jp}) & if\ C_{ip} and\ C_{jp}\ are\ binary \\ ord(C_{ip}, C_{jp}) & if\ C_{ip} and\ C_{jp}\ are\ ordinal \\ mis(C_{ip}, C_{jp}) & if\ C_{ip}\ or\ C_{jp}\ is\ missing \end{cases}$$

With $d_e(C_{ip}, C_{jp}) = \dfrac{|C_{ip} - C_{jp}|}{range_p}$

$$d_H(C_{ip}, C_{jp}) = \begin{cases} 0, & if C_{ip} = C_{jp} \\ 1, & if C_{ip} \neq C_{jp} \end{cases} for\ C_{ip}, C_{jp} \in \{yes,\ no\}$$

In *WHSM2*, a selection of relevant attributes is applied compared to *WHSM1*. This selection is performed thanks to clinical decision trees (CDT) (Figure 3) which are built from expert knowledge and clinical guidelines [12]. One clinical decision tree is available for each decision (choice of prosthesis or choice of the vascular access for TAVI procedure). In this second SM, only attributes considered in the clinical decision tree are used. As previously, no weights are considered ($w_p = 1$).

We introduced a third version of SM called Hierarchical Heterogeneous Similarity Measure (*HHSM*). The clinical decision tree was used in another way to determine the weight of the attributes. Besides the selection of relevant attributes, *HHSM* selects gradually the most similar cases. The expression of the metric constituting the *HHSM* is adapted according to each level l of the CDT:

$$diss_l(C_i, C_j) = \frac{\sum_p^{q_l} w_p \, d(C_{ip}, C_{jp})}{\sum_p^{q_l} w_p} \quad (2)$$

First, only attributes at the first level of the CDT ($l = 1$) are considered in the similarity measure (Figure 3). Next, a selection of cases is made. According to the distance metric, only half on the most similar cases are kept. From these retained cases, the update of the *HHSM* is computed according to the next level of the CDT (l is incremented by 1). This next iteration takes into account the attributes both at the previous levels and at the current level l of the CDT. The proposed *HHSM* allows selecting iteratively the most relevant attributes and the most similar cases. At a given iteration (level) only the most relevant attributes are used so that the least similar cases are directly removed for the next iteration.

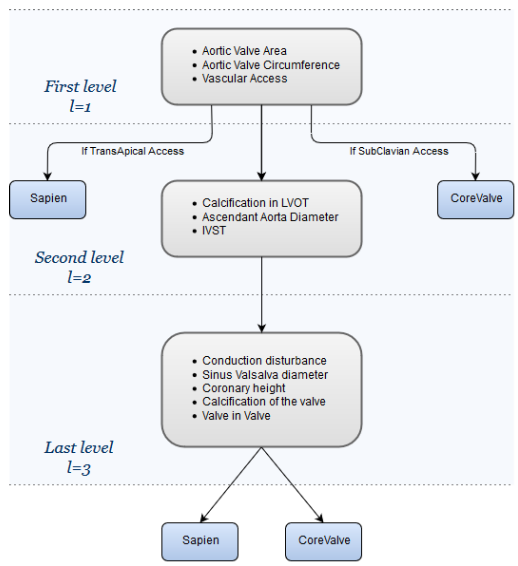

Figure 3: Example of a simplified clinical decision tree used for the choice of prosthesis.

4. Results

The different similarity measures were evaluated from a case-base containing the data (attributes) of 100 patients who underwent a TAVI procedure. A leave-one-out cross validation was performed to evaluate the similarity measures for two decision making: type and size of prosthesis.

Table 1 represents the percentage of correct solutions which appear at least once for the type and the size of the prosthesis when three most similar cases are selected ($k = 3$ in the k-NN algorithm). The correct solution about the size and the type of the prosthesis appears at least once in almost 90% of cases when the CBR is operated with the *HHSM*. When $k = 5$, this result tends towards 100%. Results show that considering a SM based on the clinical decision tree defined for a specific decision improves the retrieve step. Indeed, the percentage of correct solutions appearing at least once for the choice of prosthesis increases in *WHSM2* and *HHSM*.

Table 1: Percentage of correctly classified cases when three most similar patients are selected.

Similarity Measure	Prosthesis type	Prosthesis size
WHSM1	84%	82%
WHSM2	88%	88%
HHSM	89%	90%

5. Conclusion

This paper compared three similarity measures used to obtain a set of similar cases to help physicians to take decision for the planning of TAVI procedure. This first work has shown that constructing a dedicated similarity measure improves the CBR performance. Using clinical decision tree defined specifically for TAVI procedure improves the CBR retrieval process.

Further works will include the analysis of *HHSM* in a case-base containing more cases and additional types of attributes. Moreover, *HHSM* based CBR and its evaluation have to be extended to address the whole decision process involved in TAVI planning.

Acknowledgements

This work was partially supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 689617: EurValve, and by the French National Research Agency (ANR) in the framework of the Investissement d'Avenir Program through Labex CAMI (ANR-11- LABX-0004).

References

[1] Behbahani M, Saghaee A, Noorossana R. A case-based reasoning system development for statistical process control: Case representation and retrieval. Computers & Industrial Engineering 2012;63:1107–17.

[2] Avramenko Y, Kraslawski A. Similarity concept for case-based design in process engineering. Computers & Chemical Engineering 2006;30:548–57.

[3] Begum S, Ahmed MU, Funk P, Xiong N, Folke M. Case-Based Reasoning Systems in the Health Sciences: A Survey of Recent Trends and Developments. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 2011;41:421–34.

[4] El-Fakdi A, Gamero F, Meléndez J, Auffret V, Haigron P. eXiTCDSS: A framework for a workflow-based CBR for interventional Clinical Decision Support Systems and its application to TAVI. Expert Systems with Applications 2014;41:284–94.

[5] Wilson DR, Martinez TR. Improved heterogeneous distance functions. Journal of Artificial Intelligence Research 1997;6:1–34.

[6] Lesot MJ, Rifqi M, Benhadda H. Similarity measures for binary and numerical data: a survey. International Journal of Knowledge Engineering and Soft Data Paradigms 2009;1:63.

[7] Choi S-S, Cha S-H, Tappert CC. A survey of binary similarity and distance measures. Journal of Systemics, Cybernetics and Informatics 2010;8:43–48.

[8] Liao TW, Zhang Z, Mount CR. Similarity measures for retrieval in case-based reasoning systems. Applied Artificial Intelligence 1998;12:267–88.

[9] Núñez H, Sànchez-Marrè M, Cortés U, Comas J, Martínez M, Rodríguez-Roda I, et al. A comparative study on the use of similarity measures in case-based reasoning to improve the classification of environmental system situations. Environmental Modelling & Software 2004;19:809–19.

[10] Gu D, Liang C, Zhao H. A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis. Artificial Intelligence in Medicine 2017.

[11] Aamodt A, Plaza E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications 1994:39–59.

[12] The Task Force for the Management of Valvular Heart Disease of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS). 2017 ESC/EACTS Guidelines for the management of valvular heart disease. European Heart Journal 2017.

Address for correspondence.

H. Feuillâtre.
LTSI INSERM U1099, Université de Rennes 1,
Campus de Beaulieu, 35042 Rennes Cedex, France.
helene.feuillatre@univ-rennes1.fr