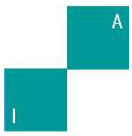
Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial
ISSN: 1137-3601
revista@aepia.org
Asociación Española para la Inteligencia Artificial
España

Coelho, Ana; Marques, Paulo; Magalhães, Ricardo; Sousa, Nuno; Neves, José; Alves, Victor

A Knowledge Representation and Reasoning System for Multimodal Neuroimaging Studies

Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 20, núm. 59, marzo, 2017, pp. 1-12
Asociación Española para la Inteligencia Artificial
Valencia, España

Inteligencia Artificial 20(59) (2017), 42-52
doi: 10.4114/intartif.vol20iss59pp42-52



INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

A Knowledge Representation and Reasoning System for Multimodal Neuroimaging Studies

Ana Coelho[1], Paulo Marques[2], Ricardo Magalhães[2], Nuno Sousa[2], José Neves[1] and Victor Alves[1]

[1] Algorithm Centre, University of Minho, Braga, Portugal acoelho.tkd@gmail.com, {jneves, valves}@di.uminho.pt

[2] Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal {paulo.c.g.marques, ricardomagalhaes, njcsousa}@med.uminho.pt

Abstract Multimodal neuroimaging analyses are of major interest for both research and clinical practice, enabling the combined evaluation of the structure and function of the human brain. These analyses generate large volumes of data and consequently increase the amount of possibly useful information. Indeed, BrainArchive was developed in order to organize, maintain and share this complex array of neuroimaging data. It stores all the information available for each participant/patient, being dynamic by nature. Notably, the application of reasoning systems to this multimodal data has the potential to provide tools for the identification of undiagnosed diseases. As a matter of fact, in this work we explore how Artificial Intelligence techniques for decision support work, namely Case-Based Reasoning (CBR) that may be used to achieve such endeavour. Particularly, it is proposed a reasoning system that uses the information stored in BrainArchive as past knowledge for the identification of individuals that are at risk of contracting some brain disease.

Keywords: Multimodal Neuroimaging, Case-based Reasoning, Artificial Intelligence.

1 Introduction

Information sharing and storing is one of the fundamental practices in neurosciences research, in particular, and in modern society, in general. This paradigm, associated with the vast amount of data generated from acquisition, processing and analysis of Magnetic Resonance Imaging (MRI) data, give rise to the need for specific tools to make easier its management [1]. Multimodal neuroimaging studies can be extremely important at different levels in neuroscience research and clinical practice, from the definition of basic concepts of brain function and organization, to the identification of hallmarks of different pathologies and even surgical planning [2]–[4] .

Neuroimaging studies usually involve groups of different areas where each one undergoes a multimodal imaging protocol with several different imaging modalities (functional and structural acquisitions). This generates a large amount of data and its processing and analysis will make it even heavier.

On the one hand, there are various approaches to manage and store relevant medical information, specially the medical imaging one [5]–[8]. One of these approaches stands for the BrainArchive, which was developed in order to organize, maintain and share data of multimodal brain MRI studies in an efficient and automated way [1]. It stores all the information available for each participant/patient, namely, the studies that the subjects underwent and the results of the processing and analysis of such studies. On the other hand, it must be stated that the information stored in the system is dynamic by nature.

Indeed, reasoning systems can be used to extract information from available knowledge. In a system like BrainArchive, knowledge can be extracted from the multimodal brain data, and be used to deduce more

ISSN: 1137-3601 (print), 1988-3064 (on-line)

information and possible conclusions, instead of simply storing it, i.e., it provides tools for the identification of undiagnosed diseases, and for aiding in neuroscience BrainArchive studies.

BrainArchive denotes a platform to store neuroimaging data, that was developed with the goal to provide an efficient and automated way of organizing, storing and sharing data of multimodal brain MRI studies. It is based in a NoSQL architecture, and the running application recognizes the files to be stored through the use of a specific hierarchy of folders and file nomenclature. This allows to export/import a full study (comprising one or more subjects, each consisting of multiple files) in a single step. Thus the sharing process of data is facilitated, since it is not necessary to export or import each file individually. Also, the system is distributed in order to store data as documents, therefore enabling users to upload and retrieve files to/from the system in different locations. This enhances the research process, through the simplification and reduction of the time spent organizing and sharing information.

Exported files are characterized by a set of meta-information (e.g. study identifier, subject identifier, analysis and type of file) that is saved automatically. This information is used to filter the files in the repository, so the user can extract only the desired data.

Since the data stored in the BrainArchive system is of a medical nature, security issues and access permissions were also considered when the application was under development, and will be provided to the inference mechanisms under use. So, to login in the application, authentication is required and data transfer is done using Secure Sockets Layer (SSL), between the BrainArchive application and the repository. Furthermore, each user has the possibility to change other users' access permissions to the data exported by him. In this way it is possible to manage the availability of files in the repository, making it only accessible to the users who inserted the data, specific users or even to every user.

BrainArchive features make it a good solution to facilitate the user in the process of organizing and sharing neuroimaging data. The use of specific hierarchy of folders and file nomenclature reduces the time spent in this process, once large volumes of information are identified without requiring the manual setting of the individual`s files. of what each individual file represents. Also, it enables data recovery, reuse of information and the access to data independently of the system used or location. Thus, the process of sharing data between researchers and/or institutions is simplified.

The data types stored in BrainArchive are very diverse since the information stored is the result of different acquisition types (e.g. structural, functional and diffusion). It is possible to find information about brain regions, volumes and areas, brain activation during specific tasks, functional and structural connectivity, among others. All these different types of information can be combined in a reasoning system with the goal to get the right background to the identification of undiagnosed diseases.

In this paper, it will be presented an Artificial Intelligence based Decision Support System to Multimodal Neuroimaging Studies centred on a formal framework based on Logic Programming for Knowledge Representation and Reasoning, complemented with a Case-based approach to computing. This approach will allow to combine a reasoning system with a platform like BrainArchive and thus extract knowledge from the multimodal brain data stored in BrainArchive. The extracted knowledge can be further used to infer more information and possible different outcomes, like the identification of undiagnosed diseases.

2 Knowledge Representation

Logic Programming (LP) has been used for knowledge representation and reasoning, representing a point of convergence in the disciplines of Logic, Mechanical Theorem Proving and Computer Science. It may be given in terms of elements of Model Theory [9]–[11], or Proof Theory [12], [13]. In the present work the Proof Theoretical approach is followed as an extension to LP. Indeed, an Extended Logic Program is a finite set of clauses in the form:

{

$\neg p \leftarrow \text{not } p, \text{not exception}_p$ (that stand for the predicate's clousure)

$p \leftarrow p_1, \cdots, p_n, \text{not } q_1, \cdots, \text{not } q_m$

$?\,(p_1, \cdots, p_n, \text{not } q_1, \cdots, \text{not } q_m) \quad (n, m \geq 0)$

$$\text{exception}_{p_1}$$
$$\cdots$$
$$\text{exception}_{p_j} \ (0 \leq j \leq k), \text{ being k and integer}$$

$$\} :: \text{scoring}_{value}$$

where "?" is a domain atom denoting falsity, the p_i, q_j, and p are classical ground literals, i.e., either positive atoms or atoms preceded by the classical negation sign ¬ [12]. Indeed, ¬ stands for a strong declaration that speaks for itself, and not denotes negation-by- failure, or in other words, a flop in proving a given statement, once it was not declared explicitly. Under this formalism, every program is associated with a set of abducibles [10], [11], given here in the form of exceptions to the extensions of the predicates that make the program, i.e., clauses of the form:

$$\text{exception}_{p_1} \ \cdots \ \text{exception}_{p_j} \ (0 \leq j \leq k), \text{ being k and integer}$$

that stand for information or knowledge that cannot be ruled out. On the other hand, clauses of the type:

$$?\ (p_1, \cdots, p_n, \text{not } q_1, \cdots, \text{not } q_m) \ \ (n, m \geq 0)$$

also named invariants or restrictions to complain with the universe of discourse, set the context under which it may be understood. The term scoring_{value} stands for the relative weight of the extension of a specific predicate with respect to the extensions of peers ones that make the inclusive or global program.

2.1 Quantitative Knowledge

In order to set one's approach to knowledge representation, two metrics will be set, namely the Quality-of-Information (*QoI*) of a logic program that will be understood as a mathematical function that will return a truth-value ranging between 0 and 1 [12], [13], once it is fed with the extension of a given predicate, i.e., $QoI_i = 1$ when the information is *known* (*positive*) or *false* (*negative*) and $QoI_i = 0$ if the information is *unknown*. For situations where the extensions of the predicates that make the program also include *abducible* sets, its terms (or clauses) present a $QoI_i \in]0, 1[$, in the form:

$$QoI_i = {}^{1}/_{Card} \tag{1}$$

if the *abducible* set for *predicates i* and *j* satisfy the *invariant*:

$$?\left(\left(exception_{p_i};\ exception_{p_j}\right), \neg\ \left(exception_{p_i};\ exception_{p_j}\right)\right)$$

where ";" denotes " *logical or*" and "*Card"* stands for set cardinality, being $i \neq j$ and $i, j \geq 1$ (a pictorial view of this process is given in Figure 1(a), as a pie chart).

On the other hand, the clauses cardinality (*K*) will be given by $C_1^{Card} + \cdots + C_{Card}^{Card}$, if there is no constraint on the possible combinations among the abducible clauses, being the *QoI* acknowledged as:

$$QoI_{i_{1 \leq i \leq Card}} = {}^{1}/_{C_1^{Card}}, \cdots, {}^{1}/_{C_{Card}^{Card}} \tag{2}$$

where C_{Card}^{Card} is a card-combination subset, with *Card* elements. A pictorial view of this process is given in Figure 1(b), as a pie chart.

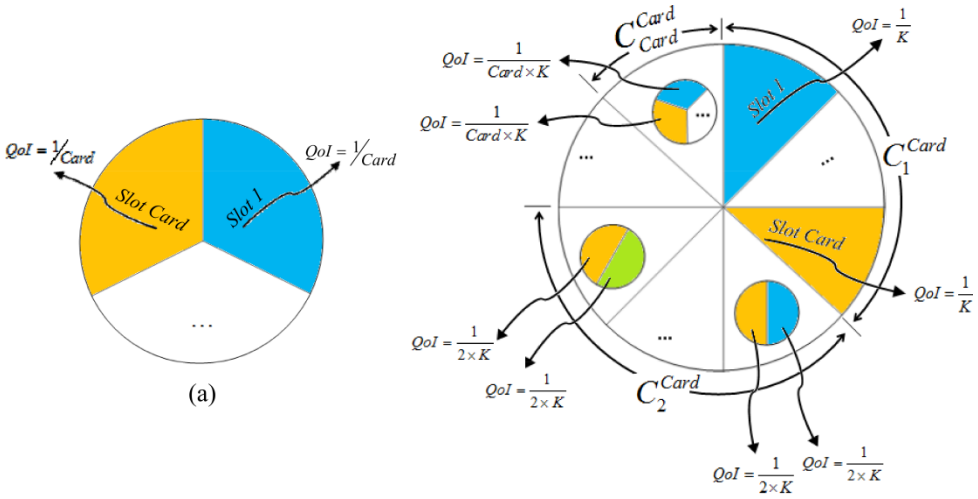


Figure 1. *QoI's* values for the abducible set for $predicate_i$ with (a) and without (b) constraints on the possible combinations among the abducible clauses.

However, a term's *QoI* also depends on their attribute's *QoI*. In order to evaluate this metric, look to Figure 2, where the segment with bounds 0 and 1 stands for every attribute domain, i.e., all the attributes range in the interval [0, 1]. [*A*, *B*] denotes the range where the unknown attributes values for a given predicate may occur (Figure 2):

Figure 2. Setting the *QoIs* of each attribute's clause.

$$QoI_{attribute_i} = 1 - \|A - B\| \quad (3)$$

where ||A–B|| stands for the modulus of the arithmetic difference between *A* and *B*. Therefore, one may have (Figure 3):

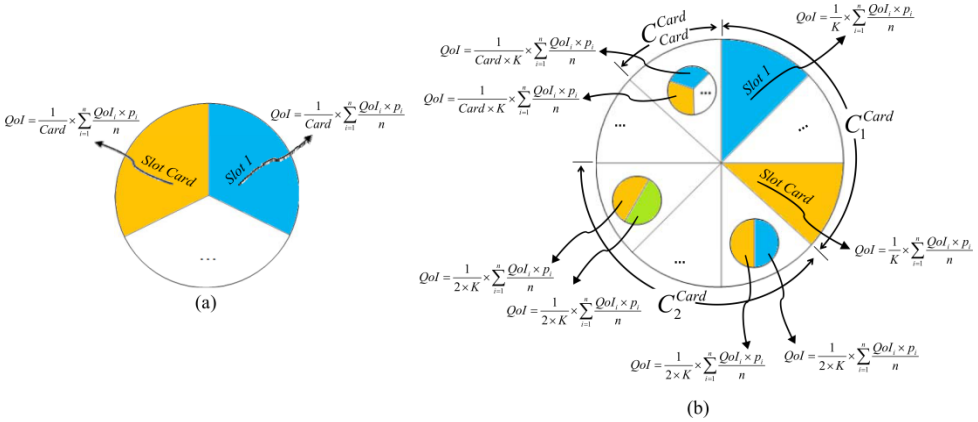


Figure 3. *QoI's* values for the abducible set for predicatei with (a) and with (b) constraints on the possible combinations among the abducible clauses. $\sum_{i=1}^{n} (QoI_i \times p_i)/n$ denotes the *QoI's* average of the attributes of each clause (or term) that sets the extension of the predicate under analysis. n and pi stand for, respectively, for the attribute's cardinality and the relative weight of attribute pi with respect to its peers ($\sum_{i=1}^{n} p_i = 1$).

Under this setting, another metric has to be considered, which will be denoted as *DoC* (*Degree-of-Confidence*), that stands for one's confidence that the argument values or attributes of the terms that make the extension of a given predicate, having into consideration their domains, are in a given interval [14]. The *DoC* is figured using $DoC = \sqrt{1 - \Delta l^2}$, where Δl stands for the argument interval length, which was set to the interval [0, 1] (Figure 4).

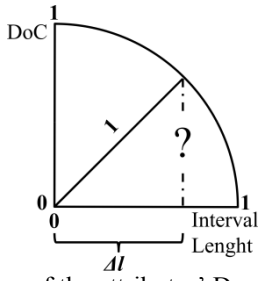


Figure 4. Evaluation of the attributes' Degree of Confidence.

Thus, the universe of discourse is engendered according to the information presented in the extensions of such predicates, according to productions of the type:

$$predicate_i - \bigcup_{1 \le j \le m} clause_j \left(\left(\left(A_{x_1}, B_{x_1} \right) \left(QoI_{x_1}, DoC_{x_1} \right) \right), \cdots, \left(\left(A_{x_l}, B_{x_l} \right) \left(QoI_{x_l}, DoC_{x_l} \right) \right) \right) :: QoI_j :: DoC_j \quad (4)$$

where $\bigcup$, *m* and *l* stand, respectively, for *set union*, the *cardinality* of the extension of *predicate*$_i$ and the number of attributes of each clause [14]. On the other hand, either the subscripts of the QoI_s and the DoC_s, or those of the pairs (A_s, B_s), i.e., x_1, ..., x_l, stand for the attributes' values ranges.

2.2 Qualitative Knowledge

In present study both qualitative and quantitative data/knowledge are present. Aiming at the quantification of the qualitative part and in order to make easy the understanding of the process, it will be presented in a graphical form. Taking as an example a set of *n* issues regarding a particular subject (where there are *k* possible choices, i.e., *none*, *low*, ..., *high* and *very high*), let us itemized an unitary area circle split into *n* slices (Figure 5). The marks in the axis correspond to each of the possible options. If the answer to issue 1 is *high* the correspondent area is $\pi \times \left(\sqrt{\frac{k-1}{k \times \pi}} \right)^2 / n$, i.e., $(k-1)/(k \times n)$ (Figure 5(a)). Assuming that in the issue 2 are chosen the alternatives *high* and *very high*, the correspondent area ranges between $\left[\pi \times \left(\sqrt{\frac{k-1}{k \times \pi}} \right)^2 / n,\ \pi \times \left(\sqrt{\frac{k}{k \times \pi}} \right)^2 / n \right]$, i.e., $\left[(k-1)/(k \times n),\ k/(k \times n) \right]$ (Figure 5(b)). Finally, in issue *n* if no alternative is ticked, all the hypotheses should be considered and the area varies in the interval $\left[0,\ \pi \times \left(\sqrt{\frac{k}{k \times \pi}} \right)^2 / n \right]$, i.e., $\left[0,\ k/(k \times n) \right]$ (Figure 5(c)). Thus, the total area is the sum of the partial ones (Figure 5(d)), i.e., $\left[(2k-2)/(k \times n),\ (3k-1)/(k \times n) \right]$.

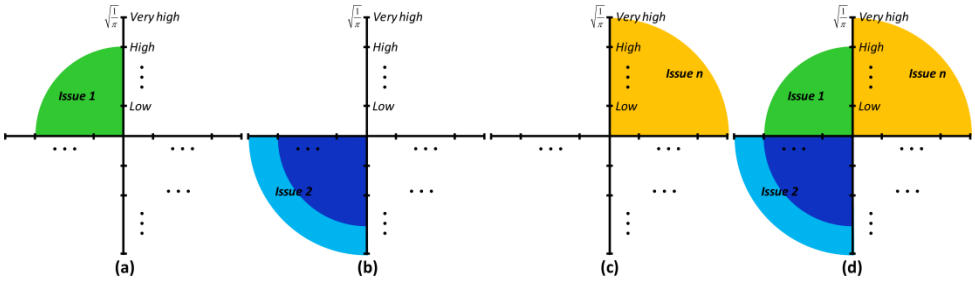


Figure 5. A view of a qualitative data/knowledge processing.

3 Implementation and considerations

As described above, the present work aims at taking advantage of the data stored in BrainArchive as the basis for a reasoning system capable of suggesting diagnoses. This system was developed based on pre-existing neuroimaging data processing workflows for structural, functional and diffusion MRI data. Nevertheless, it is also flexible enough to store data from several different studies and consequently hosting data from subjects with a variety of physiological conditions, such as aging, stress and pathophysiological conditions such as Alzheimer's Disease (AD), Major Depression (MD) or Obsessive Compulsive Disorder (OCD). For practical reasons, in the present work, we will focus on data from an OCD study containing information from 106 subjects. The data processing details and biomarkers extracted will be further detailed below.

3.1 Dataset

The original dataset used for this study was composed by the demographic and clinical characterization, structural, resting-state function, task-related function and diffusion MRI images from 106 subjects. Considering the amount of participants and the amount of attributes that can be extracted, it may be said that it stands for a rich one and appropriate for the addressed problem. Specifically, it encompasses qualitative data such as the one provided by the clinical observations and the quality of that data, quantitative data such as the one provided by the MRI data or neuropsychological tests and missing data as some participants were unable to perform some tests/acquisitions or the image quality was not suitable for analysis.

In particular, in the OCD case, there is no clear neuroimaging biomarker of the pathology. Nevertheless, several different multimodal biomarkers, with possibly different discrimination value, have been suggested. For the present work, a combination of biomarkers of structural gray matter volume derived from structural MRI images, functional connectivity from resting-state fMRI images and white matter fractional anisotropy (FA) extracted from diffusion MRI images is used. Based on a recent meta-analysis [21], we will consider the Gray Matter (GM) volume of bilateral putamen, left post central gyrus, right temporal and right insula volumes. Based on the same meta-analysis, FA in the superior longitudinal fasciculus, body of the corpus callosum and forceps minor were also extracted. Finally functional connectivity between the putamen and the orbitofrontal cortex, part of the cortical-striatal-thalamic-cortical (CSTC) know to be involved in the pathophysiology of OCD was also used [22].

3.2 Processing MRI Data

In order to obtain interpretable results from MRI data it is necessary to perform several data pre-processing tasks.

Concerning structural data, regional volumes were obtained through Voxel Based Morphometry (VBM) performed using SPM8[1]. Briefly, the structural images were segmented into different tissue classes using the New Segmentation procedure. Then, a study specific template was built using the Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL), the images were registered to the Montreal Neurological Institute (MNI) standard space, scaled with the Jacobian determinant and smoothed with 10 mm full-width at half-maximum (FWHM) Gaussian kernel. Regional relative GM volumes were then extracted for each of the regions specified above.

[1] http://www.fil.ion.ucl.ac.uk/spm/

Regarding the diffusion data, all the data processing procedures were performed with the FMRIB Software Library (FSL v5.0[2]). Eddy-current distortions and motion correction were initially performed. Afterwards, the Diffusion Tensor Imaging (DTI) model was fitted to the data and FA maps were calculated. Finally, FA images were co-registered and normalized to MNI standard space through the Tract-Based Spatial Statistics (TBSS) procedures. Similarly to the volumetric data, regional FA values were extracted for the referred tracts of interest.

The functional data was also pre-processed. Pre-processing included slice timing correction, motion correction and scrubbing, normalization to MNI standard space, regression of motion parameters, motion outliers and WM and CSF confounding signals, smoothing with 8 mm FWHM Gaussian filter and band-pass temporal filtering (0.01-0.08 Hz). Functional connectivity between pairs of regions was set as the Pearson correlation between the regional time-series.

The above mentioned data is only part of the large volume of information that are organized and stored in BrainArchive and can be used for the development of a reasoning system.

It is now possible to build up a knowledge database given in terms of the extensions of the relations (or tables) depicted in Figure 7, which denote a situation where one has to manage information in order to evaluate the OCD risk. Indeed, the *Structural Gray Matter Volumes*, *Functional Connectivity* and *White Matter Fractional Anisotropy* tables are populated with the responses to the issues presented in the dataset, where some incomplete, default and/or unknown data are present. For instance, in the former case the *Functional Connectivity* is unknown (depicted by the symbol $\perp$). The values for all tables are between 0 and 1. In order to quantify the information present in these tables the procedures already described above were followed.

[2] http://fsl.fmrib.ox.ac.uk/fsl/

Structural Gray Matter Volumes				
#	Bilateral Putamen	Left Post Central Gyrus	Right Temporal	Right Insula
1	0.008	0.012	[0.016 , 0.018]	0.010
2	⊥	[0.010 , 0.013]	0.016	0.013
...	...	...	...	...
106	0.010	0.015	0.017	[0.010 , 0.013]

Functional Connectivity	
#	Putamen and Orbitofrontal Cortex
1	⊥
2	0.50
...	...
106	[0.30 , 0.40]

White Matter Fractional Anisotropy			
#	Superior Longitudinal Fasciculus	Body of the Corpus Callosum	Forceps minor
1	[0.65 , 0.67]	0.82	0.70
2	0.70	0.85	0.60
...	...	...	...
106	0.85	[0.85 , 0.90]	0.55

Demographic and Clinical Attributes			
#	Gender	Age	Frequency of symptoms
1	M	45	Frequent
2	F	55	Very Frequent
...	...	...	...
106	M	62	Frequent

Figure 7. A fragment of the knowledge base for *OCD Risk Assessment*.

Applying the algorithm presented in [16] to the table or relation's fields that make the knowledge base for the OCD Risk Assessment (Figure 7), and looking to the DoCs values obtained as described in [16], it is possible to set the arguments of the predicate ***ocd risk assessment*** (*ora*) referred to below, whose extensions denote the objective function with respect to the problem under analysis:

$$ora: S_{tructural}G_{ray}M_{atter}V_{olumes},\ F_{unctional}C_{onnectivity},W_{hite}M_{atter}F_{ractional}A_{nisotropy} \rightarrow \{0,1\}$$

where 0 (zero) and 1 (one) denote, respectively, the truth values *false* and *true*. The algorithm presented in [16] encompasses different phases. In the former one the clauses or terms that make extension of the predicate under study are established. In the subsequent stage the arguments of each clause are set as continuous intervals. In a third step the boundaries of the attributes intervals are set in the interval [0, 1] according to a normalization process given by the expression $(Y - Y_{min})/(Y_{max} - Y_{min})$, where the Y_s stand for themselves. Finally, the DoC is evaluated as described in section 2.

3.3 Case-Based Reasoning

Case-based reasoning (CBR) is an Artificial Intelligence technique that solves problems based on similar past cases. A new problem is solved by finding a previous similar situation and by reusing and adapting knowledge of that situation [18]. CBR can be described by the CBR-cycle (Figure 8). This cycle is constituted by four main steps [26]:

1. Retrieve similar cases to the problem description
2. Reuse a solution suggested by a similar case
3. Revise or adapt that solution to better fit the new problem if necessary
4. Retain the new solution once it has been confirmed or validated

Another feature of CBR is that every time a new problem has been solved, the solution for that problem is retained, making it available for future problems. Thus, CBR knowledge is constantly updated, which allows the use of incomplete, unknown, or even, self-contradictory data.

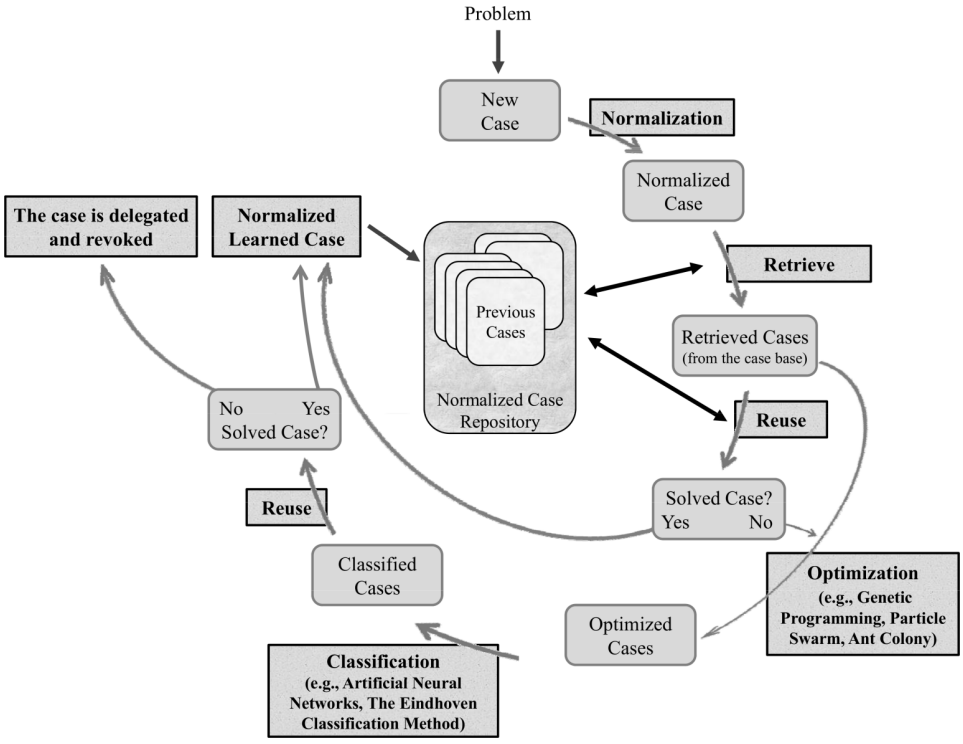


Figure 8. The CBR Cycle [18].

A soft computing approach to model the universe of discourse based on a CBR methodology for problem solving is now set. Contrasting with other problem solving methodologies (e.g., Decision Trees or ANNs), in a CBR based methodology for problem solving relatively little work is done offline [23]. Undeniably, in almost all the situations the work is performed at query time. The main difference between this new approach and the typical CBR one relies on the fact that not only all the cases have their arguments set in the interval [0, 1], but it also caters for the handling of incomplete, unknown, or even self-contradictory data or knowledge. Thus, the classic CBR cycle was changed (Fig. 1), being the Case Base given in terms of triples that follow the pattern:

$$\text{Case} = \{< \text{Raw}_{\text{data}}, \text{Normalized}_{\text{data}}, \text{Description}_{\text{data}} >\}$$

where Raw_{data} and $\text{Normalized}_{\text{case}}$ stand for themselves, and $\text{Description}_{\text{data}}$ is made on a set of strings or even in free text, which are analysed with string similarity algorithms, namely the Dice coefficient [24]. When confronted with a new case, the system is able to retrieve all cases that meet such a structure and optimize such a population, i.e., it considers the attributes DoC's value of each case or of their optimized counterparts when analysing

similarities among them. Thus, under the occurrence of a new case, the goal is to find similar cases in the CaseBase. Then, the algorithm given in [16] is applied to a new case that presents the feature vector :

$$S_{tructural}G_{ray}M_{atter}V_{olumes}: 0.65,\ F_{unctional}C_{onnectivity}: 0.40, W_{hite}M_{atter}F_{ractional}A_{nisotropy}: 0.80$$

having in consideration that the cases retrieved from the Case Base must satisfy the invariant:

$$\bigcap_{i=1}^{n}\left(\mathrm{B_i},\ \mathrm{E_i}\right) \neq \emptyset \qquad (6)$$

which denotes that the intersection of the attributes range in the cases that make the Case Base repository (Bi), and the equals of the new case (E*i*), cannot be empty. Then, the computational process may be continued, and one may have:

$$\underbrace{\mathrm{ocd}_{\mathrm{risk}_{\mathrm{new}}}\left((1,1),(1,0.99),(1,0.96),(1,1),(1,0.99),(1,1),(1,1)\right)}_{\text{new case}} :: 1 :: 0.99$$

Then, the new case can be depicted on the Cartesian plane in terms of its QoI and DoC, and by using k-means clustering methods [25], it is feasible to identify the clusters that intermingle with the new one. The new case is compared with every retrieved case from the cluster using a similarity function sim, given in terms of the average of the modulus of the arithmetic difference between the arguments of each case of the selected cluster and those of the new case. Thus, one may have:

$$\mathrm{ocd}_{\mathrm{risk}_1}\left((1,1),(1,0),(1,0.96),(1,1),(1,1),(1,1),(1,0.96)\right) :: 1 :: 0.84$$

$$\mathrm{ocd}_{\mathrm{risk}_2}\left((1,1),(1,0.99),(1,0.83),(1,1),(1,1),(1,0.97),(1,1)\right) :: 1 :: 0.97$$

$$\vdots$$

$$\underbrace{\mathrm{ocd}_{\mathrm{risk}_j}\left((1,1),(1,1),(1,0.96),(1,1),(1,1),(1,0),(1,1)\right) :: 1 :: 0.85}_{\text{normalized cases from retrieved cluster}}$$

Assuming that every attribute has equal weight, the dissimilarity, in terms of DoC, between $\mathrm{ocd}_{\mathrm{risk}_{\mathrm{new}}}^{\mathrm{DoC}}$ and $\mathrm{ocd}_{\mathrm{risk}_1}^{\mathrm{DoC}}$, i.e., $\mathrm{ocd}_{\mathrm{risk}_{\mathrm{new}\to 1}}^{\mathrm{DoC}}$, may be computed as follows:

$$\mathrm{ocd}_{\mathrm{risk}_{\mathrm{new}\to 1}}^{\mathrm{DoC}} = \frac{\|1-1\| + \|0.99-0\| + \cdots + \|1-1\| + \|1-0.96\|}{7} = 0.15$$

Thus, the similarity for $\mathrm{thromb}_{\mathrm{risk}_{\mathrm{new}\to 1}}^{\mathrm{DoC}}$ is 1 – 0.15 = 0.85. Regarding QoI the procedure is similar, returning $\mathrm{ocd}_{\mathrm{risk}_{\mathrm{new}\to 1}}^{\mathrm{QoI}} = 1$. Thus, one may have:

$$\mathrm{ocd}_{\mathrm{risk}_{\mathrm{new}\to 1}}^{\mathrm{QoI,\ DoC}} = 1 * 0.85 = 0.85$$

These procedures should be applied to the remaining cases of the retrieved cluster in order to obtain the most similar ones, which may stand for possible solutions to the problem. This approach allows physicians to define the most appropriate similarity threshold to address the problem (i.e., it gives the user the possibility to narrow the number of selected cases with the increase of the similarity threshold). Description allows physicians to assess patients' information such as previous therapy and their actual clinical status. Furthermore, the proposed system displays to physician clinical data of the target patient and their relatives in order to signalizing thrombotic events and making appropriate recommendations. It must be also stated that parameter settings, detail experimental data descriptions or implementations of the different modules, as well as issues like correctness, appropriateness and repeatability of the experimental results were already treated in others publications, namely in [19], [20].

4 Conclusions

This work presents an Artificial Intelligence based Decision Support System to Multimodal Neuroimaging Studies centred on a formal framework based on Logic Programming for Knowledge Representation and Reasoning, complemented with a Case-based approach to computing, that caters for the handling of incomplete, unknown or even self-contradictory information. It may set the basis for an overall approach to such systems in the future, in this and other domains. Indeed, it has the potential to be disseminated across others prospective areas, therefore validating a universal attitude. Indeed, under this line of thinking the cases' retrieval and optimization phases were heightened when compared with existing tactics or methods. Additionally, under this approach the users may define the cases weights attributes on-the-fly, letting them to choose the appropriate strategies to address the problem (i.e., it gives the user the possibility to narrow the search space for similar cases at runtime). Indeed, each

modality of brain imaging has its own benefits, and the combination of the various techniques, each one with their own biomarkers, allowed medical imaging to step out a minority role to a central role in terms of diagnose. Indeed, reasoning systems can be used to extract information from available knowledge. In a system like BrainArchive, knowledge can be extracted from the multimodal brain data, and be used to deduce more information and possible conclusions, instead of simply storing it, i.e., it provides tools for the identification of undiagnosed diseases, and may be added to neuroscience BrainArchive studies.

Future treatments will have greater opportunity to succeed, introduced in early stages of the disease and having the medical image to evaluate their benefits.

Acknowledgements

This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013.

References

[1] F. Fernandes, P. Marques, R. Magalhães, N. Sousa, and V. Alves, "Enabling data storage and availability of multimodal neuroimaging studies – A NoSQL based solution."

[2] S. Goldman, L. M. O'Brien, P. A. Filipek, I. Rapin, and M. R. Herbert, "Motor stereotypies and volumetric brain alterations in children with Autistic Disorder," *Res. Autism Spectr. Disord.*, vol. 7, no. 1, pp. 82–92, 2013.

[3] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal Classification of Alzheimer's Disease and Mild Cognitive Impairment," *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011.

[4] J. Radua, S. Borgwardt, A. Crescini, D. Mataix-Cols, A. Meyer-Lindenberg, P. K. McGuire, and P. Fusar-Poli, "Multimodal meta-analysis of structural and functional brain changes in first episode psychosis and the effects of antipsychotic medication," *Neurosci. Biobehav. Rev.*, vol. 36, no. 10, pp. 2325–2333, 2012.

[5] A. T. Reid, D. Bzdok, S. Genon, R. Langner, V. I. Müller, C. R. Eickhoff, F. Hoffstaedter, E. C. Cieslik, P. T. Fox, A. R. Laird, K. Amunts, S. Caspers, and S. B. Eickhoff, "ANIMA: A data-sharing initiative for neuroimaging meta-analyses," *Neuroimage*, vol. 124, pp. 1245–1253, 2016.

[6] G. A. Book, M. C. Stevens, M. Assaf, D. C. Glahn, and G. D. Pearlson, "Neuroimaging data sharing on the neuroinformatics database platform," *Neuroimage*, vol. 124, pp. 1089–1092, 2016.

[7] M. R. Hodge, W. Horton, T. Brown, R. Herrick, T. Olsen, M. E. Hileman, M. McKay, K. A. Archie, E. Cler, M. P. Harms, G. C. Burgess, M. F. Glasser, J. S. Elam, S. W. Curtiss, D. M. Barch, R. Oostenveld, L. J. Larson-Prior, K. Ugurbil, D. C. Van Essen, and D. S. Marcus, "ConnectomeDB-Sharing human brain connectivity data," *Neuroimage*, vol. 124, pp. 1102–1107, 2016.

[8] D. S. Marcus, T. R. Olsen, M. Ramaratnam, and R. L. Buckner, "The Extensible Neuroimaging Archive Toolkit," *Neuroinformatics*, vol. 5, no. 1, pp. 11–34, 2007.

[9] V. Lifschitz, "On the declarative semantics of logic programs with negation," in *Foundations of deductive databases and logic programming*, J. Minker, Ed. Morgan Kaufmann Publishers Inc., 1988, pp. 177–192.

[10] A. C. Kakas, R. A. Kowalski, and F. Toni, "Abductive logic programming," *J. Log. Comput.*, vol. 2, no. 6, pp. 719–770, 1992.

[11] L. M. Pereira and H. T. Anh, "Evolution prospection," *Stud. Comput. Intell.*, vol. 199, pp. 51–63, 2009.

[12] J. F. M. Neves, "A logic interpreter to handle time and negation in logic data bases," *Proc. 1984 Annu. Conf. ACM fifth Gener. Chall.*, 1984.

[13] J. Neves, J. Machado, C. Analide, A. Abelha, and L. Brito, "The Halt Condition in Genetic Programming," *Prog. Artif. Intell. 13th Port. Conf. Aritficial Intell. {EPIA} 2007, Work. {GAIW}, {AIASTS}, {ALEA}, {AMITA}, {BAOSW}*, vol. 4874, pp. 160–169, 2007.

[14] P. Lucas, "Quality Checking of Medical Guidelines through Logical Abduction," *Inf. Sci. (Ny).*, vol. 2003, pp. 1–11.

[15] J. Machado, A. Abelha, P. Novais, J. Neves, and J. Neves, "Quality of service in healthcare units," *Ijcaet*, vol. 2, no. 4, pp. 436–449, 2010.

[16] F. Fernandes, H. Vicente, A. Abelha, J. Machado, P. Novais, and J. Neves, "Artificial Neural Networks in Diabetes Control," pp. 362–370, 2015.

[17] M. Figueiredo, J. Neves, G. Gomes, and H. Vicente, "Assessing the Role of General Chemistry Learning in Higher Education," *Procedia - Soc. Behav. Sci.*, vol. 228, no. June, pp. 161–168, 2016.

[18] D. Martins, H. Vicente, V. Alves, and J. Neves, "Age Prediction through Pelvis X-Ray Images: A Case

Based approach to Problem Solving," 2016.
[19] M. Figueiredo, J. Neves, and H. Vicente, "A Soft Computing Approach to Quality Evaluation of General Chemistry Learning in Higher Education.," *Methodol. Intell. Syst. Technol. Enhanc. Learn. Adv. Intell. Soft Comput.*, vol. 478, pp. 81–89, 2016.
[20] C. Coelho, M. R. Martins, N. Lima, H. Vicente, and J. Neves, "An Assessment to Toxicological Risk of Pesticide Exposure," *Build. Sustain. Heal. Ecosyst. Commun. Comput. Inf. Sci.*, vol. 636, 2016.
[21] G. K. Eng, K. Sim, and S. H. A. Chen, "Meta-analytic investigations of structural grey matter, executive domain-related functional activations, and white matter diffusivity in obsessive compulsive disorder: An integrative review," *Neurosci. Biobehav. Rev.*, vol. 52, pp. 233–257, 2015.
[22] J. C. Beucke, J. Sepulcre, T. Talukdar, C. Linnman, K. Zschenderlein, T. Endrass, C. Kaufmann, and N. Kathmann, "Abnormally High Degree Connectivity of the Orbitofrontal Cortex in Obsessive-Compulsive Disorder.," *JAMA psychiatry*, vol. 70, no. 6, pp. 619–29, 2013.
[23] D. Carneiro, P. Novais, F. Andrade, and J. Zeleznikow, "Using Case Based Reasoning to Suppor t Alter native Dispute Resolution," vol. 79, no. Dcai, 2010.
[24] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," vol. 26, no. 3, pp. 297–302, 1945.
[25] M. Figueiredo, L. Esteves, J. Neves, and H. Vicente, "A data mining approach to study the impact of the methodology followed in chemistry lab classes on the weight attributed by the students to the lab work on learning and motivation," *Chem. Educ. Res. Pr.*, 2016.
[26] I. Watson, "Case-based reasoning is a methodology not a technology," *Knowledge-Based Syst.*, vol. 12, no. 5–6, pp. 303–308, 1999.