Engineering Applications of Artificial Intelligence 55 (2016) 165–185

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Understanding effects of cognitive rehabilitation under a knowledge discovery approach

Alejandro García-Rudolph[a], Karina Gibert[b,*]

[a] Institut Guttmann – Hospital de Neurorehabilitació, Camí de Can Ruti s/n 08916, Badalona, Barcelona Spain
[b] Dep. of Statistics and Operations Research, Universitat Politècnica de Catalunya – BarcelonaTech, Jordi Girona 1-3, 08034 Barcelona, Spain

A R T I C L E I N F O

Article history:
Received 17 January 2016
Received in revised form
29 April 2016
Accepted 16 June 2016

Keywords:
Cognitive rehabilitation
Traumatic brain injury
Knowledge discovery
Motifs
Clustering-based on rules
Post-processing

A B S T R A C T

Traumatic brain injury (TBI) is the leading cause of death and disability in children and young adults worldwide. Cognitive rehabilitation (CR) plans consist of a sequence of CR tasks targeting main cognitive functions. There is not enough on-field experience yet regarding which specific intervention (tasks or exercise assignment) is more appropriate to help therapists to design plans with significant effectiveness on patient improvement. The selection of specific tasks to be prescribed to the patient and the order in which they might be executed is currently decided by the therapists based on their experience.

In this paper a new data mining methodology is proposed, combining several tools from Artificial Intelligence, clustering and post-processing analysis to identify regularities in the sequences of tasks in such a way that treatment profiles (classes) can be discovered. Due to the cumulative effect of rehabilitation tasks, small variations within the sequence of tasks performed by the patient do not significantly change the final outcomes in rehabilitation and makes it difficult to find discriminant rules by using the traditional machine learning inductive methods. However, by relaxing the formalization of the problem to find patterns that might include small variations, and introducing motif discovery techniques in the proposed methodology, the complexity of the neurorehabilitation phenomenon can be better captured and a global structure of successful treatment task sequences can be devised.

Following this, the relationship between the discovered patterns and the CR treatment response are analyzed, offering a richer perspective than that provided by the single task focus traditionally used in the CR field.

The paper provides a definition of the whole methodological approach proposed from a formal point of view, and its application to a real dataset. Comparisons with traditional AI approaches are also presented and the contribution of the proposed methodology to the AI field discussed.

1. Introduction

According to the World Health Organization (WHO), traumatic brain injury (TBI) is the leading cause of death and disability in children and young adults worldwide and it is involved in nearly half of all trauma deaths (Schroeter et al., 2011). In Europe, brain injuries from trauma are responsible for more years of disability than any other cause (Maas et al., 2009). The incidence is increasing in lower income countries; the WHO predicts that TBI and road traffic accidents will be the third cause of disease and injury worldwide by 2020. Cognitive impairments due to TBI are substantial sources of morbidity for affected individuals, their family members, and society. Disturbances of attention, memory, and executive functioning are the most common consequences of TBI.

Neurorehabilitation is the process that exploits the cerebral plasticity to reduce brain deficit. Cognitive rehabilitation (CR) aims to reduce the impact of disabling conditions and tries to improve the cognitive deficits caused by TBI. From Luria's theory back in 1978, there is a common believe that direct retraining of damaged cognitive processes through repeated stimulation and activation of the targeted brain areas can help patient recovery. For maximum activation to occur, the patient must face tasks *just barely too difficult* for him (Green and Bavelier, 2005). Designing a CR treatment for a given patient therefore means determining the correct sequence of CR tasks to be asked of the patient in a quite precise trade-off between *enough stimulating* and *sufficiently achievable* tasks, which is far from intuition, and still is both an empirical and theoretical open problem in the area. It has been seen that similar patients respond differently to similar CR treatments. Literature reports single task approaches to this purpose, analysing the

* Corresponding author.
E-mail addresses: agarciar@guttmann.com (A. García-Rudolph), karina.gibert@upc.edu (K. Gibert).

http://dx.doi.org/10.1016/j.engappai.2016.06.007
0952-1976/

associations between the performance of a certain task and the response to the CR treatment. However, although there is some empirical knowledge, traditional approaches do not seem to provide sufficient scientific evidence about the factors determining a favourable outcome, and there is still a limited scientific base to support the effectiveness of CR (Cicerone et al., 2011). In fact, most of the work found in this field adopts traditional pre-post analysis with intervention studies where a specific treatment is applied to a sample of patients and contrasted with a control group. But none of this analysis includes detailed characteristics of different CR programs in the model. They only use assessment of the patient before and after the treatment and characteristics of the lesion and of the patient to predict improvement. These works, although useful to prove the effectiveness of CR, do not contribute to a better design of CR programs for a specific patient.

In this work, the underlying structure of the CR phenomenon has been analyzed in depth and it has been seen that the CR field has some specific characteristics that make a successful application of traditional methods difficult:

- Patients following a CR program are not performing a single task, neither a single type of task, but a certain complex combination of them that are likely to be interrelated or synergistics. A single task approach cannot take into account the complex interactions among tasks.
- Cognitive tasks, even when specifically designed to target a particular cognitive function, might also have side-effects on other cognitive functions (Cicerone et al., 2011). This makes it difficult to examine the isolated effect of a single task in a specific cognitive function, and no clear evidence appear when all tasks are integrated into a traditional model.
- The additional effect of a single task might be affected by the cumulative effect of the sequence of previous tasks executed under the treatment, and this might determine that the order of execution is relevant in the treatment.
- The effect of a single task may be too subtle to be detected, whereas the effects of the whole CR treatment may be sufficient to be detectable, taking into account the cumulative effect of rehabilitation already mentioned.

From a structural point of view, these characteristics resemble those in nutritional epidemiology, where a global approach has been adopted in recent years, and all nutrients are analyzed together due to the high degree of interaction (Hu, 2002). This points to analyzing the overall CR treatment, by considering all kinds of interactions among tasks together, instead of using the traditional single task approach. Therefore, CR treatment in this work will be considered as a sequence of cognitive tasks and data mining methods will be used to determine the multivariate associations between a CR treatment (or relevant subsequences) and the degree of response of the patient, under this new perspective. Analyzing CR tasks as *treatment patterns* offers an innovative perspective in neurorehabilitation, and describing their relationship with their clinical outcome provides a practical approach to evaluate the effects of rehabilitation treatments. It can also enhance our conceptual understanding of CR treatments practice, and might be useful to provide guidance for cognitive treatment interventions.

To this end, an innovative data mining methodology is proposed to analyze the underlying structure of CR processes and to better determine the most suitable CR treatment for a given patient. In a first approach, one would be tempted to identify the task sequences associated with the improvement of different patients, by using a traditional classifier to find patterns associated with the different treatment responses found in different patients. However, this approach has shown serious limitations in this field, providing extremely weak patterns that do not really help in clinical practice, as will be discussed later.

In this particular field, because of the cumulative effect of the tasks mentioned above, it is reasonable to think that the effect of a certain sequence of CR tasks can behave robustly to slight variations of the sequence. Thus, small variations of the sequence of tasks performed might keep the global effect of the treatment unaltered. This means that the model to be built should admit a certain level of variation around every relevant pattern. However, working with a task's global profile for treatments is neither useful, as the order of executions becomes relevant in CR. These characteristics have already been encountered in the bioinformatics fields, particularly in *transcription factor binding sites* (TFBS) field, where slightly different sequences of DNA are associated with a certain biological function. *Motif discovery or motif finding* methods are used in this field, to represent those weak patterns. Similarly, motif discovery methods will be introduced in our proposed methodology to identify patterns of CR treatments, where slight variations in the treatment program might be packed into a single CR motif with a similar therapeutic effect, and might be associated with a certain response level.

This paper introduces the new Sequence of Activities Improving Multi-Area Performance (SAIMAP) methodology, as an innovative combination of pre-processing tools, clustering, AI methods, motif discovery and post-processing techniques. SAIMAP is a hybrid methodological frame where useful patterns can be found from data. SAIMAP works for domains with high order interactions among variables and sequential information along time that involves cummulative effects. This provides a complex structure, for which most of the classical data mining approaches do not perform very well. The proposed methodology is general for problems with the structure described before, although in the paper it is applied to the particular field of finding design guidelines for CR treatments. SAIMAP first finds groups of similar treatments, then makes a local characterization of each group by using motif discovery methods, and finally analyzes the relationships between those typical treatments and the evaluation of patients' improvement after treatment. Statistical tests and multiple boxplots are used to relate the discovered groups with patient' characteristics, level of impairment and associated with specific treatment patterns.

The structure of the paper is the following: Section 2 introduces the state of the art, organized in sections related to the different research areas involved in this multidisciplinar research, from both the application and methodological point of view. In Section 3 methodological issues are provided: first, the formalization of the general problem addressed is defined, with a clear presentation of the structural components of the problem (our methodological proposal addresses scenarios where individuals perform sequences of predefined set of activities with high order interactions among them and cumulative effects). The main goal is to identify a reduced set of characteristic sequences of activities profiling groups of individuals who behave similarly. The second part of Section 3 introduces the SAIMAP methodology as our proposal to address this problem. SAIMAP is composed by 13 formal steps, where sequential patterns are induced from data fitting the structure defined in Section 3.1. In Section 4 our methodology is applied to a specific real case, regarding cognitive rehabilitation treatments of traumatic brain injury patients; the inputs are sequences of cognitive rehabilitation tasks performed by the patients along the CR treatment; preprocessing activities are detailed and sequential patterns of CR tasks are obtained following the SAIMAP steps. The discovered patterns, are interpreted through motif discovery tools and associated with several criteria measuring improvement in a predefined set of impact areas that might be targeted in parallel by a single task (in the particular case