Case-Based Reasoning for Pediatrics Developmental Disorders

eiTan LaVi[a], Yuval Shahar[b], Mitchell Schertz[c], Shmuel Einav[a]

[a] *Biomedical Engineering Department, Tel-Aviv University, Israel*
[b] *Medical Informatics Research Center, Department of Information Systems Engineering, Ben Gurion University, Beer Sheva, Israel*
[c] *Institute for Child Development ,Kupat Holim Meuhedet, Central Region, Herzeliya, Israel*

Abstract and Objective

While 1 in 5 children suffers from Pediatrics Developmental disorders [1] (PDD), PDD specialists are operating in IT darkness; their domain is ontologically and taxonomically ill-defined and no prototypical cases or clinical decision making models are available to assist them [2]. PDD is pressing for Medical Informatics solutions, yet no prior research has been dedicated to developing Decision-Support Systems (DSS) or methodologies for PDD [2, 3]. This ongoing research is being conducted on a real-life case-base from a major PDD clinic in Israel[b], with the primary objective of constructing a robust Case-Based-Reasoning (CBR) agent to ***perform CBR retrieval to predict diagnoses of new cases****. Heretofore we have implemented and tested the retrieval and reuse modules of the CBR cycle. Ultimately, the significance of such work is in helping to lay the foundations for a CBR system which allows clinicians to objectively utilize the whole of their collective past experience in order to individually produce better decisions in diagnosis and treatment of developmental disorders.*

Keywords:

Development-disorders, CBR, LSA, LSI, DSS, Clustering-based feature weighing, Case-based retrieval

Introduction

PDD refer to any delay in development based on that expected for a given age level or stage of development. These disabilities constitute a substantial impairment and are caused by a myriad of biological and non-biological factors [4]. Currently, PDD has no formal clinical model for decision making, no clear or widely approved taxonomy, no agreed upon names for the different pathological states and not even an agreement as to what exactly are the different pathological states that exist [2]. There is, therefore, also no agreement regarding treatment methods for the various syndromes [2]. Having no robust clinical model to aid him, a PDD specialist relies mainly on his own (available memory of) past experience, signifying the need for a CBR decision-support methodology in this domain.

Methods

We are working with a case base consisting of 8022 cases, each containing 182 attributes, about 80 of which are free-text fields which encompass the most clinically significant information. The case base is extremely sparse and noisy; the free-text fields are written by several physicians, often alternating between Hebrew and English and writing with multiple abbreviations and spelling errors. Further, the prediction task is difficult as each case is assigned a combination of diagnoses (3 in average) from a list of 235 possible morbidities. Heretofore our system has been developed in the **R software environment for statistical computing and graphics**. The developed algorithms are constructed as follows: **(a)** an **import & preprocessing module** which automatically classifies a type for each attribute and prepares the case base for extensive KDDM processing; **(b)** a **transformation module** to synthesize new and meaningful attributes from the existing ones; **(c)** a **generic multilingual text-mining module** to construct multilingual dictionaries and log-entropy weighted text-document matrices for each textual attribute; **(d)** a **generic CBR retrieval preprocessor module** to construct similarity matrices for all attribute types, including novel NA handling as well as methodologies for robust and generic LSI-based construction of similarity matrices. **(e)** A novel **CBR retrieval & prediction module** utilizing novel methodologies which integrate several weighing schemes, scoring schemes and various K values (the number of "nearest cases" to retrieve) in performing retrieval. The retrieved cases' diagnoses are then processed through novel reuse methodologies to obtain the final prediction probability scores for all diagnoses for each test case. Future work will use the LSI-based similarity matrices to perform clustering of cases for each free-text attribute, and draw on the quality of clustering (in terms of diagnoses homogeneity) to assert the clinical weight of each attribute.

Results & Conclusions

Our retrieval-based predictions were tested using a Leave One Out methodology. Typical [**mean AUC ROC, mean P Value, mean SAR value**] (±STD) obtained by averaging results for 300 test cases, for a specific set of retrieval and reuse methodologies, for K=500, were **[0.93**±0.095**, 0.035**±0.0.074**, 0.86**±0.046**]** respectively. We believe that CBR approaches can greatly help in building an information structure for PDD.

References

[1] Costello EJ, Mustillo S, et al. "Prevalence and Development of Psychiatric Disorders in Childhood and

Adolescence." Arch Gen Psychiatry, 2003, 60(8): 837-844.

[2] Asbeh P, et al. Creating Consistent Diagnoses List for DD Using UMLS. Next Generation Information Technology & Systems, 2006: 333-336

[3] Ramnarayan et al. "Paediatric clinical decision support systems." Arch Dis Child, 2002, 87(5): 361-362

[4] Stone EM. American Psychiatric glossary. Washington, DC, American Psychiatric Press, 1998.