

Decision Support Tailored to Business Workflows

Thomas Tamisier, Yoann Didry, Oliver Parisot, Jérôme Wax, Fernand Feltz

Centre de Recherche Public - Gabriel Lippmann, L-4422 Belvaux, Luxembourg

tamisier@lippmann.lu

In Decision Support, there is often a gap between the handling of problems in terms of raw business knowledge and the implementation of formal reasoning towards a unique result [1,2]. Developed through a long-time partnership with the administration in Luxembourg, the Cadral system intends to bridge this formal gap. First, its knowledge base is built from any kind of business procedures and information reference. Second, it uses data interpretation to tune the knowledge according to operational data, and to refine the outputs for the subsequent stages of a collaborative workflow.

The main application of Cadral is the processing of family benefits applications. Instead of the whole modeling of legal knowledge, Cadral is based on an explicit drawing of the mental procedures that governs the processing of the applications by interpreting operational data retrieved from the application files and centralized databases of the national administration along with relations between these procedures to the legislation [3]. All the knowledge is therefore organized into 3 kinds of data: procedural data (processing rules), reference data (law), and application data (individual cases). The procedural knowledge, processed by a resolution engine written in the Soar architecture [4], manages metadata to retrieve and interpret the two other datasets. In particular, individual parameters relevant to a demand, including the presence of required certificates, are dispatched on several tables, and special care must be taken in order to recombine the information. Managing this heterogeneous information through data tables eases the cooperation between the Fund's collaborators for running and updating Cadral.

To adapt to business requirements, Cadral refines the decisions with 3 machine-learning modules:

(1) **Clustering** is based on similarities drawn from key criteria selected by experts when modeling the rules [5]. It is used for two improvements. First, adapting the processing workflow to specific profiles requires special care [6]. In addition to processing application files, Cadral is also used for socio-economic prediction on projected demographic data or administrative and legal framework. The second use of clustering allows interpolating or correcting missing or unreliable data in predictive mode [7].

(2) A variant of clustering, **relation setting** adapts the processing workflow to input data based on pre-established criteria, calculated by the systems and hand-managed by the users. Relation setting is mainly used to refine the processing of specific input profiles (i.e. descriptive data relevant to benefit request files) requiring special care. For instance, Cadral manages a historisation module recording the previous computed results along with errors relevant to profiles' particulars. This historical information is then available to feed a relation base between profiles and possible processing errors.

(3) **Decisional intelligence** is used by Cadral to calculate the operational value of data not directly available (confidential, not recorded...). Manipulating this kind of data is quite different from manipulating missing data, because we never have values for them, so we can't use classical approximation or imputation, and we have to guess their values through mere "discovery" techniques. Hence, Cadral allows building fictive data structures containing values of all available data, and uses neural networks [8] to learn the particulars of data to discover.

References

- [1] P. McCorduck. *Machines Who Think* (2nd ed.). A. K. Peters, 2008.
- [2] T. Redman. The impact of poor data quality on the typical enterprise. *Communications of the ACM*, 1998.
- [3] T. Tamisier & al. A Collaborative Reasoning Maintenance System for a Reliable Application of Legislations. *Proc. of the 6th Int'l CDVE Conference, LNCS, Springer*, 2009.
- [4] J. Laird & al. SOAR: an Architecture for General Intelligence. *Artificial Intelligence*, v. 33-1, 1987.
- [5] V. Kumar, M. Steinbach, P.-N. Tan. *Introduction to Data Mining*. Addison-Wesley, 2006.
- [6] L. Candillier. *Contextualisation, Visualisation et Evaluation en Apprentissage non Supervisé*. PhD thesis, University of Lille-3, France, 2006.
- [7] K. Lakshminarayan, S. Harp, R. Goldman, T. Samad. Imputation of missing data using machine learning techniques. <http://www.aaai.org/Library/KDD/1996/kdd96-023.php>
- [8] Y. Wang, H.-J. Xing. Knowledge Discovery and Integration Based on A Novel Neural Network Ensemble Model. *Proc. of the International Conference on Semantics, Knowledge and Grid*, 2006.