Generation and Representation of Concepts in the Frame-Based Language Objlog+: from Probabilistic Concepts to Prototypes

COLETTE FAUCHER
(LSIS, Polytech'Marsseille)

Modeling the knowledge concerning a given domain in a classical frame-based language requires to be able to represent this knowledge in terms of concepts and instances. However, this representation is not always available at once and it is the occurrence of numerous observations of real entities illustrating a concept that allows to build the abstract representation of this concept. Moreover, within a frame-based language, it is impossible to model such observations that are neither concepts, nor instances. This article aims at presenting our solution to these problems.

First, we have designed a fully extensible new frame-based language called Objlog+, (Faucher, 2001).

Frame-based languages are knowledge representation languages inspired by Minsky's works, (Minsky, 1975). A *frame* in these languages, designates either a concept (which is represented by a *generic frame* often called a *class*), or a concrete entity which illustrates one or more classes (represented by a *specific frame* called an *instance*). A frame, class or instance, is a data structure composed of *slots* (which express the properties of the entity described by the frame or relations between the frame and other ones). Slots are themselves described by means of *facets* (e.g. *Domain*, *Value*, etc.). Facets contribute to the description of the semantics of the slot. Classes are organized within a hierarchy. Relations, called *structural links*, connect frames. In particular, the link *A-kind-of* connects the classes within the hierarchy, the link *Is-a* allows to connect an instance to the class(es) it belongs to.

This trend of knowledge representation stemming from works in cognitive psychology has the advantage of matching the way human beings represent knowledge in their minds. Formal Concept Analysis, (Fensel, 2008), for example, is a technique that leads to a more mathematical representation of knowledge by means of lattices that is less familiar to the human mind.

One of the main features of Objlog+ is its new acceptation of the frame notion that has been simplified and reduced to a simple three-leveled data structure, slot/facet/value with no implicit semantics.

To attach a semantics to frames, it is necessary to define particular frame categories. Creating a category of frames C consists in defining a semantics common to all the frames of C by means of the following elements (they are specified as slots or slot values grouped together in a particular frame, which represents the category C and constitutes the super-frame common to all the frames of this category):

(i) particular slots, called **"specific slots" of category C**, which occur in the description of each frame of category C.

(ii) a notion of **global consistency**, expressed by a relation that must be verified by all the frames of category C. It checks the consistency of a frame relative to category C.

(iii) a notion of **local consistency**, expressed by a relation that must be verified by each slot of the frames of category C, if it is different from the specific slots. It checks that the definition of this slot is authorized in a frame of category C.

(iv) one or more **basic operations (methods)** that apply to all the frames of category C.

There are predefined categories like Parameterized-Frame and Non-Parameterized-Frame, Prototype, Instance, Filter, etc.

In particular, a new frame family has been defined to represent observations, that is, descriptions of real entities by means of valued slots.

Building frame hierarchies from observations is achieved in two steps.

During the first step, FORMVIEW, an incremental concept formation algorithm takes, as main input, observations, to generate multi-perspective probabilistic concept hierarchies, (Furtado, 1996), (Furtado, 1998). FORMVIEW is a COBWEB-like algorithm, (Fisher, 1987), influenced by researches in cognitive psychology on *basic level, probabilistic concepts* and *typicality effects*, (Fisher, 1988), (Murphy, 2004), (Rosch, 1976). FORMVIEW is also based on cognitive psychology works which assume that concept formation is a goal-driven process, (Barsalou, 1983), (Stepp, 1986). Each hierarchy generated corresponds to a perspective defined by a goal.

A probabilistic concept C is a conjunction of characteristics defined by a quadruplet: (j, (j), PD_v, PP_v), where j is an attribute, (j) is a set of values of the attribute j, PD_v is the set of the values of the conditional probabilities P(j=v|C) (*predictability*) for each value v from (j), PP_v is the set of the values of the conditional probabilities P(C|j=v) (*predicteveness*) for each value v from (j)).

FORMVIEW presupposes that the goals of categorization are supplied by several experts and uses a GDN (Goal Dependence Network), (Michalski, 1986), including property relevances relative to different goals as well as implications between initial observed properties (typically surface properties) and properties dependent on the experts' domains (usually functional properties). An observation is then represented, for a given goal, by properties defined by the observer and possibly by some others inferred through the GDN. The categorization process weights observation properties based on their relevance and their importance regarding their predictive power for other properties, (Seifert, 1989).

FORMVIEW uses a category quality measure that takes into account the relevance of observation properties and the categories generated in other perspectives that cover the same set of observations than categories of the current perspective. Such categories are linked together by means of *bridges*, (Furtado, 1998).
All the inputs and outputs of FORMVIEW are modelized by means of predefined or new frame families of Objlog+.

FORMVIEW is a supervised learning tool that benefits from experts' help at several stages of its functioning.

We must note that frame-based representation is particularly relevant because it allows the representation of several hierarchies reflecting each one a certain goal, whereas for building an ontology, (Uschold, 1995), for example, the designer must postulate a given point of view or a specified problem to solve for which the ontology will be used. The ontology is then not general and cannot be reused easily for another purpose. Moreover, in our approach, the goal of categorization is explicitly defined, which is not the case when one builds an ontology.

The second step of the generation of frame hierarchies is achieved via a method transforming probabilistic concept hierarchies into prototype hierarchies, a prototype representing a concept to which typical or default information is added This transformation can be seen in two dimensions: the vertical and the horizontal dimensions, both using the information stored in the probabilistic concepts.

The vertical dimension consists in the definition of the levels of the frame hierarchies, that is, which frames should be maintained in the hierarchy and which should not, according to their predictive power (determination of *preference points*), (Fisher, 1989).

The horizontal dimension consists in the definition of the frames in term of attributes and valued facets. The frequencies of occurrence of properties and their probabilities (predictability and predictiveness) present in the probabilistic concepts are used for the creation of slots and facets which define default values, sufficient and necessary properties, values of exception and correlations of properties within concepts, that enrich the usual representation of prototypes. At the same time, the initial observations are transformed into examples attached to probabilistic concepts, then into instances of prototypes.

Within a unified framework, Objlog+, a double representation of concepts is then maintained at the same time, using probabilistic concepts the description of which can evolve and improve and prototypes that offer a more intelligible representation and thus, help the experts in the task of naming the generated prototypes.

Barsalou, L.W. (1983). Ad Hoc Categories. Memory and Cognition 11(3).

Faucher, C. (2001). Easy definition of new facets in the frame-based language Objlog+. Data and Knowledge Engineering 38,3. 223-263.

Fensel D., Facca F. (2008). Formal Concept Analysis, STI. Innsbruck.

Fisher, D.H. (1987). Knowledge Acquisition via Incremental Conceptual Learning. Machine Learning 2.2.

Fisher, D. H. (1988). A Computational Account of Basic Level and Typicality Effects. Seventh National Conference on Artificial Intelligence.

Fisher, D. H. (1989). Noise-Tolerant Conceptual Clustering. IJCAI.

Furtado J.J., Faucher C., Chouraqui E. (1996). A knowledge acquisition tool for multi-perspective concept formation. Lecture Notes in Computer Science 1076. Springer-Verlag. 229-244.

Furtado, J.J., Faucher, C., Chouraqui, E. (1998). Construction of frame hierarchies using machine learning. Advances in Classification Research 6. Ray Schwartz ed. Information Today Inc. (ITI). Medford. NJ. 37-52.

Michalski, R., Carbonnel, J., Mitchell, T. (1986). Machine Learning, An Intelligence Approach. Vol. II. Morgan Kaufmann. CA.

Minsky, M. (1975). A Framework for Representing Knowledge in the Psychology of Computer Vision, P.H. Winston Ed. McGraw Hill.

Murphy, G.L. (2004). The Big Book of Concepts. MIT Press.

Rosch, E. (1976). Principles of Categorization. In Rosch, E., Lloyd, B. (Eds). Cognition and categorization. Hillsdale. NJ: Erlbaum.

Seifert, C. (1989). A Retrieval Model Using Feature Selection. 6th International Workshop on Machine Learning. Morgan Kauffmann.

Stepp, R., Michalski, R. (1986). Conceptual Clustering: Inventing goal-oriented classifications of structured objects. In Michalski, R., Carbonnel, J., Mitchell, T. (Eds). Machine Learning, An Intelligence Approach. Vol II. Morgan Kaufmann. CA.

Uschold M., King M. (1995). Towards a Methodology for Building Ontologies. Workshop on Basic Ontological Issues in Knowledge Sharing. IJCAI.