An Evolutionary Computational Model of Prototype-Based Categorization: an Application on Clinical Semeiotics

Francesco Gagliardi (francesco.gagliardi@libero.it)
Department of Physical Sciences — University of Naples "Federico II"
Via Cintia — I-80126 Napoli, Italy

Abstract

The aim of this paper is to present a software artifact for machine supported understanding and modelling of prototype-based categorization. This software system is able to perform discovery of syndromes (seen as prototypes) into a given data set of clinical observations. A new genetic algorithm is used to achieve an unsupervised categorization of observation via adaptive learning of number and features of prototypes. Its evolutionary learning is oriented to maximize specificity and distinctiveness of categories. Experimental results show that prototype-based categorization of clinical observation is suitable for syndrome-based categorization. The trichotomy of categorizations (superordinate, basic and subordinate) is explained by trade off between specificity and distinctiveness. Moreover the natural basic level is also related to epistemic adequacy of found prototypes.

Keywords: Categorization; Artificial Intelligence; Prototype Theory; Syndrome Discovery; Basic Level, Real-Life Problem.

In cognitive psychology the problem of the categorization has been formulated, according to prototypes theory, as a problem of similarity between a prototype and the members of the category that this represents.

Cluster analysis is a machine learning task for partitioning a given data set of observations in groups of similar item.

We have realized an evolutionary machine learning system for cluster analysis, which embeds prototype theory of categorization.

The aim is to use this software artifact for machine-supported explanations of adaptive categorization skill observed in human mind.

The multiple realizability of intelligent behaviour is one useful and objective explanatory strategy of cognitive sciences.

Categorization and Prototype Theory

Categorization is one of the most vexed unsolved issues in cognitive science field.

Traditional theories about categorization argue that people categorize using the common features of the members. These features may be viewed as *defining features* because they are singly necessary and jointly sufficient to define the category (Katz, J.J., 1972; Katz, J.J. & Fodor, 1963).

In other words, each feature is an essential element of the category; together, the properties uniquely define the category.

The prototype theory as well as the exemplar-based models of categorization (see Minda, J.P. & Smith, J.D. (2002) for a comparison between models) overcome the constraints of these classical theories.

Notably the prototype theory (Rosch, E. 1975; Rosch, E., & Mervis, C.B. 1975) asserts that people categorize on the basis of how close something is to the *prototype* or ideal member of a category. For example, a robin is closer to the bird prototype than an ostrich is, but they are both closer to it than they are to the prototype of a fish, so we call them both birds, only it takes longer to say an ostrich is a bird than it take to say a robin is a bird, because the ostrich is further from the prototype.

Besides, it is well established that people can categorize the same objects at different levels of abstraction.

Rosch et al.'s (1976) seminal research isolated three "natural" levels of object categorization: the superordinate (e.g. furniture), the basic (e.g. chair), and the subordinate (e.g. Chippendale chair).

Of these, the basic level is known to have a privileged status that is often attributed to the organization of categories in memory.

The origin of the bias to the basic level is still a matter of debate. In categorization, researchers have proposed that categories at the basic level are more differentiated that is, "... have the most attributes common to members of the category and the least attributes shared with members of other [contrasting] categories." (Rosch et al., 1976, p. 435)

The first component of this differentiation definition has been called the specificity (Murphy & Brownell, 1985), or the informativeness (Murphy, 1991) of a category, and the second component the distinctiveness of a category (Murphy & Brownell, 1985; Murphy, 1991).

The difference between distinct types of categorizations would thus stem from distinct differentiations at these two levels.

Both these aspects of categorization have been embedded in our model as explained in next sections.

Cluster Analysis and Categorization

Cluster analysis (Jain et al. 1999) (Kaufman and Rousseeuw 1990) is the unsupervised classification (natural grouping) of observations (patterns, data items, or feature vectors) into groups (clusters) based on similarity. Intuitively, patterns within a valid cluster are more similar to each other than they are to a pattern belonging to a different cluster.

There is an obvious analogy between cluster analysis and the cognitive problem of categorization. The usual algorithms used in such field of the data analysis, (for instance statistical-iterative ones) are considered

unsupervised by contrast to classification ones because they not require training set. However they require some knowledge that the human must supply to the system, as for instance number of categories. Moreover, there is often no explicit maximization of specificity and distinctiveness.

Then these algorithms cannot be considered, *ipso facto*, as computational models of the human mind categorization.

Prototype-Based Cluster Analysis

According to the prototype theory, an observation will be categorized as an instance of a category if it is sufficiently similar to the prototype. Exactly what is meant by similarity to a prototype can be a complex issue, and there are actually different theories of how this similarity should be measured (Smith, E.E. & Medin, 1981). Moreover, many researchers suggest that some features should be weighted more heavily as being more central to the prototype than are other features (e.g. Komatsu, 1992).

In our computational model we have defining a simply sharp clustering criterion based on the number of similar features shared between an observation and the prototype, counted by range weighted city block distance.

We can define quantitatively similarity between item and prototype, thanks to the definition of a distance between observations. Moreover we can define quantitatively the specificity and the distinctiveness of categories as homogeneity intra-cluster and separation inter-cluster, respectively.

Formalization: Definitions and Notations

Let us denote **S** as the universe of all possible elements, said items, with L attributes. Each element is represented by an L-ple: $X \equiv \left(x^{(1)},....,x^{(L)}\right)$, where $X \in \mathbf{S}$ and $x^{(i)}$ denotes its i–th attribute. Let us consider a data set with cardinality *n*, i.e. the subset of S, which must be partitioned. Let us denote this set with $\mathbf{DS} \equiv \{X_1,.....,X_n\}$, with, in general: $\mathbf{DS} \subseteq \mathbf{S}$.

Distance between Observations. Distance between observations is defined based on the type of attributes. For binary attributes we use the Hamming distance, while for numerical attributes we use the linear distance: $\delta^{(i)}(x^{(i)},y^{(i)}) = \left|x^{(i)} - y^{(i)}\right|$. We have chosen as definition of distance between items a normalized *city-block* (or *Manhattan*) distance:

$$d(X,Y) = \frac{1}{L}\sum_{1}^{L}{}_i\left(\frac{1}{R^{(i)}}\cdot\delta^{(i)}(x^{(i)},y^{(i)})\right)$$

where $R^{(i)}$ is the range of i-th attribute:

$$R^{(i)} \equiv \delta^{(i)}\left(\underset{j\in\{1,...,n\}}{\mathfrak{Max}}\left[x_j^{(i)}\right], \underset{j\in\{1,...,n\}}{\mathfrak{Min}}\left[x_j^{(i)}\right]\right)$$

($x_j^{(i)}$ is the i–th attribute of the j–th observation). $\forall X, Y \in \mathbf{DS}$ it results $0 \le d(X,Y) \le 1$

Distance between Clusters. If P_i and P_k represent the prototypes for clusters $\mathbf{C}_i$ and $\mathbf{C}_k$, distance between those clusters is defined as distance between prototypes:

$$d(\mathbf{C}_i,\mathbf{C}_k) = d(P_i,P_k)$$

Prototypes and Sharp Clustering for Cluster Representation. We represent a cluster $\mathbf{C}_k$ through an item of **S** called prototype P_k which, in general, does not coincide with cluster centroid. A generic item $X_i \in$ **DS** is assigned to the cluster $\mathbf{C}_k$ whose representative is the closest. Formally:

if $\forall X_i \;\; \exists\tilde{k} : \forall k \neq \tilde{k} \;\; d(X_i,P_{\tilde{k}}) < d(X_i,P_k)$ then $X_i \in \mathbf{C}_{\tilde{k}}$

A clustering can be denoted with the list of the representatives of its clusters: $\mathbf{CL} \equiv \{P_1,.....,P_m\}$

Homogeneity and Separability. Homogeneity for a cluster $\mathbf{C}_k$ is defined as:

$$H(\mathbf{C}_k) \equiv -\frac{\sum_{\forall X\in\mathbf{C}_K} d(X,P_k)}{w_k}$$

where P_k is the cluster prototype and w_k is the cardinality of k-th cluster. Hence we can define clustering homogeneity as weighted average of homogeneity of clusters, and separability for a clustering as the weighted average of distances among clusters based on above metrics:

$$H(\mathbf{CL}) \equiv \frac{\sum_1^m{}_i w_i \cdot H(\mathbf{C}_i)}{\sum_1^m{}_i w_i}$$

$$S(\mathbf{CL}) \equiv \frac{\sum_1^{m-1}{}_i \sum_{i+1}^m{}_j w_i \cdot w_j \cdot d(\mathbf{C}_i,\mathbf{C}_j)}{\sum_1^{m-1}{}_i \sum_{i+1}^m{}_j w_i \cdot w_j}$$

Due to the above definitions of distance, $-1 \le H(\mathbf{CL}) \le 0, \;\; 0 \le S(\mathbf{CL}) \le 1$.

A high value for *H* is usually related to a clustering with a quite high number of clusters, whereas a high value for *S* usually means that the clustering has a quite low number of clusters. So, maximization of both has contrasting needs.

Prototypes based categorization can be seen as a problem of optimal location for prototypes in observation space rather than finding partitioning for the given data set. Then prototype localization can be performed driven by some measure of goodness as the maximization of intra–cluster homogeneity, inter–cluster separability, or both so that members of the same cluster are as similar as possible one another and as different as possible from members in other clusters.

Being concurrent maximization of homogeneity and separability impossible then categorization can be performed maximizing only a trade off between intra–cluster homogeneity and inter–cluster separability.

This trade-off is embedded in our model by linear dependency of fitness function from both components.

Syndrome Discovery in Clinical Semeiotics

Semeiotics is the medical discipline that studies signs and symptoms addressing towards diagnosis of pathologies. One of its goals is discovery of syndromes.

The term syndrome is the association of several clinically recognizable features which often occur together, in similar way and without explicit reference to its pathological factors.

Therefore a syndrome can be considered as a prototype of a clinical observations class.

A syndrome can be expression of a specific pathology or of totally different pathologies. Discovery of a syndrome is important because it addresses next etiological phase, i.e. investigation of underlying physio–pathological causes, with the aim to determine possible diseases causing syndrome appearance.

For instance SARS (Severe Acute Respiratory Syndrome) is a recent example of a syndrome discovery (a new category among clinical observation) that was later explained with the identification of a causative coronavirus.

Hence discovery of a syndrome is the first phase of scientific discovery process in the medical domain.

Evidently syndrome based categorization are not composed of classical immutabile categories, so they are, by contrast, better represented from dynamically evolving prototypes (cf. Smith, E. E., 1988; Smith, E. E., 1995).

Moreover syndrome detection is useful in facilitating differential diagnosis and in outbreak surveillance.

The Considered Data Set

We have used the "dermatology" data set of UCI machine learning repository (Blake & Merz 1998), which contains clinical dermatological observations. It has been chosen because it also presents classification into pathologies. So we are able to compare syndromes discovered by our system with known pathologies. This concurs us to validate *a posteriori* the etiological quality, i.e. epistemic adequacy, of prototypes founds.

Pathologies present in the data set are six (see Table 1).

Table 1: Pathologies contained in the data set.

Id.	Pathologies	No. of Observation
1	Psoriasis	112
2	Seboreic dermatitis	61
3	Lichen planus	72
4	Pityriasis rosea	49
5	Chronic dermatitis	52
6	Pityriasis rubra pilaris	20

This data set consists of 366 clinical cases, each with 34 attributes, 12 of which clinical and 22 hysto–pathological. Any attribute, apart from family anamnesis and age, is expressed by an integer value in [0, 3] where 0 means that the attribute is not present, 3 states that it is present at its maximum degree, 1 and 2 denote intermediate values.

Family anamnesis is a boolean attribute, has value 1 if any pathology has been observed in the family, 0 otherwise. Age is expressed by an absolute integer value.

Cluster Validation

Cluster validity analysis is the assessment of a clustering procedure's output. Often this analysis uses a domain specific criterion of optimality (Jain et al. 1999).

We define a Pathology Addressing Index, *PAI,* which can be useful to evaluate the degree of utility of discovered syndromes in non–ambiguously addressing etiological investigation of underlying physio–pathological causes of disease, which is supposed unknown.

For these reasons the index *PAI* can quantify the epistemic adequacy of a considered group of prototypes in clinical domain.

Pathology Addressing Index

Given a data set of clinical observations, each item belongs to only one of the p pathologies (classes) based on medical opinion contained in the data set. By sharp clustering, each item is assigned to only one of s syndromes (represented by a cluster). Then a table between s syndromes and p pathologies can be built that corresponds to a matrix **A**, where element a_{ij} is the number of elements in the data set assigned to syndrome j which are affected by pathology i.

Based on this matrix we define a Pathology Addressing Index (*PAI*). Firstly, we define *PAI* for the generic j-th syndrome as follows:

$$PAI^{(j)} = \sum_{1}^{p}{}_{i} \frac{a_{ij}}{\alpha^{(j)}} \quad \text{where} \quad \alpha^{(j)} = \max_{i \in \{1..p\}} a_{ij}$$

This index is 1 in the best case, in which the syndrome represents only one pathology, and is p in the worst case, in which the syndrome addresses uniformly towards all considered pathologies.

Then, we define *PAI* for the whole categorization as the weighted average for $PAI^{(j)}$:

$$PAI = \sum_{1}^{q}{}_{j} w^{(j)} \cdot PAI^{(j)} \Big/ \sum_{1}^{q}{}_{j} w^{(j)}$$

where w_j denotes the weight of j–th syndrome, i.e. the number of clinical cases assigned to j–th category. *PAI* can vary within *1* and p, and the higher its value the more ambiguously on average each syndrome addresses towards p pathologies rather than towards one. *PAI* does not depend explicitly on s, and can tend easily to 1 as the number of syndromes increases. In fact, any clinical case could be considered as a degenerate syndrome on its own.

For the considered data set we have: $1 \leq PAI \leq 6$

It can be observed that it is possible to compute this index only if pathologies are already known. Therefore it can be used for validation of clustering algorithm and it cannot is be used from a clustering algorithm, and of course it was not designed for this.

The Evolutionary Learning System Implemented

We have created an evolutionary learning system, called SDS (*Syndrome Discovery System*), based on a new Genetic Algorithms (GAs) (Holland 1975, Goldberg 1989), named

Self–sizing Genome Genetic Algorithm (De Falco, Della Cioppa, Gagliardi, & Tarantino 2005) developed for unsupervised cluster analysis.

It is self–sizing with respect to the length of genotype encoding the prototypes. This algorithm finds the number of categories and categories themselves through a direct search of the prototypes and then segmentation of data set is achieved with sharp clustering concept.

Genotype variable length is an essential feature of our algorithm which allows us not to require *a priori* knowledge about number of prototypes, because this value is implicitly encoded in the genotype itself, as its length, so it is a part of the solution to find, together, of course, with prototypes themselves. Our system can control on its own the length of genotypes present in the population, thanks also to the definition of a new crossover operator, named gene-pooling.

The adaptive prototypes search is driven by a fitness function that simultaneously maximize homogeneity and separability of the found categories by means of a suitable parameter called *scale factor*.

Genotype

It is constituted by more chromosomes, each encoding a prototype. A prototype belongs to the environment set **S** of the given data set **DS**, so it is a list of attributes according to same order used for objects of the data set. Formally, a genotype is a list (with variable length) of chromosomes defined as:

$$\mathbf{G}^{(k)} \equiv \left(P_1; \ldots; P_k\right) \equiv \left(\left(x_1^{(1)}, \ldots, x_1^{(m)}\right); \ldots\ldots; \left(x_k^{(1)}, \ldots, x_k^{(m)}\right)\right)$$

where P_j is the prototype of j–th cluster, $x_i^{(j)}$ is the i–th attribute of j–th cluster and k is not set a priori. Any attribute is represented by values of same type as attribute in its own variation range.

Fitness Function

After performing data set segmentation based on a genotype **G**, its quality must be evaluated by using a fitness function *f*. Choice of fitness function is critical especially as concerns number of categories *N*.

Following Yip (2002) we have define it as linear function of homogeneity and separation as follows:

$$f_{fitness} = H(\mathbf{G}) + \mu \cdot S(\mathbf{G})$$

where μ represents a scaling factor ($0 \leq \mu$).

The following limit cases exist:

$$\mu \to 0 \;\Rightarrow\; f_{fitness} = H(\mathbf{G}) \Rightarrow \;\; N \to +\infty$$

$$\mu \to +\infty \;\Rightarrow\; f_{fitness} = S(\mathbf{G}) \;\Rightarrow \;\; N \to 1$$

So categorization is formulated as a problem of direct maximization of trade-off between H and S independently of number of categories *N*, yet as μ varies a control on *N* will be indirectly achieved. Given the above ranges for H (**G**) and S (**G**), and constrained $\mu \leq 1$, it turns out that: $-1 \leq f_{\text{fitness}}(\mathbf{G}) \leq +1$.

Genetic Operators

Selection. Selection is based on roulette wheel method.

Mutation Operator. It acts on single prototype attributes of the offspring achieved by gene–pooling, and modifies them with a given probability. This probability is normalized with respect to whole genotype length to avoid that genotypes with more chromosomes are more exposed to mutations.

Gene-pooling Operator. Selected two parents, this operator creates an offspring (prototypes list), with a variable number of chromosomes chosen from the set of chromosomes contained in both parents. *Gene-pooling* does not impose limits on maximal and minimal length of offspring. This operator is of the type: $\left(\mathbf{G}^{(\mathbf{k}_1)}, \mathbf{G}^{(\mathbf{k}_2)}\right) \xrightarrow{k} \mathbf{G}^{(\mathbf{k})}$ with $k \in [2, k_1 + k_2]$. Probability distribution for offspring genotype length k is defined as uniform.

This operator is also based on deterministic internal selection for the choice of chromosomes. Selection takes into account weights (i.e. clusters cardinality), which chromosomes, seen as prototypes, have in parents genotypes.

Given two genotypes $\mathbf{G}_1^{(\mathbf{k}_1)}$ *and* $\mathbf{G}_2^{(\mathbf{k}_2)}$, Algorithm 1 shows gene–pooling procedure.

Algorithm 1: Gene–Pooling Algorithm

1. *Create genetic soup* $\mathbf{GS}^{(\mathbf{k}_1+\mathbf{k}_2)} = \mathbf{G}_1^{(\mathbf{k}_1)} \cup \mathbf{G}_2^{(\mathbf{k}_2)}$
2. *Eliminate duplicate chromosomes* in $\mathbf{GS}^{(\mathbf{k}_1+\mathbf{k}_2)}$ thus obtaining $\mathbf{GS}^{(\tilde{\mathbf{k}})}$
3. *Order chromosomes* in $\mathbf{GS}^{(\tilde{\mathbf{k}})}$ based on weight
4. *Randomly choose* number k of chromosomes for the offspring, $k \in [2, k_1 + k_2]$
5. *Give* the first k prototypes of $\mathbf{GS}^{(\tilde{\mathbf{k}})}$ to offspring genotype $\mathbf{G}^{(\mathbf{k})}$

Our system, thanks to the mechanism of selection internal to crossover operator and thanks to the particular type of fitness, shows the emergent ability to control lengths of genotypes in the population and accordingly it finds the number of categories.

Experimental Results and Discussion

Several tests has been implemented to changing operational parameters of the SDS, particularly we analyse here the test suite performed to varying scale factor μ, that are responsible of the trade-off between homogeneity and separation.

We have taken into account for μ the set of values in the range [0.00; 1.00] with step 0.05. For each such value our genetic algorithm has been run twenty times.

In all experiments we have chosen population size $P_s = 100$ and maximum number of generations $N_g = 250$.

The aim is to study variation of the solutions found, homogeneity, separability and in particular the number of categories with theirs epistemic adequacy as a function of scale factor μ.

The results achieved are shown in Fig.1 and Fig.2. In the former we show the dependence from parameter μ of homogeneity and separability of solutions found.

In the next one we show the dependence from parameter μ of categorizations cardinality and PAI of solutions found.

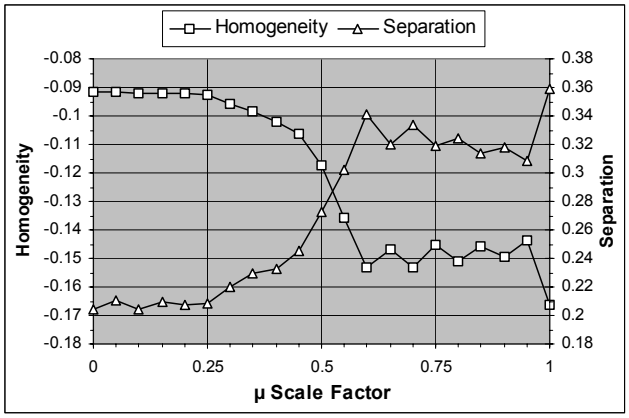

Figure 1: Average values for H, S, as a function of μ scale factor

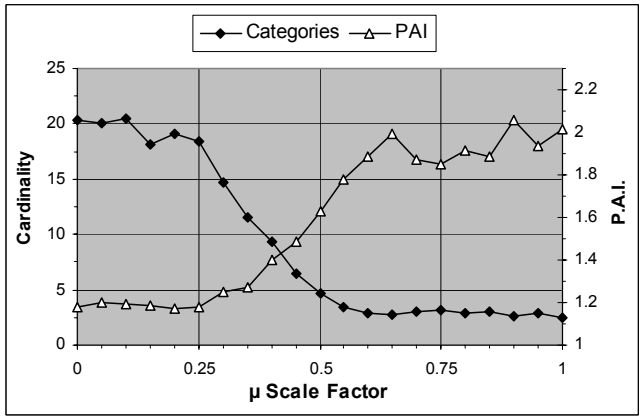

Figure 2: Average values for cardinality and PAI as a function of μ scale factor

From simple theoretical point of view we expect that as μ increases the homogeneity, separability and cardinality change with continuity while in this data we can discriminate three regions.

Two are the limit cases. In fact, we see that for $\mu \leq 0.25$ obtained categorizations have about the same values for homogeneity, separation, number of categories and PAI, likewise for $\mu \geq 0.60$ but with reverse values.

The third and most interesting region is the intermediate one, i.e. that for $0.25 \leq \mu \leq 0.60$. In this region an almost monotonic behaviour can be observed, decreasing for homogeneity and number of categories and increasing for separation and PAI.

We try to explicate this data. In first region ($\mu \leq 0.25$) obtained categorizations magnify more homogeneity rather that separation. These categorizations have high specificity, low distinctiveness and high cardinality, so we can consider them as subordinate categorizations.

In fact the achieved prototypes-syndromes have a good effectiveness, they often individualize an only pathology: the values of PAI are close to 1, the ideal value.

Otherwise in second region ($\mu \geq 0.60$) our system maximizes the separation mainly. Obtained categorizations have low specificity, high distinctiveness and low number of categories, and then they can be considered some categorizations of superordinate type. In fact the prototype-syndromes gotten have a not so good effectiveness, they individualize around two pathologies (PAI ≈ 1.9).

Even if they can be considered some valid categories since they bring values of PAI well distant from the worse possible value (PAI = 6).

In intermediate region we obtain a continuous range of possible categorizations from subordinate ones to superordinate ones, varying from those with great specificity and high number of categories to those with great distinctiveness and small number of categories.

We can observe that the natural hierarchy of categories could in part arise from the trade off between specificity and distinctiveness.

The basic categorizations would be fixed in the middle of intermediary region where a right compromise can be individualized also in relationship to the etiological utility of the found prototypes, quantified thanks to the Pathology Addressing Index.

In fact PAI has a almost constant behaviour for up to values of μ equal to 0.35 therefore it grows linear up to values of μ equal to 0.65 for then to be almost constant again. This behaviour of PAI is not completely lined up with the three regions individualized first.

Therefore we observe that for values of $\mu \approx 0.35$-0.40, the cut into halves of the number of categories to 9-12 (respect the around 20 of the subordinate) produces a worsening of prototype usefulness that are negligible or light: PAI increase from about 1.19 (for $\mu \leq 0.25$) to 1.27 – 1.40.

We could identify for this middle values of μ (as above recognized) of the intermediary region, the basic level observing that this natural level emerges either from the trade-off between homogeneity and separation either from maximizing the etiological usefulness (to minimize the PAI) in comparison to the cognitive effort of increasing the number of prototypes.

In other terms for $\mu \approx 0.35$ our model exhibits a 'critical' point beyond which a magnification of specificity and consequently increase of number of categories is not rewarded by an increase of etiological usefulness of the found prototypes.

This second aspect of basic level can be synthesized through the well known principle of parsimony "*Pluralitas non est ponenda sine necessitate"* expressed by Ockham which translates into English and into medical domain as *syndromes should not be posited without necessity*.

Here, we argue that the basic level would be also characterized by the methodological reductionism, over that

from the right compromise between specificity and distinctiveness.

Concluding Remarks and Future Works

In this work we have adopted an '*artifact approach*' for modelling and understanding a cognitive skill of human mind, according to well rooted perspective of the *'robot approach*' (Hull, C.L. 1943) and mechanistic explanation of intelligent behaviour (Cordeschi, R. 2002).

We have modeled the syndrome discovery process as an adaptive prototype-based categorization.

This assumption has been evaluated through the definition of a domain specific index for cluster validation that measures the effectiveness of the prototype-syndrome founds for addressing etiological investigations.

We have provided an evolutionary computational cognitive model of the prototypes theory of categorization able to reproduce several levels of categorizations spontaneously.

This model shows a possible explanation of the emergency of the categorization levels as trade-off between specificity and distinctiveness. Also, basic level is linked to a point of equilibrium between the aetiological utility of the prototypes and the principle of parsimony.

Moreover, our model, based on a genetic algorithm, implements an *evolutionary* machine learning model for an *adaptive* human learning, so according to E. Mach and his *biological view of knowledge* (Pléh, C. 1997) we can argue a possible algorithmic uniformity between adaptive individual learning and evolutionary learning by species selection.

Future research will need conducted to deeply analyse and model the interplay of trade-off between specificity and distinctiveness with epistemic adequacy of prototypes in clinical field and other ones.

Acknowledgments

I wish to thank Prof. V. Cordeschi for the improving comments provided on an earlier version of this paper.

References

Blake, C.L., & Merz, C.J. (1998). UCI Repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Department of Information and Computer Science, University of California. Irvine, CA.

Cordeschi R. (2002). *The Discovery of the Artificial. Behavior, Mind and Machines Before and Beyond Cybernetics.* Dordrecht: Kluwer Academic Publishers.

De Falco, I., Della Cioppa, A., Gagliardi, F., & Tarantino, E. (2005). A New Variable–Length Genome Genetic Algorithm for Data Clustering in Semeiotics. Accepted paper to *The 20 th ACM Symposium on Applied Computing.* March 13-17, 2005, Santa Fe, New Mexico.

Goldberg, D.E. (1989). *Genetic algorithms in search optimization and machine learnin*g. Reading, MA: Addison–Wesley.

Holland, J.H. (1975). *Adaptation in natural and artificial system*s, (2nd ed.). The University of Michigan Press.

Hull, C.L. (1943). *Principles of behaviour*. NY: Appleton-Century.

Jain, A.K., Murty, M.N., & Flynn, P.J. (1999). Data clustering: a review. *ACM Computing Surveys, 31(3),* 266-323.

Katz, J.J. (1972). *Semantic theory*. New York: Harper & Row.

Katz, J.J., & Fodor, J.A. (1963). The structure of semantic theory. *Language*, *39*, 170-210.

Kaufman, L., & Rousseeuw, P.J. (1990). *Finding groups in data. An introduction to cluster analysi*s. New York: Wiley and Sons.

Komatsu, L.K. (1992). Recent views on conceptual structure. *Psychological Bulletin, 112*(3), 500-526.

Minda, J.P. & Smith, J.D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 275-292.

Murphy, G.L. (1991). Parts in objects concepts: Experiments with artificial categories. *Memory & Cognition, 19*, 423-438.

Murphy, G.L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11*, 70-84.

Pléh, C. (1997). Ernst Mach and Daniel Dennett: Two Evolutionary Models of Cognition. *Proceedings of the nineteenth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Rosch, E. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology*, *104(3)*, 192-233.

Rosch, E., & Mervis, C.B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382-452.

Smith, E. E. (1988). Concepts and thoughts. In R.J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought* (pp. 19-49). New York: Cambridge University Press.

Smith, E. E., (1995). Concept and Categorization. In E.E. Smith & D. N. Osherson (Eds.), *An Invitation to cognitive science* (Vol.3, *Thinking*) (2nd ed., pp. 3-33). Cambridge, MA: MIT Press.

Smith, E. E., & Medin, D.L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Yip, A.M. (2002), A scale dependent data clustering model by direct maximization of homogeneity and separation. *Proc. Mathematical Challenges in Scientific Data Mining IPAM*. January 14–18, 2002.