COMENIUS UNIVERSITY BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS, AND INFORMATICS

Conceptual Spaces with Weighted Prototypes

An Application to Categorization

Diploma Thesis

Alexander Moravčík

Advisor: RNDr. Martin Takáč

2005

Declaration

I do hereby declare that this thesis has been composed by me and that the work described within is my own, unless stated otherwise. I certify it by my signature below.

Alexander Moravčík

Acknowledgement

I would like to express thanks to my advisor RNDr. Martin Takáč for his helpful hints, comments, opinions, and patience, as well as the literature with which he has provided me.

Abstract

This thesis introduces a new integrated formalization of the conceptual spaces framework proposed by Gärdenfors. We extend the notion of conceptual space generated by a set of prototypes by adding a weight vector to each prototype. This will allow us to represent non-uniformly scaled concepts and, as a special case, partial concepts. We also present a novel Weighted Prototypes-Based Categorization Algorithm (WPCA), which is designed to serve as a basis of cognitive apparatus of agents that engage in language games.

Contents

1 Introduction

1.1 Evolutionary Approach to Linguistics

There are many good reasons for studying properties and mechanisms of human natural languages. In the context of computer science, understanding them would provide us with an excellent man-machine communication interface. It should not be forgotten that natural language is still the one and only symbolic medium that enables us to manipulate information of arbitrary content.

Research in linguistics in the 20th century mostly concentrated on the description of syntactic and phonological structures of a language. Moreover, linguists like de Saussure viewed the language as a synchronic phenomenon, deliberately ignoring its historical evolution.

When we look at the domain of computational linguistics, we find out that today's natural language processing systems are almost completely crippled of the acquisition module. The learning process in these systems involves employment of a specially trained person who actually feeds in all the carefully prepared data [9].

Classical linguistics as well as computational linguistics misses the crucial fact that language is a complex adaptive system [15], [20]. Because language is constructed and reconstructed repeatedly by its users, thinking in the terms of idealized speaker or "the" grammar unavoidably leads to a blind alley [17]. To get over this problem, we adopt the hypothesis that language changes and evolves over time, mostly under the pressure of linguistic interactions. A survey of this theory and two opponent ones (i.e., genetic evolution and genetic assimilation) can be found in [15].

1.2 Language Games

1.2.1 The Tenets of Language Games

Modeling inter-individual interactions can be effectively realized by simulated multi-agent systems, which are commonly used in field known as artificial life. Any entity that perceives its environment through sensors and acts upon that environment through actuators is called agent. Every agent has a specific internal state, which, together with the sensory input, determines its behavior. Implicit global properties of the whole population emerge spontaneously as a consequence of explicit purely locally defined interactions. This contrasts with the top-down approach of classical artificial intelligence, where even the global behavior is well prepared and programmed beforehand.

According to the game theory terminology, generic interaction between an agent and the environment (e.g., another agent) is called a game. Linguistic interactions are identified as language games, in line with Wittgenstein. These games are adaptive because they result in a revision of mental states of the participants so that they become more successful in the future games. From a different perspective, the agents act rationally since they try to be as successful as possible. What counts as success obviously depends on what are the agents required to do or what are their motivations.

One might ask how something like common language could appear in a community of individuals who do have neither a total view of the linguistic knowledge of the group nor a complete control over it. In fact, coherence in the population is achieved by a process of self-organization based on a positive feedback loop between a successful use of a certain convention and its further application in the future. If a convention is used successfully, the chance that it will be used in the same or a similar situation again increases. Conversely, the more a convention is used, the higher will be its success ratio [13], [15].

Performed experiments point out an interesting phenomenon—conceptual, phonological, lexicological, and grammatical layers of the investigated models co-evolve. For example, when an agent develops a very useful new concept (mental entity), pressure is on to lexicalize it. Similarly, a concept associated with acoustically individual speech has a natural advantage over concepts, which are difficult to distinguish after externalization.

Plausible models of language dynamics need to respect some important general criteria. A multi-agent system should be:

- distributed: no agent has a complete view of the behavior of the other agents; no agent can directly control the behavior of all his companions; all interactions are strictly local.
- open: there should be an influx and outflow of all components of the system; this includes, but is not limited to, agents, referents of meaning (e.g., objects), and concepts.

In addition, agents ought to be:

- limitedly rational: agents do not have direct access to mental states of each other; the only way information can be exchanged is by interaction.
- situated: every agent is located in a concrete environment.
- embodied: body of an agent allows, disallows, or requires it to perform certain actions; agents should be physically realized whenever possible.
- imperfect: competences of agents, such as perception, memory, or utterance production, do not work ideally all over the time; stochasticity can damage every piece of information or make an operation faulty.

Because all of these constraints are present in real environment, the more of them we build into a model, the more authentic will be the simulation results. On the other hand, doing so will most likely add to the number of dependencies in the system and make description, predictability, and interpretation of its behavior more complicated.

1.2.2 The Results of Selected Experiments with Language Games

One has to appreciate that a full-fledged evolutionary linguistic model would involve perception, conceptualization, lexicon look-up, grammatical parser and generator, speech recognition and synthesis, intention evaluation, and, of course, action. Hence, today's experiments with language games focus on certain subparts of the whole problem. For instance, when simulating the emergence of a shared lexicon, one game could look like this [5]:

1. Two agents, a speaker and a hearer, are randomly selected from the population.
2. The speaker chooses an arbitrary object from the shared context as the topic of the interaction.
3. It retrieves the most preferred of the words associated with the topic and communicates it to the hearer.
4. The hearer maps the word used by the speaker onto the most preferred object according to his own lexicon.
5. If the object decoded by the hearer and the topic are the same, the game is successful. If it is not the case or if the process failed earlier, the game is unsuccessful.

In the end of each game, both agents adaptively modify their lexicons. If the game is successful, they reinforce the used association. If the speaker does not have a word to encode the topic, it can create a new word and associate it with the topic. If the hearer does not know the word used by the speaker, or it does not associate it with the topic, it can acquire the new topic-word relation.

Although iterating this procedure leads to a common vocabulary, it is easy to notice that the limited rationality constraint is broken. When finding out the result of a game, agents telepathically inspect each other's "thoughts." The same happens if the hearer associates a new word with the topic. This problem can be solved by adding some nonverbal form of communication, e.g., grasping, pointing, or gazing. As we can see, no qualitative change would arise in this particular case, but in a more general model in which words do not have to identify objects directly, such ways of communication are quite ambiguous.

Certain important macroscopic characteristics of the population are monitored during the simulation process. In the case of lexicon formation, these can include [2]:

- mean success,
- coherence (i.e., the ratio of agents that use the same word for the same object),
- specificity (a measure that increases with relative frequency of agents using different words for diverse objects).

It has been shown that after sufficient number of games, these quantities nearly reach ideal values and do not drop later. It means that all of the agents speak the same language and that there is no synonymy or homonymy in this language. The lexicon is only insignificantly affected by any consecutive games.

Kaplan investigated the effect of system openness upon the global lexicon [5]. In his study, a new agent with an empty lexicon enters the system from time to time and replaces a randomly chosen agent. If this happens with a reasonable frequency, the new agents become a source of innovation; they can discover new, formerly ignored associations. Since they uniformly interact with all other agents in the population, after some time, their lexicons will capture global trends and suppress marginal tendencies in the system. However, if the fluctuations are without limits, the lexicon has no chance to stabilize. Consequently, communicative success, coherence, and specificity suffer.

Kaplan's experiment is even more appealing because if an agent does not have a word for a particular object, it is allowed to use an existing word associated with a resembling object. Thus, the words become names of whole classes of similar objects instead of labeling only single objects and the lexicon develops into an implicitly represented taxonomic hierarchy.

In experiments performed by Steels and McIntyre, agents were situated in a two-dimensional space [26]. The probability with which two agents interact is based on their respective distance: agents close to each other have a higher chance to communicate than agents apart have. Spatial distribution of the agents is generated randomly, but inclines to clustering. Simulations show that although the success in communication advances as before, inspection of the lexicons of different agents reveals the fact that agents actually develop two languages. They use the first one preferentially within their cluster; the second language, although weaker, is shared among distinct clusters.

At this point, we recall non-situated agents again. We have mentioned that a stable global lexicon emerges in the system after some time and this lexicon is partially resistant to alterations of the population. The situation changes if we incorporate stochasticity in the model. Steels and Kaplan added noise to both the linguistic and nonlinguistic form of communication, as well as to retrieval of information from associative memory, and found out that there are upper bounds on the amount of stochasticity admissible in the initial phase of lexicon formation [21]. If these bounds are exceeded, no self-organization takes place. However, once a language established, malfunction of one component is partially counterbalanced by the other components and by context. Stochasticity introduces and maintains variations in the language—associations strengthen and weaken, new words appear—but communicative success along with coherence remains high.

Later on, Steels and Kaplan modeled agents able to utilize partially damaged information [22]. An agent in the role of hearer takes into account that the perceived word and the word actually used by the speaker might not be identical; they could be only similar. The same holds for uncertainty of extra-linguistic hints. Consequently, the global lexicon will never be completely coherent, but the communication is more resistant to noise and inaccuracies. Interestingly enough, periods of stabilized lexicon were observed, followed by phases of strong reorganization. Such behavior is known to be characteristic of natural languages.

The Talking Heads experiment headed by Steels was a much more large-scaled project [17], [23], [24], [25]. The experimental setup consisted of two steerable cameras connected to computer equipment and oriented towards a whiteboard on which simple, colored geometrical figures, like triangles, circles, and rectangles, were pasted. Each of the two agents that take part in a language game "embodies" in one of the cameras. The speaker then tries to describe the selected object in such a way that it is clearly separated from the other objects. Hearer's task is to find out which of the shapes has been chosen as the topic of the interaction. To make the agents even more situated, similar environments were assembled in several different sites in the world. The agents were able to travel from one location to another place through internet, but only those sharing the same physical environment were allowed to communicate. Authors even set up a public web page that let anyone create his or her own agents, supply them with initial lexicons, place them into the system, and monitor their statistics.

During the experiment, several thousands of words and hundreds of concepts emerged. The core vocabulary consisted of about one hundred words, including eight essential words expressing basic colors (red, green, blue, bright) and locations (left, right, up, down).

Up to now, we have dealt with experiments focused on lexicon, but this is not the only subject, to which this methodology can be applied. Phonetics and phonology are research domains of de Boer who investigates, for example, the emergence of realistic vowel systems [1]. In the model described by Steels and Oudeyer, agents not only develop a common repertoire of individual phonemes, but they also agree on shared syllables [27]. Authors emphasize the overall importance of mimicking the constraints of articulatory apparatus and memory.

The emergence of grammar is still an open issue although some pilot experiments already appeared. Steels models grammar, seen as "structural regularities in the language," using frames. Slots of these structures, which can also contain other frames, have to meet certain syntactic as well as semantic restrictions. Unsurprisingly, constraint propagation and satisfaction methods are used to maintain validity of the frame hierarchy. The hierarchy itself can represent either syntactic [16] or semantic knowledge [19]. It is remarkable that these experiments were performed on visually grounded autonomous robotic agents.

1.3 The Concept of Concepts

If one is concerned with communication, the notion of semantic, i.e., the relation between the words of a language and their meanings, becomes prominent. Two fundamentally different approaches to explicating the meaning of a word or an expression exist. According to the realist semantics, the meanings of terms are somewhere out there in the world. On the other hand, cognitivist theory claims that meanings are mental entities in the head. It has been shown that the realist way is severely flawed because it is quite common for the meaning of a concept to change over time and between contexts. Furthermore, not all mentally constructed objects must have a direct correspondence in the world [4].

However, opting for the cognitive semantics brings about another question: how should we represent concepts? Connectionism and the symbolic paradigm, the two mostly used approaches nowadays, have a number of serious disadvantages from the viewpoint of multi-agent simulations of language phenomena. Whereas the information, which is stored in the hidden layers of artificial neuron networks, is difficult to interpret, we are interested in explicitly represented concepts and the process of their formation. Furthermore, these networks learn very slowly in comparison to the rapid acquisition observed on human subjects who are able to learn a new concept even from a single exemplar. The negative aspect of symbolic representation is the unclear origin of new symbols and the relation of these symbols to the physical world (the symbol grounding problem).

Steels and his colleagues mostly use binary discrimination trees to represent concepts [14]. Each sensory channel of an agent is associated with a particular discrimination tree. The tree partitions the continuous sensory output range into a finite set of disjoint subintervals (Figure 1). If an inner node of this tree corresponds to interval $[a,b)$, its left offspring and right offspring correspond to $[a,(a+b)/2)$ and $[(a+b)/2,b)$, respectively. The root always corresponds to $[0,1)$. A concept can be defined as a partial function that takes a sensor identifier as argument and returns a node of the discrimination tree associ-

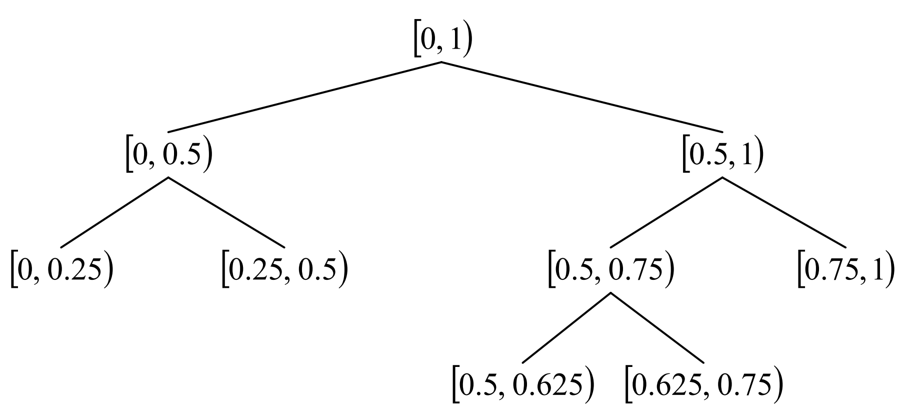

Figure 1: Binary discrimination tree.

ated with the sensor. Note that this can be a leaf node as well as an inner node.

Agents are assumed to have the same sensors, but they neither have to be equipped with the same set of discrimination trees nor are the trees given in advance. Instead, they are adaptively revised according to specific objects, which an individual agent encounters. As a result, the agents develop capacity to represent concepts with precision that adequately reflects the characteristics of the environment.

The major drawback of binary discrimination trees is that they have no known counterpart in human cognition. We present a more plausible model in the following chapter.

2 Conceptual Spaces

2.1 Metric Conceptual Spaces

As we argued in section 1.3, computationalism and connectionism are not the most appropriate paradigms if we demand explicitly represented concepts and our primary interest lies in their formation and their relation to the outer world. The theory of conceptual spaces proposed by Gärdenfors is one of few representational frameworks that meet all these requirements [4]. In this chapter, we will build the notion of conceptual space—on various levels of abstraction and with different representational strengths. Note that the presented attempt to formalize conceptual spaces is original.

We start with some general definitions.

Definition 1. *A set P is a partition of a set X if and only if the following conditions hold for all $A, B \in P$:*

$$A \neq \varnothing,$$

$$\bigcup_{C \in P} C = X,$$

$$A \neq B \Rightarrow A \cap B = \varnothing.$$

The elements of P are called the blocks of the partition.

Definition 2. *Let $X \neq \varnothing$ be a set. We call the elements of X percepts or (possible) objects and the nonempty subsets of X concepts. We say that a percept x and a concept c match just if $x \in c$. A conceptual space over X is a finite partition of X.*

It is an inherent property of human cognition that certain members within a class of objects are considered more representative of the category than others. For example, sparrows and blackbirds are judged more typical birds than are penguins and emus. The most representative members of categories are called prototypes, in conformity with the prototype theory of categorization developed by Rosch and her collaborators [11]. To determine the degree of representativeness, we need to be able to measure the distance between any two possible objects. Since it is a common assumption that similarity is an exponentially decaying function of distance, we can state that greater distance indicates greater dissimilarity [4].

Definition 3. *A metric on a set* X *is any function* $d : X \times X \to R$, *which satisfies the following conditions for all* $x, y, z \in X$ *:*

$$d(x, y) \geq 0 \text{ (non-negativity)},$$
$$d(x, y) = 0 \Leftrightarrow x = y \text{ (identity of indiscernible)},$$
$$d(x, y) = d(y, x) \text{ (symmetry)},$$
$$d(x, y) + d(y, z) \geq d(x, z) \text{ (triangle inequality)}.$$

Definition 4. *A metric space is a pair* (X, d), *where* X *is a set and* d *is a metric on* X.

Definition 5. *Let* (X, d) *be a metric space, let* $P = \{p_i\}_{i=1}^{m} \subseteq X$ *be a set of prototypes for some* $m \in Z^+$, *and let* $s : 2^P \to P$ *be a function. The conceptual space over* (X, d) *generated by* P *and by* s *is the set* $C = \{c_i\}_{i=1}^{m}$, *where*

$$c_i = \{x \in X : p_i = s(\{p_j : d(x, p_j) = \min_{k=1}^{m} d(x, p_k)\})\}.$$

The function s *is called the selection function. We say that* $x \in X$ *and* p_i *match exactly when* x *and* c_i *match.*

Remark 1. *The domain of the selection function in Definition 5 can be safely restricted to*

$$\{P' \subseteq P : \exists x \in X : P' = \{p_j : d(x, p_j) = \min_{k=1}^{m} d(x, p_k)\}\},$$

where the free variables have the same denotation as in the definition.

Lemma 1. *Let* (X,d) *be a metric space, let* $P \subseteq X$ *be a set of prototypes such that* $|P| \in Z^+$, *and let* $s : 2^P \to P$ *be a function. Then the conceptual space over* (X,d) *generated by* P *and by* s *is a conceptual space over* X.

So far, we have concentrated on percepts, concepts, and conceptual spaces with no internal structure. From now, we will presume that possible objects are feature vectors whose components are real numbers.

Remark 2. *The set* $\overline{R} = R \cup \{-\infty, +\infty\}$ *is called the set of (affinely) extended real numbers. Interval notation, order relations, and arithmetic operations on* $\overline{R}$ *are extended intuitively, e.g.,* $\overline{R} = [-\infty, +\infty]$, $x < +\infty$ *for all* $x \in R$, $x + \infty = \infty$ *if* $x \neq -\infty$. *As a convention, we accept* $0 \cdot (\pm\infty) = \pm\infty \cdot 0 = 0$.

Definition 6. *The Minkowski metric* d_λ^n *is defined for all* $n \in Z^+$ *and all* $\lambda \in [1, \infty]$ *as*

$$d_\lambda^n(\mathbf{x}, \mathbf{y}) = \lim_{l \to \lambda} \sqrt[l]{\sum\nolimits_{i=1}^n |x_i - y_i|^l} ,$$

where $\mathbf{x} = (x_i)_{i=1}^n \in R^n$ *and* $\mathbf{y} = (y_i)_{i=1}^n \in R^n$. *For special values of the parameter* λ, *this simplifies to:*

$$d_1^n(\mathbf{x}, \mathbf{y}) = \sum\nolimits_{i=1}^n |x_i - y_i| \text{ (the Manhattan metric),}$$

$$d_2^n(\mathbf{x}, \mathbf{y}) = \sqrt{\sum\nolimits_{i=1}^n (x_i - y_i)^2} \text{ (the Euclidean metric),}$$

$$d_\infty^n(\mathbf{x}, \mathbf{y}) = \max\nolimits_{i=1}^n |x_i - y_i| \text{ (the Chebyshev metric).}$$

Lemma 2. (R^n, d_λ^n) *is a metric space for all* $n \in Z^+$ *and all* $\lambda \in [1, \infty]$.

Corollary 1. *Suppose that* $P \subseteq R^n$, *where* $n \in Z^+$, *is a set of prototypes such that* $|P| \in Z^+$ *and that* $s : 2^P \to P$ *is a function. Then the conceptual space over* (R^n, d_λ^n) *generated by* P *and by* s *is a conceptual space over* R^n *for all* $\lambda \in [1, \infty]$.

A conceptual space over a metric space generated by a set of prototypes is also known as the Voronoi tessellation (Figure 2). If the tessellation is based on the Euclidean

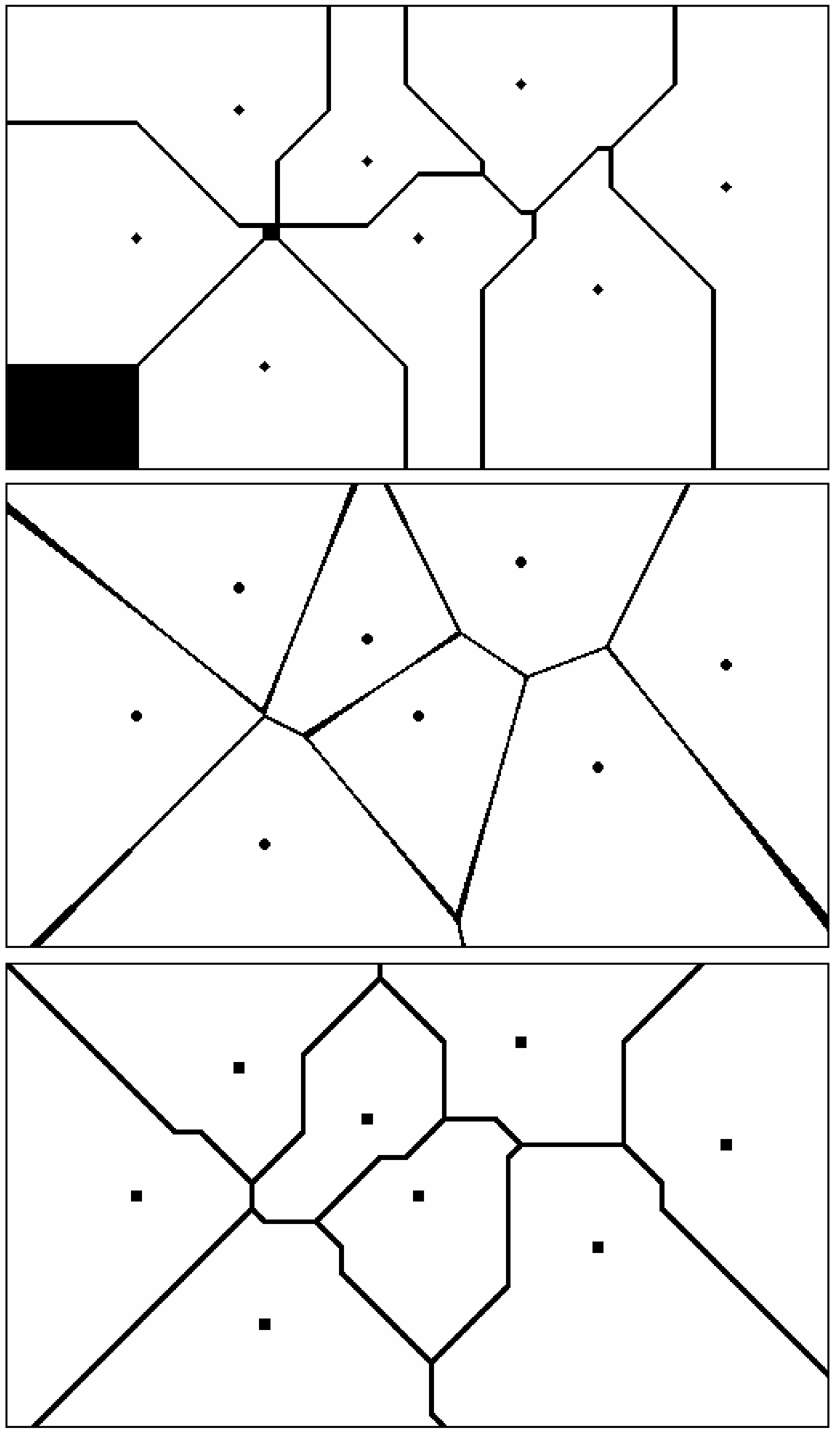

Figure 2: Voronoi tessellation based on the Manhattan metric (top), the Euclidean metric (middle), and the Chebyshev metric (bottom).

metric, the concepts are guaranteed to be convex, i.e., the line segment joining any pair of objects that match a concept lies entirely in the concept [8]. Neither the Manhattan metric nor the Chebyshev metric assures convexity of the Voronoi regions. However, the concepts produced by these metrics are always star-convex with respect to the prototypes, i.e., a line segment from a prototype to any object matching the prototype is contained in this concept.

We now define a useful generalization of the Minkowski metric.

Definition 7. *The (multiplicatively) weighted Minkowski metric* $d_\lambda^n(\mathbf{w})$ *is defined for all* $n \in Z^+$, *all* $\lambda \in [1,\infty]$, *and all weight vectors* $\mathbf{w} \in (0,\infty)^n$ *as*

$$d_\lambda^n(\mathbf{w})(\mathbf{x},\mathbf{y}) = \lim_{\alpha \to \lambda} \sqrt[\alpha]{\sum_{i=1}^n \left(|x_i - y_i|/w_i\right)^\alpha} ,$$

where $\mathbf{w} = (w_i)_{i=1}^n$, $\mathbf{x} = (x_i)_{i=1}^n \in R^n$ *and* $\mathbf{y} = (y_i)_{i=1}^n \in R^n$.

Lemma 3. $(R^n, d_\lambda^n(\mathbf{w}))$ *is a metric space for all* $n \in Z^+$, *all* $\lambda \in [1,\infty]$, *and all* $\mathbf{w} \in (0,\infty)^n$.

Corollary 2. *Suppose that* $P \subseteq R^n$, *where* $n \in Z^+$, *is a set of prototypes such that* $|P| \in Z^+$ *and that* $s : 2^P \to P$ *is a function. Then the conceptual space over* $(R^n, d_\lambda^n(\mathbf{w}))$ *generated by* P *and by* s *is a conceptual space over* R^n *for all* $\lambda \in [1,\infty]$ *and all* $\mathbf{w} \in (0,\infty)^n$.

The role of the weight vector in the generalized Minkowski metric can be viewed from two different perspectives. First, the non-weighted version of the metric assumes that all features have identical scales. If this is not the case, the distance (dissimilarity) reported by the metric will be inappropriate. The solution is to appropriately stretch or compress the relevant dimensions by using weights, which are greater than one or less than one, respectively. For example, we could apply the weight vector $(\sup X_i - \inf X_i)_{i=1}^n$, where $X_i \subseteq R$ is the set of all possible values of the i^{th} feature. Using the weight vector $(\sigma_i)_{i=1}^n$, where σ_i is the standard deviation of the probability distribution of the i^{th} feature, is not only viable even if X_i is unbounded for some i, but it also to large extent substitutes standardization of the feature variables utilized, e.g., in [10].

From a cognitive point of view, the weights allow to adjust the level of attention given to individual dimensions [4]. Changes of context, which generally result in altered "salience" of dimensions, throw light upon subjects' judgments being reported to violate metric properties. For example, it has been demonstrated that Tel Aviv is judged more similar to New York than New York is similar to Tel Aviv [30]. This can be explained by an unaware change of context from geographical to political and vice versa, which causes the weight vector to change, too.

2.2 Conceptual Spaces with Weighted Prototypes

When we read Definition 5 with a keen eye, we realize that the assumption that d is a metric is unnecessarily restrictive. We do not need to know the distance between arbitrary two objects, because according to the definition, one of these objects is always a prototype. This fact helps us to define a novel, generalized form of a conceptual space over $\left(R^n, d_\lambda^n(\mathbf{w})\right)$ generated by a set of prototypes.

Remark 3. *By symbol* $\overline{d}_\lambda^n(\mathbf{w})$*, we denote the (affine) extension of the (multiplicatively) weighted Minkowski metric if* $n \in Z^+$*,* $\lambda \in [1,\infty]$*, and* $\mathbf{w} \in [0,\infty]^n$*. Note that if* $\mathbf{w} \notin (0,\infty)^n$*,* $\overline{d}_\lambda^n(\mathbf{w})$ *itself is not a metric.*

Definition 8. *Let* $PW = \{(\mathbf{p_i}, \mathbf{w_i})\}_{i=1}^m \subseteq R^n \times [0,\infty]^n$ *be a set of (multiplicatively) weighted prototypes for some* $m, n \in Z^+$ *and let* $s : 2^{PW} \to PW$ *be a function. The conceptual space over* $\left(R^n, d_\lambda^n\right)$ *generated by* PW *and by* s *is the set* $C = \{c_i\}_{i=1}^m$*, where*

$$c_i = \left\{\mathbf{x} \in R^n : (\mathbf{p_i}, \mathbf{w_i}) = s\left(\left\{(\mathbf{p_j}, \mathbf{w_j}) : \overline{d}_\lambda^n(\mathbf{w_j})(\mathbf{x}, \mathbf{p_j}) = \min_{k=1}^m \overline{d}_\lambda^n(\mathbf{w_k})(\mathbf{x}, \mathbf{p_k})\right\}\right)\right\}$$

for all $\lambda \in [1,\infty]$.

Lemma 4. *Let $PW \subseteq R^n \times [0,\infty]^n$ be a set of (multiplicatively) weighted prototypes such that $|P|, n \in Z^+$ and let $s : 2^{PW} \to PW$ be a function. Then the conceptual space over (R^n, d_λ^n) generated by PW and by s is a conceptual space over R^n for all $\lambda \in [1,\infty]$.*

When we conceive concepts as Voronoi regions, which start to grow from prototypes at the same time, by introducing the weights of prototypes, we allow the regions to have different growth rates. Furthermore, the growing process can be non-uniform, i.e., a region can be expanded with dissimilar speed along each dimension. The topology of conceptual spaces generated by weighted prototypes can be quite rich: the concepts can

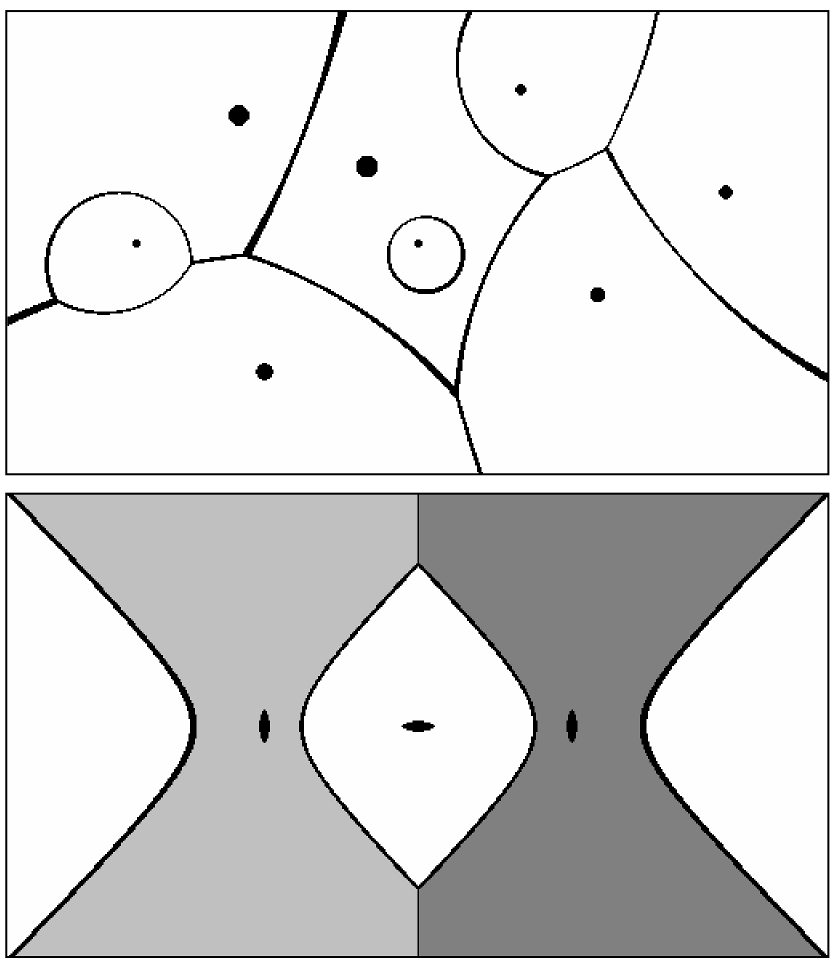

Figure 3: Conceptual spaces generated by sets of weighted prototypes can contain hierarchical concepts (top) and disconnected concepts (bottom).

form (implicit) hierarchies and even do not have to be connected (Figure 3). Note that Definition 8 allows the weights to be infinite, which can be used to represent partial concepts. For example, concepts of adjectives can be modeled by prototypes whose weight vectors contain only one finite component. In addition, zero weights make the corresponding features mandatory. As a special case, prototypes with all their weights being zero can stand for percepts, i.e., degenerated concepts.

3 Categorization Based on Weighted Prototypes

3.1 The Categorization Algorithm

In this chapter, we propose a novel categorization algorithm, which takes advantage of the representational strength of conceptual spaces with weighted prototypes—the Weighted Prototypes-Based Categorization Algorithm (WPCA). It is designed to serve as a basis of cognitive apparatus of agents that engage in language games. As a part of a game, agents perceive a scene consisting of a number of objects, which are guaranteed to be taken from (pairwise) distinct categories. One of the objects is considered focal, the remaining objects create context. The agents' task is to conceptualize the focus uniquely, i.e., in such a way that the prototype matching the focus does not match an object from the context.

The key idea of the algorithm is to keep the mean examples of categories as prototypes' positions and the mean absolute deviations from the prototypical members as weights. The WPCA also maintains two supplemental measures with each prototype. The first one is the prototype's importance, which increases whenever the prototype and the focal percept match. The second quantity, which is the success of the prototype, is determined by the number of unambiguous matches of focus. Formally, a conceptual space is represented by a multiset

$$Q = \{(\mathbf{p}_i, \mathbf{w}_i, suc_i, imp_i)\}_{i=1}^{m} \subseteq R^n \times [0,\infty]^n \times [0,1] \times [0,1].$$

Let $\lambda_p, \lambda_w, \lambda_{suc}, min_{suc}, \lambda_{imp}, min_{imp} \in [0,1]$; typically, these parameters are close to zero. The WPCA works as follows:

1. $Q := \varnothing$.
2. Generate the next context $Y \subseteq R^n$ and the focus $\mathbf{x} \in R^n$.
3. If $Q = \varnothing$, then $Q := \left\{\left(\mathbf{x}, (0)_{i=1}^n, 1, 1\right)\right\}$.
4. Suppose that $\mathbf{x}$ and $\mathbf{q} := (\mathbf{p}, \mathbf{w}, suc, imp) \in Q$ match. $Q := Q \setminus \{\mathbf{q}\}$. If any $\mathbf{y} \in Y$ and $\mathbf{q}$ match, then $success := 0$; otherwise, $success := 1$.
5. $\mathbf{p}' := (1 - \lambda_p)\mathbf{p} + \lambda_p \mathbf{x}$, $\mathbf{w}' := (1 - \lambda_w)\mathbf{w} + \lambda_w abs(\mathbf{x} - \mathbf{p})$, $imp' := (1 - \lambda_{imp}) \cdot imp + \lambda_{imp}$.
6. $suc' := (1 - \lambda_{suc}) \cdot suc + \lambda_{suc} \cdot success$. If $suc' \geq min_{suc}$, then $Q' = \varnothing$, else $suc' := 1$, $imp' := imp/2$, $Q' = \{(\mathbf{p}', \mathbf{w}', suc', imp')\}$.
7. $Q'' := \varnothing$. For each $(\mathbf{p}'', \mathbf{w}'', suc'', imp'') \in Q$ such that $(1 - \lambda_{imp}) \cdot imp'' \geq min_{imp} / |Q|$, do $Q'' := Q'' \cup \{(\mathbf{p}'', \mathbf{w}'', suc'', (1 - \lambda_{imp}) \cdot imp'')\}$. $Q := Q''$.
8. $Q := Q \cup Q' \cup \{(\mathbf{p}', \mathbf{w}', imp', suc')\}$.
9. Repeat from step 2.

Some steps of the algorithm need further explanation. In the 3rd step, a first prototype is created. By using the percept to construct the prototype, we can speed up the process under most circumstances. In step 4, we find the prototype that matches the focus. Step 5 is the heart of the WPCA: we shift the prototype towards the focal object, and update the weights so that they became closer to the absolute deviation of the perceived features from the current prototype position. In step 6, if the success measure of the prototype is below a given limit, the prototype is split, i.e., its success is reset to one, its importance is set to half of the value before the division, and a copy of the prototype is created. The loop in step 7 removes all prototypes, which are used infrequently. Note that the admissible lower bound of prototypes' importance depends on the current prototype count.

The presented categorization algorithm is derived from the Distributed Clustering Algorithm (DCA) [6]. Both of these algorithms are intended to work incrementally, to determine the number of needed categories automatically, and to make only necessary assumptions about the distribution of the data. However, there are some major differences. In the DCA, which is an unsupervised clustering technique, there is no context or focus, only a sequence of unlabeled examples. On the other hand, the WPCA exploits (and requires) that the category from which the focal object is taken never has a representant in the context at the same time. Furthermore, the DCA uses a roundness condition when checking whether a prototype should be split. If the corresponding class of objects is not considered round enough, it will be "covered" by several prototypes. This contrasts with categorization based on weighted prototypes, which allows even non-uniformly scaled concepts to be represented by a single prototype.

3.2 The WPCA in Action

In this section, we demonstrate the behavior of the WPCA by confronting it with several tasks of different complexity. The objective here is to categorize a number of monochrome (black-and-white) raster images that capture various geometrical shapes. Each image is converted into a vector whose components can be any of the following twenty numerical features:

- area,
- perimeter,
- (total) hole area,
- (total) hole perimeter,
- compactness ($area / perimeter^2$),
- (total) hole compactness ($hole\,area / hole\,perimeter^2$ if $hole\,perimeter \neq 0$, otherwise zero),
- solidity ($area / (area + hole\,area)$),
- minimum radius (radial values are determined by measuring the distance from each perimeter point to the shape's centroid),
- maximum radius,
- range of radius ($max\,radius - min\,radius$),
- mean radius,

- standard deviation of radius,
- diameter (the maximum distance between two perimeter points),
- relative minimum radius ($min\,radius / mean\,radius$),
- relative maximum radius ($max\,radius / mean\,radius$),
- relative range of radius ($radius\,range / mean\,radius$),
- relative standard deviation of radius ($radius\,std\,dev / mean\,radius$),
- relative diameter ($diameter / mean\,radius$),
- "zero" (auxiliary feature),
- random value (auxiliary; random numbers are taken from the interval $[0,1)$ with uniform distribution).

Three testing datasets were used, each consisting of three categories of objects: triangles, squares, and circles. In the first case, the figures were generated to have ideal contours; in the second case, the shapes were hand-drawn (mouse-drawn) and hence are quite distorted (Figure 4). Both sets contained thirty-three examples of each category. The third trial involved all objects from the previous two sets. The WPCA was parameterized by the following values: $\lambda_p = \lambda_w = \lambda_{suc} = \lambda_{imp} = 0.05$, $min_{suc} = min_{imp} = 0.1$.

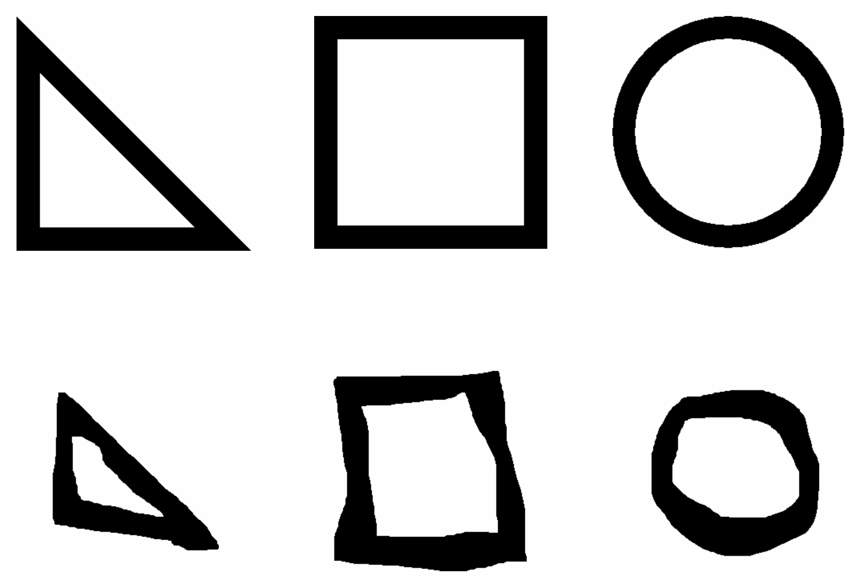

Figure 4: Ideal and distorted samples of the three categories used during the WPCA trial.

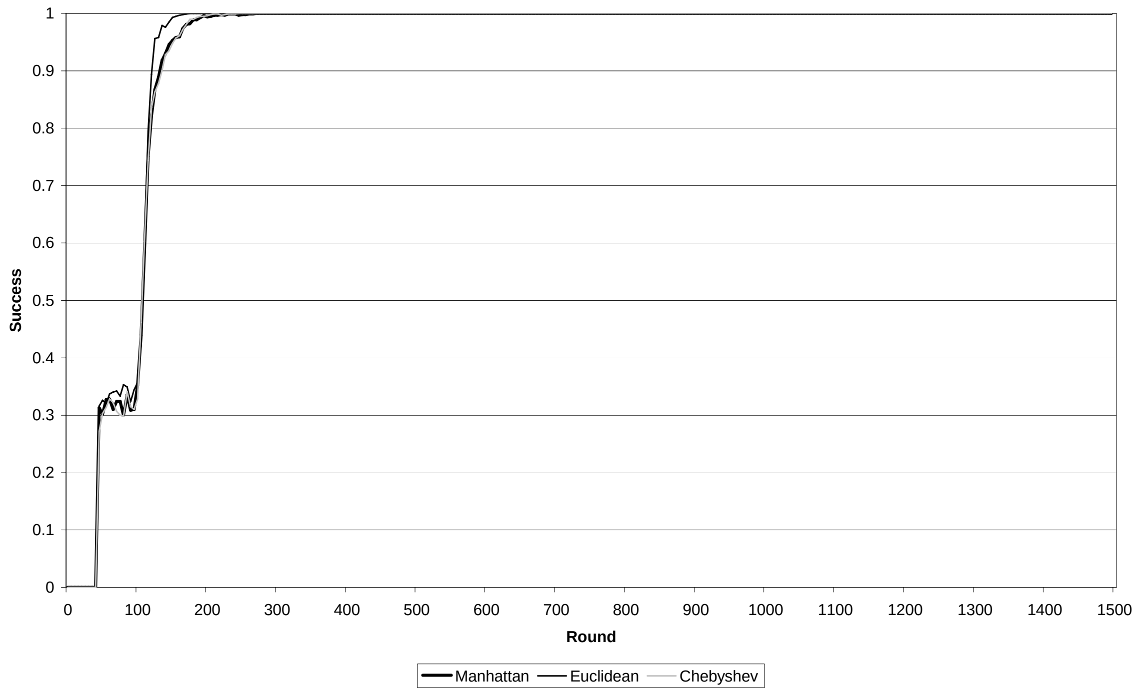

Figure 5: Results of the WPCA trial with ideal two-dimensional samples.

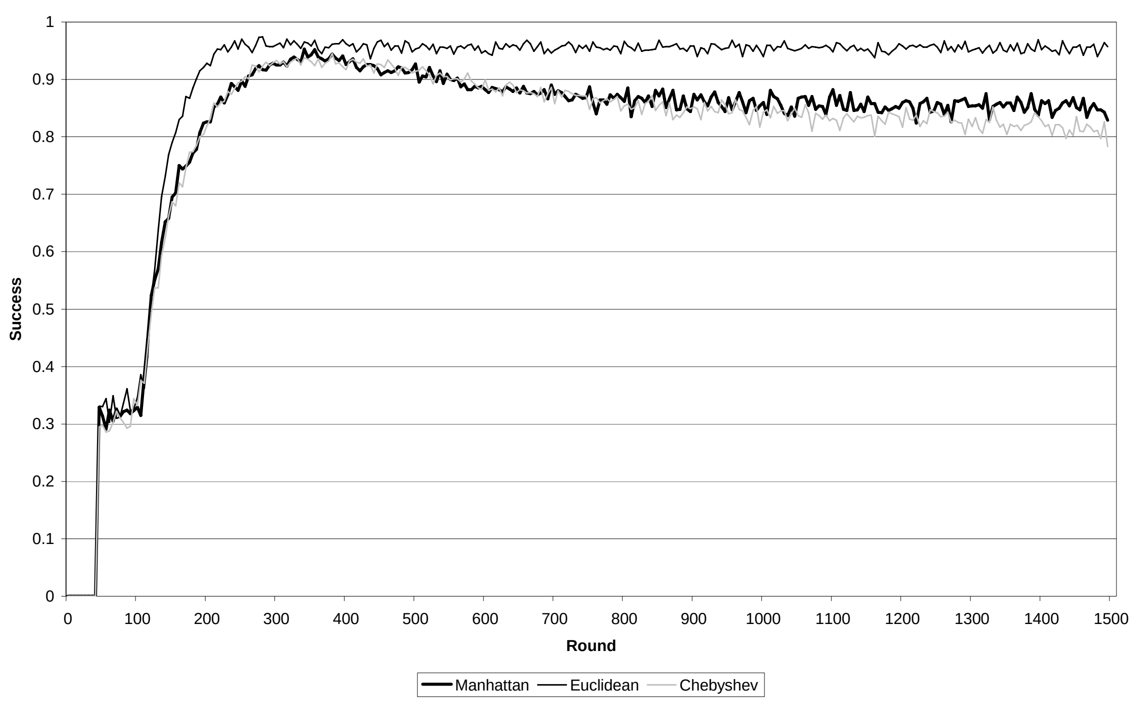

Figure 6: Results of the WPCA trial with two-dimensional distorted samples

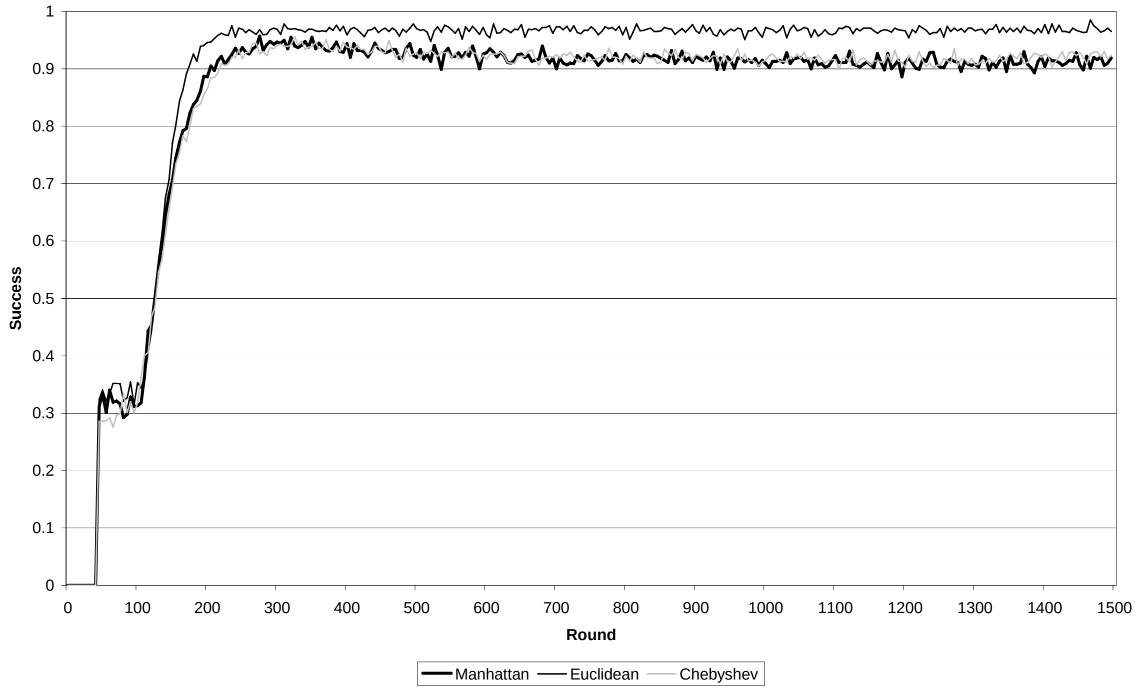

Figure 7: Results of the WPCA trial with two-dimensional combined samples.

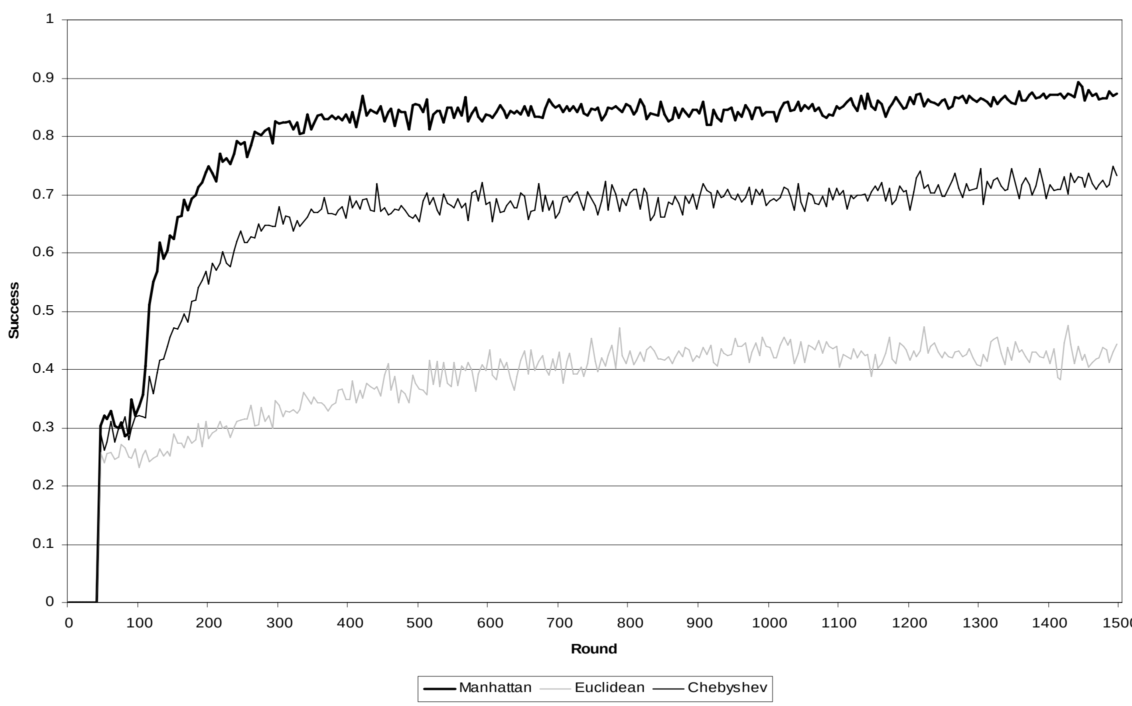

Figure 8: Results of the WPCA trial with ideal twenty-dimensional samples.

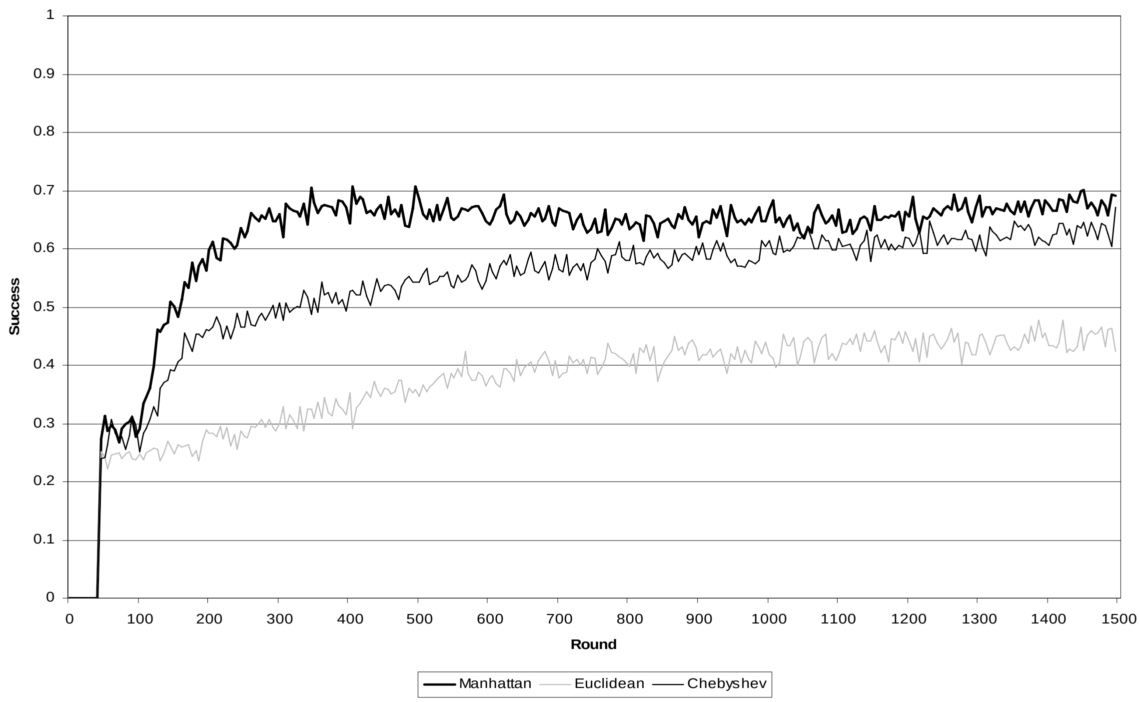

Figure 9: Results of the WPCA trial with twenty-dimensional distorted samples.

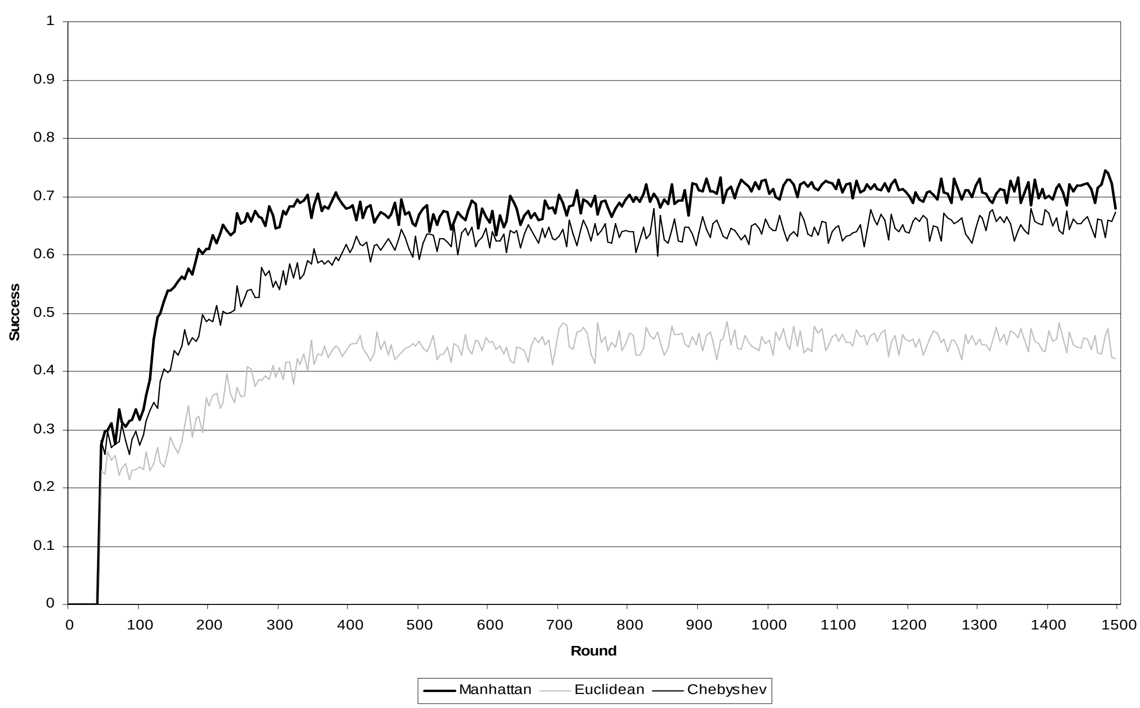

Figure 10: Results of the WPCA trial with twenty-dimensional combined samples.

Figures 5–7 show the categorization progress in the case when the feature vectors are two-dimensional and consist of relative radial standard deviation and of relative diameter. The success in categorization if all of the twenty features are used is shown on Figures 8–10. All six graphs are averaged over 1000 runs of the whole categorization process. Note that the average number of prototypes converges to three, i.e., there is one prototype for each category.

As can be seen clearly, the Euclidean metric outperforms the other two distance functions in the two-dimensional case. On the other hand, when all twenty features are employed, the best results can be achieved by using the Manhattan metric. This reflects the fact that the two selected features are quite independent. However, if we use more features, chances are that some new dependencies arise. As stated by Gärdenfors, if the Euclidean metric fits the data best, the dimensions are integral; if the Manhattan metric gives the best results, the dimensions are separable [4].

4 Conclusion

In chapter 1, we introduced the basic notions necessary to understand the required characteristics of a conceptualization scheme employed by agents, which engage in language games. In chapter 2, we not only formalized some of the ideas behind conceptual spaces proposed by Gärdenfors, but also extended the standard conceptual spaces generated by sets of prototypes by adding an individual vector of weights to each of the prototypes. These weights allow us to represent non-uniformly scaled concepts, special cases of which are partial concepts. We devoted chapter 3 to the novel Weighted Prototypes-Based Categorization Algorithm (WPCA), which is intended to be employed by agents playing language games.

As a bonus for the reader, we propose two further improvements to the WPCA. First, an additive vector of weights could be stored with each prototype that we believe would let the algorithm handle conceptual hierarchies, i.e., "overlapping" concepts. Finally, the weight vector could be replaced with a covariance matrix, which would allow a modified version of the WPCA to handle even more generally shaped and oriented concepts.

5 References

[1] De Boer, B. Investigating the Emergence of Speech Sounds. In: T. Dean (ed.), *IJCAI-99*. San Francisco: Morgan Kaufman, 1999. pp. 364–369.

[2] De Jong, E.D. The Development of a Lexicon Based on Behavior. In: H. la Poutré, J. van den Herik (eds.), *Proceedings of the Tenth Netherlands/Belgium Conference on Artificial Intelligence NAIC'98*. Amsterdam: CWI, 1998. pp. 27–36.

[3] De Jong, E.D., Vogt, P. How Should a Robot Discriminate Between Objects? A Comparison Between Two Methods. In: *Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior SAB'98*. Cambridge (MA): MIT Press, 1998. pp. 86–91.

[4] Gärdenfors, P. *Conceptual Spaces: The Geometry of Thought*. Cambridge (MA): MIT Press, 2000. A Bradford Book. ISBN 0-262-07199-1.

[5] Kaplan, F. A New Approach to Class Formation in Multi-Agent Simulations of Language Evolution. In: Y. Demazeau (ed.), *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS 98)*. Los Alamitos (CA): IEEE Computer Society, 1998. pp. 158–165.

[6] Hulth, N., Grenholm, P. *A Distributed Clustering Algorithm*. Lund: Lund University Cognitive Studies, 1998, vol. 74. ISSN 1101-8453.

[7] Kaplan, F. Semiotic Schemata: Selection Units for Linguistic Cultural Evolution. In: M. Bedau, J. McCaskill, N. Packard, S. Rasmussen (eds.), *Proceedings of Artificial Life VII*. Cambridge (MA): MIT Press, 2000. pp. 372–381.

[8] Okabe, A., Boots, B., Sugihara, K. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. New York (NY): John Wiley & Sons, 1992.

[9] Páleš, E. *Sapfo: Parafrázovač slovenčiny. Počítačový nástroj na modelovanie v jazykovede* [SAPFO: Slovak Paraphraser. A Computational Tool for Linguists]. 1. vydanie. Bratislava: VEDA, 1994. ISBN 80-224-0109-9.

[10] Raubal, M. Formalizing Conceptual Spaces. In: A. Varzi, L. Vieu (eds.), *Formal Ontology in Information Systems: Proceedings of the Third International Conference (FOIS 2004)*. Amsterdam: IOS Press, 2004, vol. 114, pp. 153–164. Frontiers in Artificial Intelligence and Applications.

[11] Rosch, E. Prototype Classification and Logical Classification: The Two Systems. In: E. Scholnik (ed.), *New Trends in Cognitive Representation: Challenges to Piaget's Theory*. Hillsdale (NJ): Lawrence Erlbaum, 1978, pp. 73–86.

[12] Smith, A.D.M. Intelligent Meaning Creation in a Clumpy World Helps Communication. *Artificial Life*, 2003, vol. 9, no. 2, pp. 175–190.

[13] Steels, L. Synthesising the Origins of Language and Meaning Using Co-Evolution, Self-Organisation, and Level Formation. In: J. Hurford, C. Knight, M. Studdert-Kennedy (eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases*. Edinburgh: Edinburgh University Press, 1997.

[14] Steels, L. Constructing and Sharing Perceptual Distinctions. In: M. van Someren, G. Widmer (eds.), *Proceedings of the European Conference on Machine Learning*. Berlin: Springer-Verlag, 1997.

[15] Steels, L. The Synthetic Modeling of Language Origins. *Evolution of Communication Journal*, 1997, vol. 1, no. 1, pp. 1–34.

[16] Steels, L. The Origins of Syntax in Visually Grounded Robotic Agents. *Artificial Intelligence*, 1998, no. 103, pp. 1–24.

[17] Steels, L. The Puzzle of Language Evolution. *Kognitionswissenschaft*, 1999, vol. 8, no. 4.

[18] Steels, L. Mirror Neurons and the Action Theory of Language Origins. *Architectures of the Mind, Architectures of the Brain*, September 2000.

[19] Steels, L. The Emergence of Grammar in Communicating Autonomous Robotic Agents. In: W. Horn (ed.), *Proceedings of ECAI 2000*. Amsterdam: IOS Press, August 2000. pp. 764–769.

[20] Steels, L. Language as a Complex Adaptive System. In: M. Schoenauer (ed.), *Proceedings of PPSN VI*. Berlin: Springer-Verlag, September 2000. Lecture Notes in Computer Science.

[21] Steels, L., Kaplan, F. Stochasticity as a Source of Innovation in Language Games. In: C. Adami, R. Belew, H. Kitano, C. Taylor (eds.), *Proceedings of Artificial Life VI*. Cambridge (MA): MIT Press, June 1998. pp. 368–376.

[22] Steels, L., Kaplan, F. Spontaneous Lexicon Change. In: *Proceedings of COLING-ACL 1998*. Montreal: ACL, August 1998. pp. 1243–1249.

[23] Steels, L., Kaplan, F. Situated Grounded Word Semantics. In: T. Dean (ed.), *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence IJCAI'99*. San Francisco (CA): Morgan Kaufmann, 1999. pp. 862–867.

[24] Steels, L., Kaplan, F. Collective Learning and Semiotic Dynamics. In: D. Floreano, J-D. Nicoud, F. Mondada (eds.), *Advances in Artificial Life (ECAL 99)*. Berlin: Springer-Verlag, 1999. pp. 679–688. Lecture Notes in Artificial Intelligence 1674.

[25] Steels, L., Kaplan, F., McIntyre, A., Van Looveren, J. Crucial Factors in the Origins of Word-Meaning. In: A. Wray (ed.), *The Transition to Language*. Oxford (UK): Oxford University Press, 2002.

[26] Steels, L., McIntyre, A. Spatially Distributed Naming Games. *Advances in Complex Systems*, January 1999, vol. 1, no. 4.

[27] Steels, L., Oudeyer, P-Y. The Cultural Evolution of Syntactic Constraints in Phonology. In: *Proceedings of the VIIth Artificial Life Conference (Alife 7)*. MIT Press, 2000.

[28] Takáč, M. *Model prototypovej konceptualizácie sveta s komunikáciou* [A Model of Prototype-Based Conceptualization of World with Communication]. 2000. Internal technical report. Institute of Informatics, Faculty of Mathematics, Physics, and Informatics, Comenius University Bratislava.

[29] Takáč, M. *Emergencia lingvistických fenoménov v jazykových hrách* [The Emergence of Linguistic Phenomena in Language Games]. 2001. Internet Distance Education Program Cognitive Sciences. <http://math.chtf.stuba.sk/kog_vedy.htm>.

[30] Tversky, A. Features of Similarity. *Psychological Review*, 1977, vol. 84, no. 4, pp. 327–352.

[31] Vogt, P. Iterated Learning and Grounding: From Holistic to Compositional Languages. In: S. Kirby (ed.), *Language Evolution and Computation: Proceedings of the Workshop/Course at ESSLLI*. 2003, pp. 76–86.