PERGAMON

Expert Systems with Applications 19 (2000) 149–158

Expert Systems with Applications

www.elsevier.com/locate/eswa

A categorical expert system "Jurassic"

A. Wichert*

Department of Neural Information Processing, University of Ulm, D-89069 Ulm, Germany

Abstract

An expert system "Jurassic" from the field of paleontology for the determination of a dinosaur species is presented. It helps the paleontologist to determine creatures from uncertain knowledge. The system is composed of 423 rules arranged in a directed acyclic graph of a depth of five. This knowledge is represented by a taxonomical arrangement of verbal categories represented by associative memories. Categorical representation is psychologically motivated and also offers an explanation of how to deal with uncertain knowledge. It is an alternative to other well known uncertainty calculi. During the categorization the model learns to favor these categories which often lead to a successful goal. This may help to speed up the search. Experiments with the availability heuristic in which parts of the knowledge base are primed or forgotten are performed.

Keywords: Availability heuristic; Hierarchical categorization; Neural networks; Semantic networks; Uncertain knowledge

1. Introduction

An expert system "Jurassic" from the field of palentology for the determination of a dinosaur species is presented. It helps the paleontologist to determine creatures from uncertain knowledge by hierarchical categorization. The goal of the hierarchical categorization is the accurate determination of a category or some categories out of present facts. This categorically motivated uncertainty calculus is an alternative to the other well known uncertainty calculuses (Duda, Gaschnig & Hart, 1979; Shafer, 1976; Shortliffe & Buchanan, 1975; Zadeh, 1975), for summary see (Lucas & van der Gaag, 1991), or other approaches (Sun, 1995).

2. Verbal categories

2.1. Feature approach

Objects can be described by a set of discrete features, such as red, round and sweet (McClelland & Rumelhart, 1985; Tversky, 1977). The similarity between them can be defined as a function of the features they have in common (Osherson, 1995; Sun, 1995). An object is judged to belong to a verbal category to the extent that its features are predicted by the verbal category (Osherson, 1987). The similarity of a category Ca and of a feature set B is given by the following formula, which is inspired by the contrast model of Tversky (Opwis & Plötzner, 1996; Smith, 1995; Tversky, 1977),

$$Simc(Ca, B) = \frac{|Ca \cap B|}{|Ca|} - \frac{|(Ca - (Ca \cap B)|}{|Ca|}$$
$$= \frac{2}{|Ca|} \cdot |Ca \cap B| - 1 \in [-1, 1]$$

$|Ca|$ is the number of the prototypical features that define the category Ca. The present features are counted and normalized so that the value can be compared. For example, the category **bird** is defined by the following features: flies, sings, lays eggs, nests in trees, eats insects. The category **bat** is defined by the following features: flies, gives milk, eats insects. The following features are present: flies and gives milk.

$$Simc(\textbf{bird},\ present\ features) = \tfrac{1}{5} - \tfrac{4}{5} = \tfrac{2}{5} \cdot 1 - 1 = -\tfrac{3}{5}$$

$$Simc(\textbf{bat},\ present\ features) = \tfrac{2}{3} - \tfrac{1}{3} = \tfrac{2}{3} \cdot 2 - 1 = \tfrac{1}{3}$$

In binary vector notation c is the number of the ones in the vector $\vec{C}$ which describes the category Ca, and $\vec{f}$ the vector which describes the present features.

$$qc^f(Ca) := Simc(\vec{C}, \vec{f}) = \frac{2}{c} \cdot \sum_{j=1}^{n} C_j f_j - 1$$

This is the quality criterion of the category Ca for the given feature set f represents the rating in the presence of a category, with $qc^f(Ca) = 1$ an absolute rating and $qc^f(Ca) = -1$ no rating at all.

* Tel.: +49-731-502-4257; fax: +49-731-502-4156.
E-mail address: wichert@neuro.informatik.uni-ulm.de (A. Wichert).

0957-4174/00/$ - see front matter
PII: S0957-4174(00)00029-4

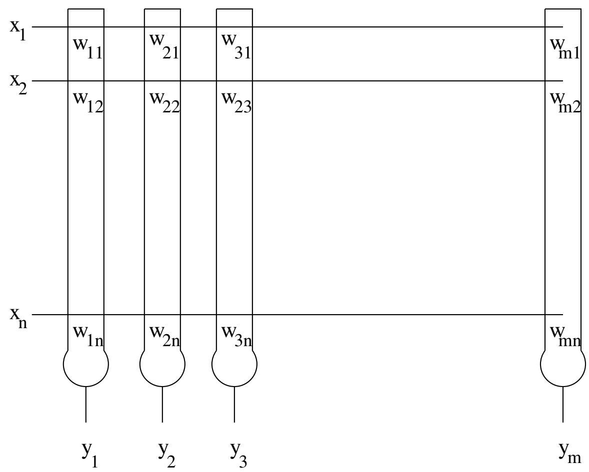

Fig. 1. The associative memory is composed of a cluster of units (Palm, 1990; Palm, Schwenker & Bibbig, 1992).

2.2. Uncertainty and salience

2.2.1. Salience of a feature

Features that discriminate among relevant facts should have a higher salience than those that do not (Smith, 1995). The features of an equal salience have a unary representation, they can only be represented as existent or non-existent. A category that is described as a set of features can be present with different grades of vagueness corresponding to the cardinal number of the set. A set of features that describes a category can be sometimes divided into subsets that represent some subcategories. Each feature can be also regarded as a kind of subcategory. If this subcategory cannot be described by other features, but, nevertheless, should have the properties of variable salience and vagueness, it is described by invisible features. To each feature a number of invisible features is assigned dependent on its salience.

2.2.2. Uncertainty of a category

An observer determines the presence of features corresponding to each category. However, if it is not possible to observe some features, the quality criterion can never reach the maximal value 1. If some features exist for a complete definition of a category, but because of the lack of available knowledge they cannot be named, they cannot be verified. These features are called the unobservable features. The certainty of the definition of a category is expressed by their number.

The number of invisible features for each feature and the number of unobservable features for each category are determined by an observer who specifies the category.

Example An example of two old sayings from country folklore:

1. If it is April and it snows much then probably the apple harvest will be bad.
2. If it is April and it rains a lot and it is very cold then the vintage will be good.

The number of invisible features as determined by the observer:

- **April** is represented by one invisible feature, as it can be present or absent.
- **Snows much** is described by two invisible features because the observer thinks that it has a higher salience than **April**. It can be either present, maybe present, or absent.
- The observer thinks that **rains a lot** has the same salience as **snows much**.
- The observer thinks that **very cold** has the greatest salience, as it is described by three invisible features. It can be either present, maybe present, maybe absent or absent.The number of unobservable features as determined by the observer:
- The observer thinks that the uncertainty of the first category which corresponds to the adjective *probably* is expressed by two unobservable features.

2.3. Representation by associative memory

2.3.1. Associative memory

The associative memory (Steinbuch, 1961, 1971) is composed of a cluster of units which represent a simple model of a real biological neuron. The unit is composed of weights which correspond to the synapses and dentrides in the real neuron. They are described by w_{ij} in Fig. 1.

The patterns are represented by binary vectors. The presence of a feature is indicated by a “one” component of the vector, its absence through a “zero” component of the vector. Two pairs of these vectors are always associated and this process of association is called learning. The first of the two vectors is called the question vector and the second, the answer vector. After learning, the question vector is presented to the associative memory and the answer vector difference is determined.

In the initialization phase of the associative memory no information is stored. Because the information is represented in the weights, they are all initially set to zero. In the learning phase, binary vector pairs are associated. Let $\vec{x}$ be the question vector and $\vec{y}$ the answer vector, so that the learning rule is:

$$w_{ij}^{\text{new}} = 1 \qquad \text{if } y_i \cdot x_j = 1$$
$$w_{ij}^{\text{new}} = w_{ij}^{\text{old}} \qquad \text{else}$$

In the retrieval phase of the associative memory, a fault tolerant answering mechanism determines qc^f for each category:

$$qc(Ca_i) = y_i = \frac{2}{\sum_{j=1}^{n} \delta(w_{ij})} \cdot \sum_{j=1}^{n} \delta(w_{ij} x_j) - 1$$

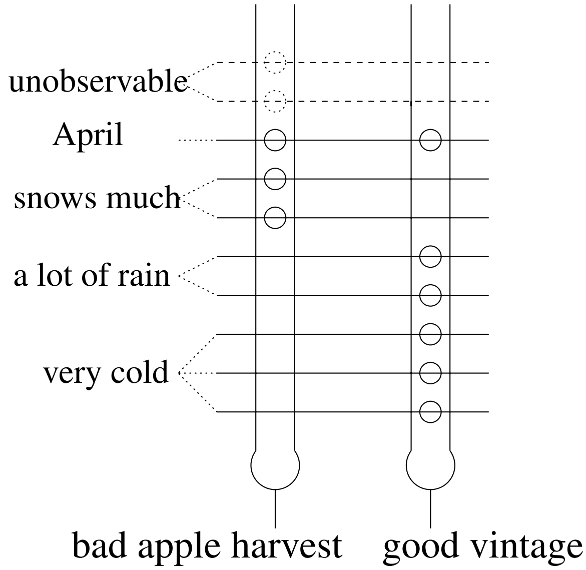

Fig. 2. Neural representation of two country sayings.

The set of features can be represented by a binary vector in which the positions represent different features. For each category a binary vector can be defined. A "one" at a certain position corresponds to a certain invisible or unobservable feature. Those vectors can be stored in an associative memory. The two country sayings are represented by an associative memory composed of two units (see Fig. 2).

2.3.2. Table of belief values

Salience of a feature is represented by the number of invisible features determined by an observer who specifies the category.

Belief in the presence of a feature is represented by the number of present invisible features determined by an observer who utilizes the category.

To facilitate the specification of the belief value in the presence of a feature a belief table is used in which different belief values are named by adjectives. The observer describes the present features with the corresponding adjectives from the table. Different tables can be specified depending on the observer and the task (Table 1 or Table 2). A chosen belief value from the table is converted to the corresponding belief value of the feature by

$$belief = \left\lfloor salience \cdot \frac{table\ belief}{table\ salience} \right\rfloor.$$

Table 1
The observer who utilizes the category can chose between three different belief values. The table salience is two

	Table belief value
Present	2
Maybe	1
Absent	0

Table 2
The observer who utilizes the category can chose between seven different belief values. The table salience is six

	Table belief value
Certainly	6
Very probably	5
Probably	4
Maybe	3
Probably not	2
Very probably not	1
Certainly not	0

Nothing is known about unspecified features. The probability of their presence is fifty percent or less depending on their salience,

$$belief\ in\ the\ not\ specified = \left\lfloor \frac{salience}{2} \right\rfloor.$$

3. Deduction systems

Problems without side effects of actions can be described by deduction systems which are a subgroup of production systems (Winston, 1992). In deduction systems the premise specifies combinations of assertions, by which a new assertion of the conclusion is directly deduced. This new assertion is added to the working memory. Deduction systems do not need strategies for conflict resolution because every rule presumably produces reasonable assertions and there is no harm in firing all triggered rules. Deduction systems may chain together rules in a forward direction, from assertions to conclusions, or backward from hypotheses to premises. During backward chaining it is ensured that all features are properly focused. Backward chaining is used if no features are present. If all features are given, forward chaining is used to prevent wasting of time pursuing hypotheses which are not specified by the features. The chained rules describe the complete problem space which can be represented by a semantic net (Quillian, 1968; Shastri, 1988). For clarity, rules can be arranged in groups (Aikins, 1986; Kahn, Kepner & Pepper, 1987) which define a taxonomy according to their dependence between the conclusion pattern and one premise pattern (see Fig. 3).

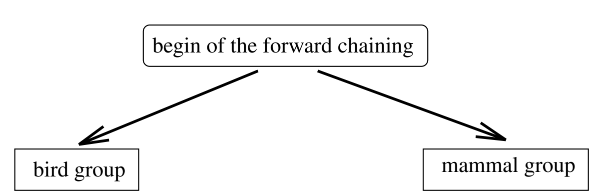

Fig. 3. Semantic net representing the taxonomy of the rules about birds and mammals.

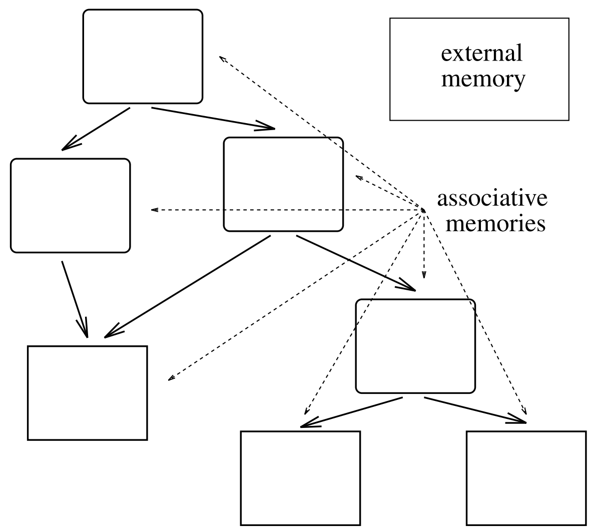

Fig. 4. Neural deductions system.

begin of the forward chaining

- If (flies(x) ∨ feathers(x)) ∧ lays eggs(x) then bird(x)
- If gives milk(x) then mammal(x)

bird group

- If bird(x) ∧ swims(x) then penguin(x)
- If bird(x) ∧ sings(x) then nightingale(x)

mammal group

- If mammal(x) ∧ long neck(x) ∧ four legs(x) then giraffe(x)
- If mammal(x) ∧ swims(x) then dolphin(x)

Uncertainty can be expressed by the certainty theory. The heuristic function can be represented by the *CF* values of the conclusion patterns in the case when conflict resolution is used. Deduction systems are often used by expert systems which were developed with the engineering aim to capture the knowledge of an expert in a certain closed domain (Jackson, 1999; Lucas & van der Gaag, 1991; Samkian, 1992). The type of knowledge captured by these expert systems was often organized by human society in different kinds of taxonomies, for example taxonomy of animals, or the taxonomy of medical illness.

The semantic net representation of the taxonomy of rules is related to the decision tree approach (Quinlan, 1986). A decision tree is used to determine a category by a function that maps each element of its domain to a category label or numerical value. At each leaf of a decision tree, one finds a category label. It is determined at nodes by tests that have a small number of possible outcomes. A decision tree with a range of symbolic category or class labels is called a classification tree. In our approach, the tests are described by rules. A decision tree with

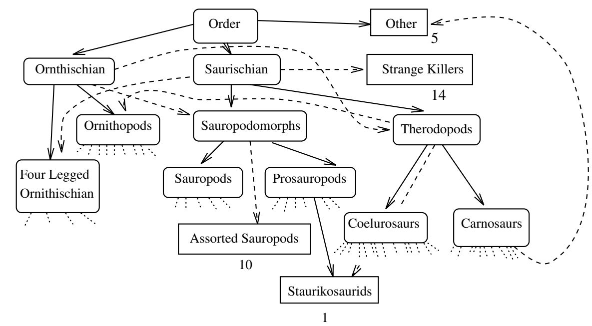

Fig. 5. Taxonomy of dinosauria. The number written below the rectangular boxes represents the categories which are not divided, the species. Uncertain categories are represented by dotted arrows.

a range of continuous numeric values is called a regression tree. Decision trees show clearly how to reach a decision. They are constructed automatically by C4.5 (Quinlan, 1993) and CART (Classification and Regression Tree) (Breiman, Friedman, Olshen & Stone, 1984) programs. Decision trees can be also converted into rules by the C4.5 program.

There is also a connection between neural networks and decision trees. Feed forward neural networks can be initialized by decision trees (Ivanova & Kubat, 1995). In this case decision trees represent some approximation of the target concept which helps to determine the initial weight setting and the architecture of the feed-forward neural network. The architecture is determined by the number of neurons and the topology of connections between them. This kind of initialization was proven successful in different real world applications (Kubat, Holte & Matwin, 1998a; Kubat, Koprinska & Pfurtscheller, 1998b).

4. Hierarchical categorization

4.1. Neural deductions system

Hierarchical categorization is performed by a neural deductive system (Wichert, 1998; Wichert & Kestler, 1998). The model is composed of connected associative memories which represent groups of rules. The problem space is represented by connections between the associative memories and those connections correspond to conclusions of the rules. The known features are represented in an external memory before the forward chaining begins (see Fig. 4). This model is appropriate for taxonomical representation of knowledge with the aid of verbal categories. Categories can be divided into subcategories, so that a taxonomy can be constructed and represented by an acyclic graph. The nodes in this graph correspond to categories and the links indicate the “is a subcategory” relation between them. The process of the hierarchical categorization can be performed by moving from more general categories to more specific categories

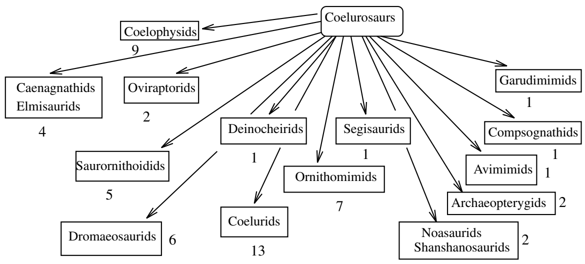

Fig. 6. Infraorder Coelurosaurs.

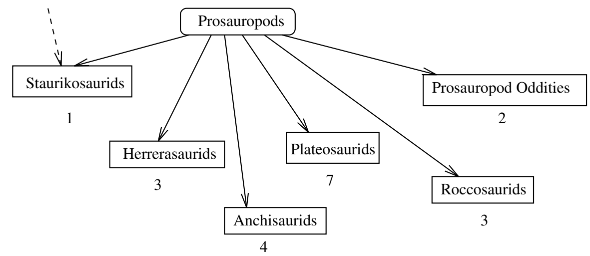

Fig. 8. Infraorder Prosauropods.

until the desired categories are reached. The rules in each associative memory are represented as described in Section 3 with the exception that the categories which are represented by the conclusion may be divided into more exact subcategories by another associative memory (see Fig. 4). The associative memory with the name of the corresponding category which it divided is called a module. A disjunction of patterns in a premise can be represented by two different rules. Negation cannot be interpreted, only represented by a corresponding feature name.

The model is composed of connected associative memories which represent groups of rules. The problem space is represented by connections between the associative memories and those connections correspond to conclusions of the rules. The known features are represented in an external memory before the forward chaining begins (see Fig. 4). This model is appropriate for taxonomical representation of knowledge with the aid of verbal categories. Categories can be divided into subcategories, so that a taxonomy can be constructed and represented by an acyclic graph. The nodes in this graph correspond to categories and the links indicate the “is a subcategory” relation between them. The process of the hierarchical categorization can be performed by moving from more general categories to more specific categories until the desired categories are reached. The associative memory with the name of the corresponding category which it divided is called a module.

4.2. System Jurassic

The goal of the system is to help paleontologists determine creatures or creature groups out of a taxonomic knowledge base which describes the dinosaurs (Haines, 1999) and using only some vague beliefs about the presence and absence of some features. In 1887 Prof. Harry Govier Seeley grouped all dinosaurs into the saurischia and ornithischia groups according to their hip design. The saurischian were divided later into two subgroups: the carnivorous, bipedal theropods and the plant-eating, mostly quadruped sauropodomorphs. The ornithischians were divided into the subgroups birdlike ornithopods, armored thyreophorans, and margginoncephalia. The subgroups can be divided into suborders and then into families and finally into genus. The genus includes the species. It must be noted that in this taxonomy many relations are only guesswork, and many paleontologists have different ideas about how the taxonomy should look. The whole knowledge base is composed of 70 modules in which 423 rules are stored. The taxonomy is modeled in the books of (Lambert, 1983, 1993) in which over fifty families and more than 340 genre of dinosaurs are described (Figs. 5–11).

4.3. Categorization

The external memory guides the search in the problem space in the direction of the most plausible category, which is the category with the highest quality criterion value. In the search the first favored category can turn out to be wrong because it is assumed that its subcategories are not present. In this case another category is examined. This search strategy corresponds to hill climbing (Winston, 1992), which is a depth-first search in which the choices are ordered according to the quality criteria values. The quality criteria values represent a heuristic measurement of the distance to the remaining goal.

In the following example the retrieval of taxonomic knowledge from a database, which tries to completely describe a knowledge area, is shown. The species of Fig. 12 is described by six features using the belief table.

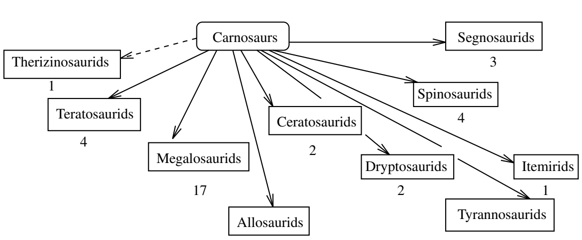

Fig. 7. Infraorder Carnosaurus.

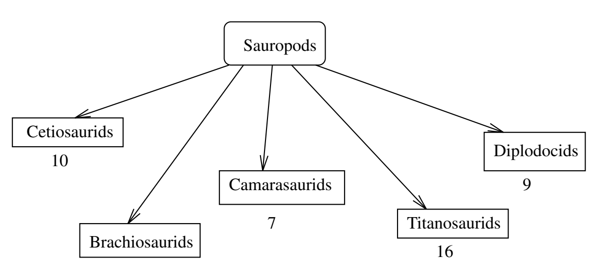

Fig. 9. Infraorder Sauropods.

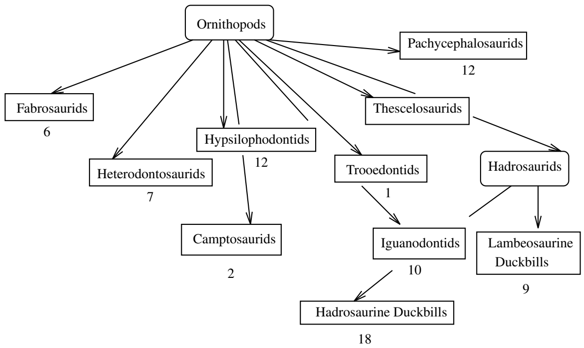

Fig. 10. Infraorder Ornithopods.

1. probably **bird hipped**
2. certainly **two legged**
3. probably **long arms**
4. certainly **long stiffly held tail**
5. probably **solid bony humps or crests**
6. very probably **length nine m**

*

(1) ! MODULE ORNITHISCHIAN with $qc = 0.33$, $\langle 1 \rangle$! $*$

(2) ! MODULE ORNITHOPODS with $qc = 0.67$, $\langle 2 \rangle$! $* *$

(3) ! MODULE PACHYCEPHALOSAURIDS with $qc = 0.51$, $\langle 3 \rangle$!

(4) ! MODULE THESCELOSAURIDS with $qc = 0.49$, $\langle 3 \rangle$!

(5) ! MODULE HADROSAURIDS with $qc = 0.48$, $\langle 3 \rangle$! $*$

(6) ! MODULE HADROSAURINE_DUCKBILLS with $qc = 0.36$, $\langle 4 \rangle$! $*$

RESULT:

MAIASAURA with $qc = 0.38$
SAUROLOPHUS with $qc = 0.33$

The first determined category represents *Maiasaura*

4.4. Hypothesis

Which object of one group is most similar to an object of another different group? We wish to know which animal of the dinosaur group is most similar to an animal of the mammal group. This is a kind of case-based reasoning

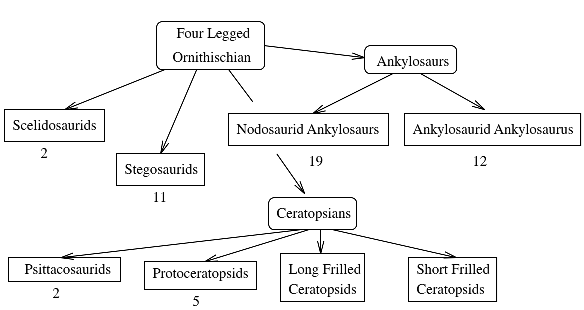

Fig. 11. Infraorder four legged Ornithishian.

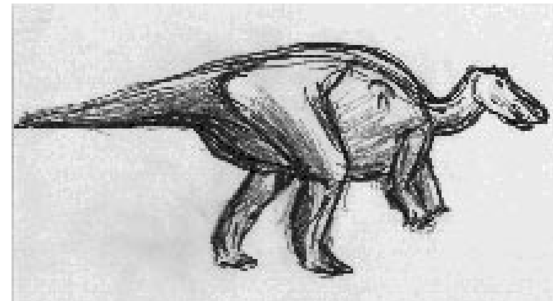

Fig. 12. Maiasaura.

(Hammond, 1989) in which specific knowledge is used to retrieve the most similar stored case. "Which dinosaur species is most similar to a human?" We describe a human being by seven features. The hierarchical categorization is performed in which the most similar stored category is determined.

1. certainly **two legged**
2. certainly not **four legged**
3. probably thin walled fragile bones
4. very probably not **big**
5. very probably **big eyes**
6. certainly **big brain**
7. certainly capsule in the skull
8. very probably **length two m**

*

(1) ! MODULE SAURISCHIAN with $qc = -0.33$, $\langle 1 \rangle$! $* *$

(2) ! MODULE THEROPODS with $qc = 0.05$, $\langle 2 \rangle$! $*$

(3) ! MODULE COELUROSAURS with $qc = 0.12$, $\langle 3 \rangle$! $* * *$

(4) ! MODULE SAURORNITHOIDIDS with $qc = 0.2$, $\langle 4 \rangle$! $*$

RESULT:

SAURORNITHOIDES with $qc = 0.25$
STENONYCHOSAURUS with $qc = 0.25$

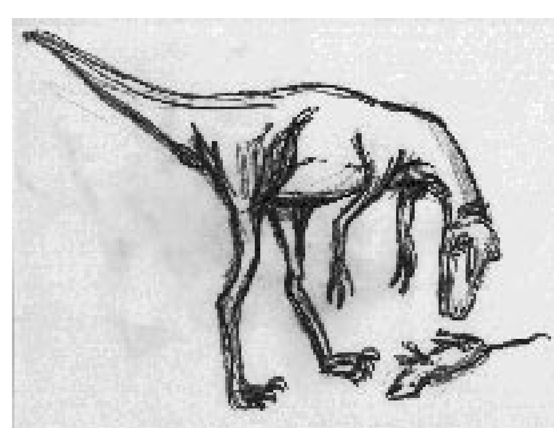

Fig. 13. Stenonychosaurus.

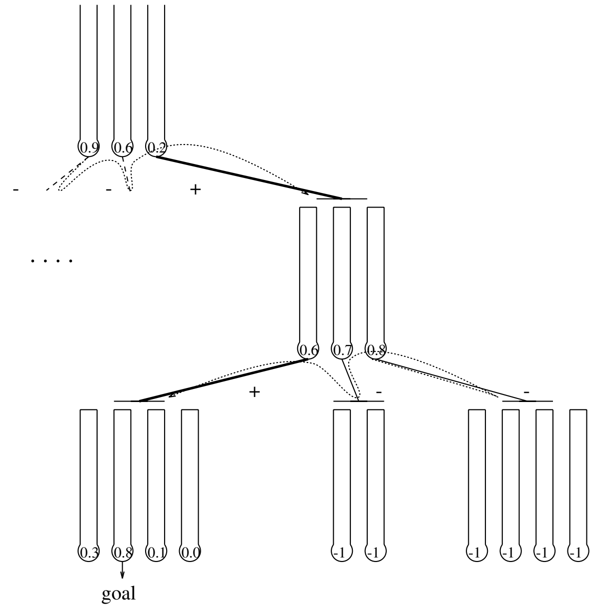

Fig. 14. The learning of the availability heuristic. Links between the categories which describe a successful search are strengthened by a small factor, shown by $+$. Links to categories where backtracking occurs are weakened by a small factor, shown by $-$.

This answer is the same as the suggestion of Dale Russell (Russell & Sequin, 1982) in the early 1980s that the Stenonychosaurus, see Fig. 13 (also now known as Troodon), could have given rise to a brainy descendant, had dinosaurs survived instead of dying out (Lambert, 1983, 1993).

4.5. Priming

During the hierarchical categorization by humans some categories come to mind more easily, because they were determined to be more frequent than the other. Some psychologists (Tversky & Kahneman, 1973) assume that the individual estimates the frequency of an event, in our case the determination of a category. This kind of heuristic is called the availability heuristic (Schwartz, 1995; Tversky & Kahneman, 1973; Wickelgren, 1977). In our model the availability heuristic corresponds to the strengthening of the links between the categories which describe a successful search by a small factor. The successful search corresponds to the path of links between the categories from the category where the search began to the results of the hierarchical categorization (see Fig. 14. Paths that corresponds to wrong search directions are weakened by a small factor. During the repeated search the categories that receive strong links are favored. The search direction corresponds to the highest value that is the sum of the quality criterion of a category and the strength of the link to it. In terms of Bayesian statistics the chosen category has the highest posterior probability which is composed of actual likelihood and prior probability (Lucas & van der Gaag, 1991; Pearl, 1989).

A paleontologist tries to determine a species from vague knowledge by forming possible hypotheses. He specifies the most probable species by declarative knowledge in the form of some of the most noticeable features and the species is determined. At the same time the corresponding taxonomic knowledge area is primed (Anderson, 1995; Neely, 1991; Squire & Kandel, 1999) by non-declarative knowledge with the aid of the availability heuristic. After this the paleontologist formulates more sophisticated features and the species of the primed taxonomic area are examined first. Thus, priming eases the formation of the final hypothesis, as more exact, possible hypotheses are formed.

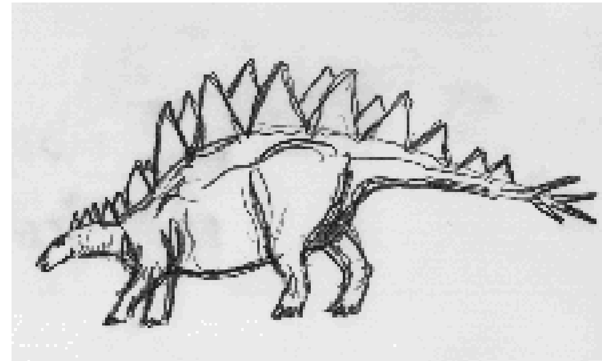

Fig. 15. Stegosaurus.

The roof lizard of Fig. 15 is described by three features, it is a stegosaurid with the largest known plates:

1. probably **bird hipped**
2. certainly **four heavy spikes**
3. certainly **plates up to seventy six cm high and seventy nine cm long**

*

(1) ! `MODULE ORNITHISCHIAN` with $qc = 0.33$, $\langle 1 \rangle$!

(2) ! `MODULE ORNITHOPODS` with $qc = 0$, $\langle 2 \rangle$!

(3) ! `MODULE THESCELOSAURIDS` with $qc = -0.02$, $\langle 3 \rangle$!

⋮

(13) ! `MODULE IGUANODONTIDS` with $qc = -0.05$, $\langle 3 \rangle$!

(14) ! `MODULE FOUR_LEGGED_ORNITHISCHIANS` with $qc = 0$, $\langle 2 \rangle$!

(15) ! `MODULE CERATOPSIANS` with $qc = -0.02$, $\langle 3 \rangle$!

⋮

(19) ! `MODULE LONG_FRILLED_CERATOPSIDS` with $qc = -0.07$, $\langle 4 \rangle$!

(20) ! `MODULE STEGOSAURIDS` with $qc = -0.05$, $\langle 3 \rangle$! * *

`RESULT:`

STEGOSAURUS with $qc = 0.1$

The search took twenty steps.

Species with triangular plates which lived in the late Jurassic are searched, the primed taxonomic area is

Table 3
Required steps before S and after learning S^L Ornithopods

Suborder	Family	Species	S	S^L
Ornithopods (α)	Fabrosaurids	*azendohsaurus*	33	5
	Heterodontosaurids	*abrictosaurus*	38	6
	Hypsilophodontids	*alocodon*	32	4
	Trooedontids	*trooedon*	39	3
	Thescelosaurids	*thescelosaurus*	30	3
	Lamb. Duckbills	*hypacrosaurus*	36	11
Four legged	Scelidosaurids	*scelidosaurus*	67	24
	Protoceratopsids	*leptoceratops*	60	17
	Ankylo. Ankylosaurs	*amtosaurus*	66	23
Sauropods (β)	Cetiosaurids	*austrosaurus*	56	36
	Titanosaurids	*alamosaurus*	53	37
	Diplodocids	*barosaurus*	50	34
	Brachiosaurids	*astrodon*	54	38
Prosuaropods	Anchisaurids	*anchisaurus*	46	30
	Roccosaurids	*riojasaurus*	48	32
	Staurikosaurids	*staurikosaurids*	41	29
	Prosaur. Oddities	*mussaurus*	44	27
	Herrerasaurids	*herrerasaurus*	45	28
Other		Marine lizard	27	65
Strange kill.		*marshosaurus*	56	68
Coelusaurus (γ)	Coelurids	*microvenator*	14	52
	Noa. Shanshanosaurids	*noasaurus*	15	53
	Segisaurids	*segisaurus*	6	44
	Avimimids	*avimimus*	17	55
Carnosaurus	Megalosaurids	*dilophosaurus*	24	62
	Allosaurids	*allosaurus*	25	63
	Tyrannosaurids	*albertosaurus*	23	61

examined first:

1. very probably not **lizard hipped**
2. certainly **late jurassic**
3. certainly **triangular plates**

*
(1) ! `MODULE ORNITHISCHIAN` with $qc = -0.33$, $\langle 1 \rangle$!

Table 4
Mean and standard deviation of required steps before and after learning α for α, β and γ and their combinations. The p values were determined by paired sample t test. Significant for $p < 0.05$ by convention

	S	S^L	S	S^L
Steps for α			*Steps for* $\alpha + \beta$	
Mean	44.56	10.67	46.56	21.5
Sdev	15.26	8.56	11.19	12.91
p	–	4.32^{-7}	–	6.9^{-9}
Steps for β			*Steps for* $\alpha + \gamma$	
Mean	48.56	32.33	33.78	34.39
Sdev	5.05	4.09	18.03	25.65
p	–	5.24^{-9}	–	0.47
Steps for γ			*Steps for* $\alpha + \beta + \gamma$	
Mean	23	58.11	38.7	33.7
Sdev	14.05	7.62	16.45	20.88
p	–	9.73^{-7}	–	0.2

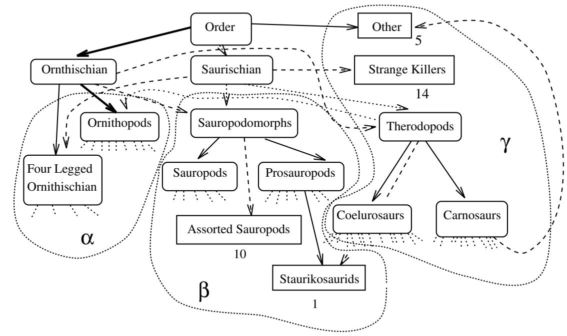

Fig. 16. Taxonomy of dinosauria after learning Ornithopods.

(2) ! `MODULE FOUR_LEGGED_ORNITHISCHIANS` with $qc = -0.33$, $\langle 2 \rangle$!
(3) ! `MODULE STEGOSAURIDS` with $qc = -0.27$, $\langle 3 \rangle$! * *
\---

`RESULT:`

KENTROSAURUS with $qc = -0.04$
TUOJIANGOSAURUS with $qc = -0.04$
DACENTRURUS with $qc = -0.12$

Dacentrurus has the lowest qc value, it lived in the middle to late Jurassic and had perhaps no plates. The search took only three steps.

4.6. Emphasis and forgetting

A medical doctor diagnosing diseases during an influenza epidemic can incorrectly diagnose malaria as influenza. This is because the medical knowledge area corresponding to the influenza disease is primed and the primed area is preferred. Symptoms which are common to both of the two diseases could indicate, to a careless doctor at first glance, a case of influenza. This human behavior of emphasizing knowledge areas which are often used and disregarding knowledge areas which are seldom used is modeled in the next experiment. Mostly this policy is useful, if one has to balance between completeness and speediness of retrieved knowledge.

The taxonomy can be divided into three groups α, β, γ (see Table 4). Each group is represented by nine randomly chosen species (see Table 3). Each species is described by two certain features which are sufficient for the categorization but insufficient to guide the search. The number of required steps S for the categorization is determined (see Table 3). The uniformed search strategy randomly prefers the group γ to group α significantly.[1] After learning the group *ornitopods* by frequent determination of five species with sufficient knowledge to guide the search with five certain features, the categorization is repeated with equal

[1] Mean difference test rejects the hypothesis that the mean difference is zero between α and γ with $p = 0.0033$.

features as before learning. The number of required steps S^L for the categorization is determined (see Table 3). There is a significant improvement for the groups α,β, and a significant deterioration for the group γ (see Table 4). Emphasis of one knowledge area leads to forgetting of another knowledge area. There is no significant change in the behavior of the whole taxonomy before and after learning, or the groups α and γ together, as they compensate for each other (see Table 4). The significance of the preference of group α to γ is very high (Fig. 16).[2]

5. Conclusions

A neural model for a deduction system based on the assembly theory was introduced. It was shown that hierarchical categorization can be performed by neural networks efficiently. The categorical representation offers an alternative to the traditional uncertainty calculus (Duda et al., 1979; Luger & Stubblefield, 1998; Shafer, 1976; Shortliffe & Buchanan, 1975; Zadeh, 1975).

Similarity is used when uncertain knowledge is represented without the need of an additional calculus. In addition, belief tables allow additionally the detachment of the uncertainty of the coded knowledge and of the actually present knowledge. The availability heuristic offers a combination of the frequency with the actual likelihood of the presence of a category. Priming eases the formation of the final hypothesis, as more exact, possible hypotheses are formed. This human behavior of emphasis of knowledge areas which are often used and disregarding of knowledge areas which are seldom used was modeled. This policy is useful, if one has to balance between completeness and speediness of retrieved knowledge.

References

Aikins, J. (1986). Prototypical knowledge for expert systems. *Artificial Intelligence*, *20*, 163–210.

Anderson, J. R. (1995). *Cognitive psychology and its implications*. (4th ed.) New York: W.H. Freeman.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*, Belmont, CA: Wadsworth.

Duda, R. O, Gaschnig, J., & Hart, P. E. (1979). Model design in the prospector consultant program for mineral exploration. In D. Michie, *Expert systems in the microelectronic age*. Edinburgh University Press.

Haines, T. (1999). *Walking with dinosaurs—a natural history*. BBC Worldwide Limited.

Hammond, K. (1989). *Case-based planning: viewing planning as a memory task*, New York: Academic Press.

Ivanova, I., & Kubat, M. (1995). Initialization of neural networks by means of decision trees. *Knowledge Based Systems*, *8*, 333–344.

Jackson, p. (1999). *Introduction to expert systems*. (3rd ed.). Reading, MA: Addison-Wesley.

Kahn, G., Kepner, A. & Pepper, J. (1987). Test: a model-driven application shell. In: *National Conference on Artificial Intelligence*, (pp. 814–818).

Kubat, M., Holte, R., & Matwin, S. (1998a). Machine learning for the detection of oil spills in satellite images. *Machine Learning*, *30*, 195–215.

Kubat, M., Koprinska, I., & Pfurtscheller, G. (1998). Learning to classify biomedical signals. In R. Michalski, I. Bratko & M. Kubat, *Machine learning and data mining: methods and applications*. New York: Wiley.

Lambert, D., (1983). *Collins guide to dinosaurs*. Diagram Visual Information Ltd.

Lambert, D. (1993). *The ultimate dinosaur book*, Dorling Kindersley.

Lucas, P., & van der Gaag, L. (1991). *Principles of expert systems*, Reading, MA: Addison-Wesley.

Luger, G. F., & Stubblefield, W. A. (1998). *Artificial intelligence, structures and strategies for complex problem solving*. (3rd ed.) Reading, MA: Addison-Wesley.

McClelland, J., & Rumelhart, D. (1985). Distributed memory and the representation of general and specific memory. *Journal of Experimental Psychology: General*, *114*, 159–188.

Neely, J. (1991). Semantic priming effects in visual word recognition: a selective review of current findings and theories. In D. Besner & G. Humphreys, *Basic processes in reading. Visual word recognition*. Hillsdale, NJ: Erlbaum.

Opwis, K., & Plötzner, R. (1996). *Kognitive Psychologie mit dem Computer*, Heidelberg/Berlin/Oxford: Spektrum Akademischer.

Osherson, D. N. (1987). New axioms for the contrast model of similarity. *Journal of Mathematical Psychology*, *31*, 93–103.

Osherson, D. N. (1995). Probability judgment. In E. E. Smith & D. N. Osherson, (2nd ed.). *Thinking* (pp. 35–75), Vol. 3. Cambridge, MA: MIT Press (chap. 2).

Palm, G. (1990). *Gehirn und Kognition*. Spektrum der Wissenschaft (pp. 164–174).

Palm, G., Schswenker, F., & Bibbig, A. (1992). *Theorie Neuronaler Netze. 1. Skript zur Vorlesung*. University of Ulm, Department of Neural Information Processing.

Pearl, J. (1989). *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Palo Alto, CA: Morgan Kaufmann.

Quillian, R. (1968). Semantic memory. In M. Minsky, *Semantic information processing* (pp. 227–270). Cambridge, MA: MIT Press.

Quinlan, J. (1986). Introduction of decision trees. *Machine Learning*, *1* (1), 81–106.

Quinlan, J. (1993). *C4.5: programs for machine learning*, San Francisco, CA: Morgan kaufmann.

Russell, D., & Sequin, R. (1982). Reconstruction of the small cretaceous theropod stenonychosaurus inequalis and a hypothetical dinosauroid. *Syllogeous*, *37*, 1–43.

Samkian, E. (1992). *VV-XPS, ein Expertensystem zur Ermittlung absetzbarer Werbungskosten und Sonderausgaben bei Einküften aus Vermitung und Verpachtung*. Master's thesis, Universität des Saarlandes, Saarbrücken, Germany.

Schwartz, N. (1995). Social cognition: information accessibility and use in social judgment. In E. E. Smith & D. N. Osherson, (2nd ed.). *Thinking*, Vol. 3. Cambridge, MA: MIT Press (chap. 10).

Shafer, G. (1976). *A mathematical theory of evidence*, Princeton, NJ: Princeton University Press.

Shastri, L. (1988). *Semantic networks: an evidential formulation and its connectionistic realization*, London: Morgan Kaufmann.

Shortliffe, E. H., & Buchanan, B. G. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences*, *23*, 351–379.

Smith, E. E (1995). Concepts and categorization. In E. E Smith & D. N. Osherson, (2nd ed.). *Thinking* (pp. 3–33), Vol. 3. Cambridge, MA: MIT Press (chap. 1).

Squire, L. R., & Kandel, E. R. (1999). *Memory. From mind to molecules*, Scientific American Library.

Steinbuch, K. (1961). Die Lernmatrix. *Kybernetik*, *1*, 36–45.

Steinbuch, K. (1971). *Automat und Mensch*. (4th ed.). Berlin: Springer.

Sun, R. (1995). A two-level hybrid architecture for structuring knowledge for commonsense reasoning. In R. Sun & L. A. Bookman,

[2] Mean difference test rejects the hypothesis that the mean difference is zero is even more significant, $p = 7.33^{-10}$.

Computational architectures integrating neural and symbolic processing (pp. 182–247). Dordrecht: Kluwer Academic (chap. 8).

Tversky, A. (1977). Feature of similarity. *Psychological Review*, *84*, 327–352.

Tversky, A., & Kahneman (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

Wichert, A. (1998). Hierarchical categorization. In M. W. Evens, *Ninth midwest artificial intelligence and cognitive science conference* (pp. 141–148). AAAI Press.

Wichert, A., & Kestler, H. A. (1998). A categorical based reanimation expert system. In E. Ifeachor, A. Sperduti & A. Starita, *Neural networks and expert systems in medicine and healthcare*. Singapore: World Scientific.

Wickelgren, W. A. (1977). *Cognitive psychology*, Englewood Cliffs, NJ: Prentice-Hall.

Winston, W. A. (1992). *Artificial intelligence*. (3rd ed.) Reading, MA: Addison-Wesley.

Zadeh, L. A. (1975). Fuzzy logic and approximate reasoning. *Synthese*, *30*, 407–428.