Concept Formation and Learning: A Cognitive Informatics Perspective

Y.Y. Yao

Department of Computer Science
University of Regina
Regina, Saskatchewan, Canada S4S 0A2
E-mail: yyao@cs.uregina.ca
URL: http://www.cs.uregina/∼yyao

Abstract

Concepts are the basic units of thought that underlie human intelligence and communication. From the perspective of cognitive informatics, a layered framework is suggested for concept formation and learning. It combines cognitive science and machine learning approaches. The philosophical issues and various views of concepts are reviewed. Concept learning methods are presented based on the classical view of concepts.

1. Introduction

Cognitive science is the study of intelligence and its computational processes in humans and animals, in computers, and in the abstract [18]. Informatics deals with gathering, storing, retrieving, manipulating, processing and interpreting recorded information. The combination of cognitive science and informatics leads to an emerging, multidisciplinary research area known as cognitive informatics [13, 24]. There is a fast growing interest in this new research initiative.

Wang [23] presented a general framework for the study of cognitive informatics. Following the definition of cognitive science given by Simon and Kaplan [18], one may view cognitive informatics as the study of information processing in humans, in computers, and in the abstract. The first two topics can be concisely described as "information processing in mind and machine", the subtitle of a book by Sowa published in 1984 [21]. Based on results from psychology, philosophy, linguistics, and neuroscience, cognitive informatics studies the internal information processing mechanisms and the natural intelligence of the brain [23]. From the point of view of computer science and particularly artificial intelligence, cognitive informatics aims at building machines and systems that simulate human ways of knowing and thinking. The findings and an in-depth understanding of human intelligence would have a significant impact on the development of the next generation technologies in informatics, computing, software, and cognitive sciences [23].

The ideas of studying intelligence in humans, in computers, and in the abstract can be used to study any particular topics of cognitive informatics. Concepts are the basic units of thought that underlie human intelligence and communication. The study of concept formation and learning is central to cognitive informatics. The objective of this paper is therefore to investigate basic issues of concept formation and learning from cognitive informatics perspectives. We propose a layered model for concept formation and learning and present two concept learning algorithms.

The study of concept formation and learning from the perspective of cognitive informatics captures both cognitive and algorithmic aspects and issues. The results suggest that in general the connections between cognitive science and informatics need to be further explored.

The rest of the paper is organized as follows. Section 2 presents a three-level framework for knowledge discovery and machine, namely, the philosophy level, the algorithm/technique level, and the application level. Sections 3 to 5 examine the issues of concept formation and learning with reference to the three levels.

2. A Layered Model for Knowledge Discovery

The notion of architectures plays an important role in the study of cognitive science, which allows the study of intelligence and its computational processes in the abstract. The fundamental design specifications of an intelligent system are referred to as its architecture [18]. The components of the architecture only abstractly represent the underlying physical structures. The architecture may be specified at different levels of abstraction. For example, one can study the architectures of digital computer at various levels. It is also possible to derive a general architecture of human cognitive system [18]. Systems based on cognitive informatics can be similarly specified. Wang *et al.* [25] presented a layered reference model of the brain, ranging from lower level functions of sensation and memory to higher level cognitive functions.

The same idea can be used in building a framework for knowledge discovery from databases, or data mining. One needs to separate the study of knowledge and the study of knowledge discovery algorithms, and in turn to separate them from the study of the utility of the discovered knowledge. A three-level framework, consisting of philosophy level, the algorithm/technique level, and the application level, has been proposed [27].

The three levels of the layered model focus on three fundamental questions. The philosophical level addresses questions about knowledge, the algorithm level concentrates on knowledge discovery methods, and the application level deals with the utility of the discovered knowledge. Their main features and functionalities are summarized below.

A. Philosophy level

The philosophy level is the study of knowledge and knowledge discovery in mind. One attempts to answer the fundamental question, namely, what is knowledge? There are many related issues to this question, such as the representation of knowledge in the brain, the expression and communication of knowledge in words and languages, the relationship between knowledge in the mind and in the external real world, and the classification and organization of knowledge [21]. One also needs to study the cognitive process in which knowledge is acquired, understood, processed, and used by humans. It is necessary to define precisely "knowledge" and the "basic unit of knowledge", which serve as the primitive notions of knowledge discovery.

The study of concepts is central to philosophy, psychology, cognitive science, inductive data processing and analysis, and inductive learning [7, 20, 21, 22]. Concepts are assumed to be basic constituents of thought and belief, and the basic units of thought and knowledge. The focus of the philosophy level is on the representation, interpretation, connection and organization of concepts, the processes of forming and learning concepts, and the the processes of reasoning with concepts.

B. Algorithm/technique level

The algorithm level is the study of knowledge and knowledge discovery in machine. One attempts to answer the question, how to discover knowledge? In the context of computers, there are many issues related to this question, such as the coding, storage, retrieval of knowledge in a computer, the implementation of human knowledge discovery methods in programming languages, and the effective use of knowledge in intelligent systems. The focus of technique level is on algorithms and methods for extracting knowledge from data.

The main stream of research in machine learning, data mining, and knowledge discovery has concentrated on the technique level. Many concept learning algorithms have been proposed and studied.

C. Application level

The ultimate goal of knowledge discovery is to effectively use the discovered knowledge. The application level therefore should focus on the notions of "usefulness" and "meaningfulness" of discovered knowledge. These notions can not be discussed in total isolation with applications, as knowledge in general is domain specific.

The notion of usefulness can be interpreted based on utility theory [5] and the notion of meaningfulness can be modeled based on explanation-oriented data mining [29, 30].

The division between the three levels is not a clear cut, and may have overlap with each other. It is expected that the results from philosophy level will provide guideline and set the stage for the algorithm and application levels. On the other hand, it is desirable that philosophical study does not depend on the availability of specific techniques, and technical study is not constrained by a particular application. The existence of a type of knowledge in data is unrelated to whether we have an algorithm to extract it. The existence of an algorithm does not necessarily imply that the discovered knowledge is meaningful and useful [27].

The three levels represent the understanding, discovery, and utilization of knowledge. Any of them is indispensable in the study of intelligence and intelligent systems. They must be considered together in a common framework through multi-disciplinary studies, rather than in isolation. Many research efforts

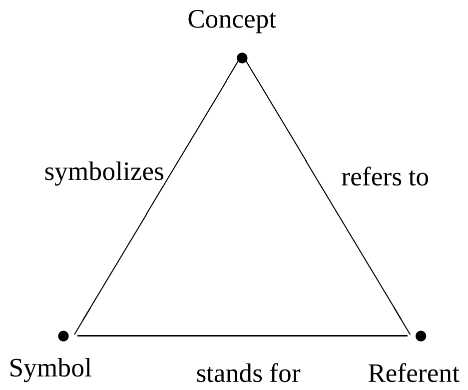

Figure 1. The meaning triangle (adopted from [21])

have been made to achieve such a goal. For example, Sowa [21] combined cognitive science and artificial intelligence approaches for modeling information processing in mind and machine based on conceptual structure. The book edited by Van Mchelen *et al.* [22] brings together the cognitive research on categories and concepts (i.e., the philosophy level study) and data analysis (i.e., the technique level study).

The general ideas of the layered framework can be immediately applied to the study of cognitive informatics. One can address different types of issues at the three levels. In the following sections, we study concept formation and learning with respect to the three levels.

3. Philosophy Level Study of Concepts

There are many theoretical views of concepts, concept formation and learning [15, 20, 21, 22]. The classical view treats concepts as entities with well-defined borderlines and describable by sets of singly necessary and jointly sufficient conditions [22]. Other views include the prototype view, the exemplar view, the frame view, and the theory view [22]. Each view captures specific aspects of concepts, and has a different implication for concept formation and learning. The applications of different views for inductive data analysis have been addressed by many authors [16, 21, 22].

In this section, we review the classical view of concepts and the basic issues related to concept formation and organization. More complete treatments of these issues can be found in the references [11, 15, 20, 21, 22].

3.1. Classical view of concepts

In the classical view, every concept is understood as a unit of thought that consists of two parts, the intension and the extension of the concept [20, 21, 22]. The intension (comprehension) of a concept consists of all properties or attributes that are valid for all those objects to which the concept applies. The extension of a concept is the set of objects or entities which are instances of the concept. All objects in the extension have the same properties that characterize the concept. In other words, the intension of a concept is an abstract description of common features or properties shared by elements in the extension, and the extension consists of concrete examples of the concept. A concept is thus described jointly by its intension and extension.

Extensional objects are mapped to intensional concepts through perception, and concepts are coded by words in speech. The two mappings of perception and speech define an indirect mapping between words and objects [12, 21]. This is depicted by the meaning triangle given in Figure 1. The peak is the concept, intension, thought, idea, or sense, the left corner is the symbol or word, and the right corner is the referent, object, or extension.

The classical view of concepts enables us to study concepts in a logic setting in terms of intensions and also in a set-theoretic setting in terms of extensions. Reasoning about intensions is based on logic [21]. Inductive inference and learning attempt to derive relationships between the intensions of concepts based on the relations between the extensions of concepts. Through the connections between extensions of concepts, one may establish relationships between concepts [26, 27].

3.2. Concept formation and structures

Human knowledge is conceptual and forms an integrated whole. In characterizing human knowledge, one needs to consider two topics, namely, context and hierarchy [15, 19]. The two topics have significant implications for concept formation and organization.

A context in which concepts are formed provides meaningful interpretations of the concepts. The theory view of concepts attempts to, to a large extent, reflect the contextual feature of concepts [11]. It is assumed that the formation of individual concepts and the overall conceptual structure depend on one's theory of a domain. One's theories and complex knowledge structures play a crucial role in concept formation, combination and learning.

Human knowledge is organized in a tower or a partial ordering. The base or minimal elements of the ordering are the most fundamental concepts and higher-level concepts depend on lower-level concepts [19]. The first-level concept is formed directly from the perceptual data [15]. The higher-level concepts, representing a relatively advanced state of knowledge, are formed by a process of abstracting from abstractions [15].

In concept formation, there are two basic issues known as aggregation and characterization [4]. Aggregation aims at the identification of a group of objects so that they form the extension of a concept. Characterization attempts to describe the derived set of objects in order to obtain the intension of the concept [4].

For aggregation, one considers two main processes called differentiation and integration [15]. Differentiation enables us to grasp the differences between objects, so that we can separate one or more objects from other objects. Integration is the process of putting together elements into an inseparable whole. As the final step in concept formation, characterization provides a definition of a concept.

4. Technique Level Study of Concepts

Based on the philosophy level study, one can build computational model for concept formation and learning. A particular concrete model is normally based on some philosophical assumptions and may not be able to cover all issues. As an illustration, we consider a simple model. The intensions are expressed as formulas of a logic language. The extensions are defined by adopting Tarski's approach through the notions of a model and satisfiability [3, 14, 26, 27]. Concept learning is modeled as search in a conjunctive concept space.

4.1. Intensions of concepts defined by a language

Traditionally, the intension of a concept is given by a set of properties. In artificial intelligence, one can define a language so that the intension of a concept is expressed as a formula of the language.

Let At be a finite set of attributes or features. For each attribute $a \in At$, we associate it with a set of values or labels V_a. Let U be a set of universe whose elements are called objects. For each $a \in At$, there is a mapping I_a connecting elements of U and elements of V_a. Furthermore, it is assumed that the mapping I_a is single-valued. In this case, the value of an object $x \in U$ on an attribute $a \in At$ is denoted by $I_a(x)$.

In order to formally define intensions of concepts, we adopt the decision logic language $\mathcal{L}$ used and studied by Pawlak [14]. Formulas of $\mathcal{L}$ are constructed recursively based on a set of atomic formulas corresponding to some basic concepts. An atomic formula is given by $a = v$, where $a \in At$ and $v \in V_a$. For each atomic formula $a = v$, an object x satisfies it if $I_a(x) = v$, written $x \models a = v$. Otherwise, it does not satisfy $a = v$ and is written $\neg x \models a = v$. From atomic formulas, we can construct other formulas by applying the logic connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$. The satisfiability of any formula is defined as follows:

$$
\begin{array}{ll}
(1) & x \models \neg\phi \text{ iff } \text{ not } x \models \phi, \\
(2) & x \models \phi \wedge \psi \text{ iff } x \models \phi \text{ and } x \models \psi, \\
(3) & x \models \phi \vee \psi \text{ iff } x \models \phi \text{ or } x \models \psi, \\
(4) & x \models \phi \rightarrow \psi \text{ iff } x \models \neg\phi \vee \psi, \\
(5) & x \models \phi \leftrightarrow \psi \text{ iff } x \models \phi \rightarrow \psi \text{ and } x \models \psi \rightarrow \phi.
\end{array}
$$

The language $\mathcal{L}$ can be used to reason about intensions. Each formula represents an intension of a concept. For two formulas ϕ and ψ, we say that ϕ is more specific than ψ, and ψ is more general than ϕ, if and only if $\models \phi \rightarrow \psi$, namely, ψ logically follows from ϕ. In other words, the formula $\phi \rightarrow \psi$ is satisfied by all objects with respect to any universe U and any information function I_a. If ϕ is more specific than ψ, we write $\phi \preceq \psi$, and call ϕ a sub-concept of ψ, and ψ a super-concept of ϕ.

4.2. Conjunctive concept space

Concept learning, to a large extent, depends on the structure of the target concepts. Typically, each learning algorithm focuses on a specific type of concepts.

Consider the class of conjunctive concepts used in version space learning method [9]. Let $\mathrm{CN}(\mathcal{L})$ denote the class of conjunctive concepts. It contains the special formula $\top$ which is satisfied by every object, the atomic formula, and formula constructed from atomic formula by only logic connective $\wedge$. Furthermore, we assume that an attribute appears at most once in each formula of $\mathrm{CN}(\mathcal{L})$.

The class $\mathrm{CN}(\mathcal{L})$ is referred to as the conjunctive concept space. For two concepts with $\phi \preceq \psi$, ϕ is called a specification of ψ and ψ a generalization of ϕ. Furthermore, ϕ is called a most general specification of ψ and ψ a most specific generalization of ϕ, if there does not exist another concept between ϕ and ψ. The conjunctive concept space can be represented as a graph by connecting a concept with its most specific generalizations and its most general specifications. At the top level, the most general concept is defined by the formula $\top$. The next level concepts are defined by atomic formulas. The combination of two atomic formulas produces the next level of concepts, and so on. Finally, at the bottom level, a most specific concept is formed by the conjunction of each atomic formula from every attribute.

Figure 2 draws part of the graph of the conjunctive space for three attributes, $\{\mathrm{Height}, \mathrm{Hair}, \mathrm{Eyes}\}$, with the following domains:

$$\begin{aligned} V_{\text{Height}} &= \{short,\ tall\}, \\ V_{\text{Hair}} &= \{blond,\ dark,\ red\}, \\ V_{\text{Eyes}} &= \{blue,\ brown\}. \end{aligned}$$

In the figure, an atomic formula is simply represented by the attribute value. For example, the atomic formula Height $= short$ is simply written as $short$.

We can classify conjunctive concepts by the number of atomic concepts in them. A concept involving k atomic concepts is called a k-conjunction. Obviously, the most general specifications of k-conjunction are $(k+1)$-conjunctions, and the most specific generalizations of k-conjunction are $(k-1)$-conjunctions.

4.3. Extensions of concepts defined by an information table

In inductive learning and concept formation, extensions of concepts are normally defined with respect to a particular training set of examples. With respect to a dataset, we can build a model based on an information table:

$$M = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}). \quad (1)$$

In this case, U is typically a finite set of objects.

If ϕ is a formula, the set $m(\phi)$ defined by:

$$m(\phi) = \{x \in U \mid x \models \phi\}, \quad (2)$$

is called the meaning of the formula ϕ in M. The meaning of a formula ϕ is therefore the set of all objects having the property expressed by the formula ϕ. In other words, ϕ can be viewed as the description of the set of objects $m(\phi)$. Thus, a connection between formulas and subsets of U is established. Obviously, the following properties hold [14]:

$$\begin{aligned} &\text{(a)} \quad m(\neg\phi) = -m(\phi), \\ &\text{(b)} \quad m(\phi \wedge \psi) = m(\phi) \cap m(\psi), \\ &\text{(c)} \quad m(\phi \vee \psi) = m(\phi) \cup m(\psi), \\ &\text{(d)} \quad m(\phi \rightarrow \psi) = -m(\phi) \cup m(\psi), \\ &\text{(e)} \quad m(\phi \equiv \psi) = (m(\phi) \cap m(\psi)) \cup \\ &\qquad\qquad (-m(\phi) \cap -m(\psi)). \end{aligned}$$

With the introduction of language $\mathcal{L}$, we have a formal description of concepts. A concept definable in a model M is a pair $(\phi, m(\phi))$, where $\phi \in \mathcal{L}$. More specifically, ϕ is a description of $m(\phi)$ in M, the intension of concept $(\phi, m(\phi))$, and $m(\phi)$ is the set of objects satisfying ϕ, the extension of concept $(\phi, m(\phi))$.

Table 1 is an example of an information table, taken from an example from Quinlan [17]. Each object is described by three attributes. The column labeled by Class denotes an expert's classification of the objects.

Object	Height	Hair	Eyes	Class
o_1	short	blond	blue	+
o_2	short	blond	brown	-
o_3	tall	red	blue	+
o_4	tall	dark	blue	-
o_5	tall	dark	blue	-
o_6	tall	blond	blue	+
o_7	tall	dark	brown	-
o_8	short	blond	brown	-

Table 1. An information table

A concept $(\phi, m(\phi))$ is said to be a sub-concept of another concept $(\psi, m(\psi))$, or $(\psi, m(\psi))$ a super-concept of $(\phi, m(\phi))$, in an information table if $m(\phi) \subseteq m(\psi)$. A concept $(\phi, m(\phi))$ is said to be a smallest non-empty concept in M if there does not exist another non-empty proper sub-concept of $(\phi, m(\phi))$. Two concepts $(\phi, m(\phi))$ and $(\psi, m(\psi))$ are disjoint if $m(\phi) \cap m(\psi) = \emptyset$. If $m(\phi) \cap m(\psi) \neq \emptyset$, we say that the two concepts have a non-empty overlap and hence are related.

4.4. Relationship between concepts in an information table

Based on the notions introduced so far, we can study a special type of knowledge represented by relationship between overlapping concepts. This type of knowledge is commonly referred to as rules. A rule can be expressed in the form, $\phi \Rightarrow \psi$, where ϕ and ψ are intensions of two concepts. A crucial issue is therefore the characterization, classification, and interpretation of rules. It is reasonable to expect that different types of rules represent different kinds of knowledge derivable from a database.

In data mining, rules are typically interpreted in terms of conditional probability [31]. For a rule $\phi \Rightarrow \psi$, its characteristics can be summarized by the following contingency table:

	ψ	$\neg\psi$	Total
ϕ	a	b	$a+b$
$\neg\phi$	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d=n$

$$\begin{aligned} a &= |m(\phi \wedge \psi)|, & b &= |m(\phi \wedge \neg\psi)|, \\ c &= |m(\neg\phi \wedge \psi)|, & d &= |m(\neg\phi \wedge \neg\psi)|. \end{aligned}$$

Different measures can be defined to reflect various aspects of rules.

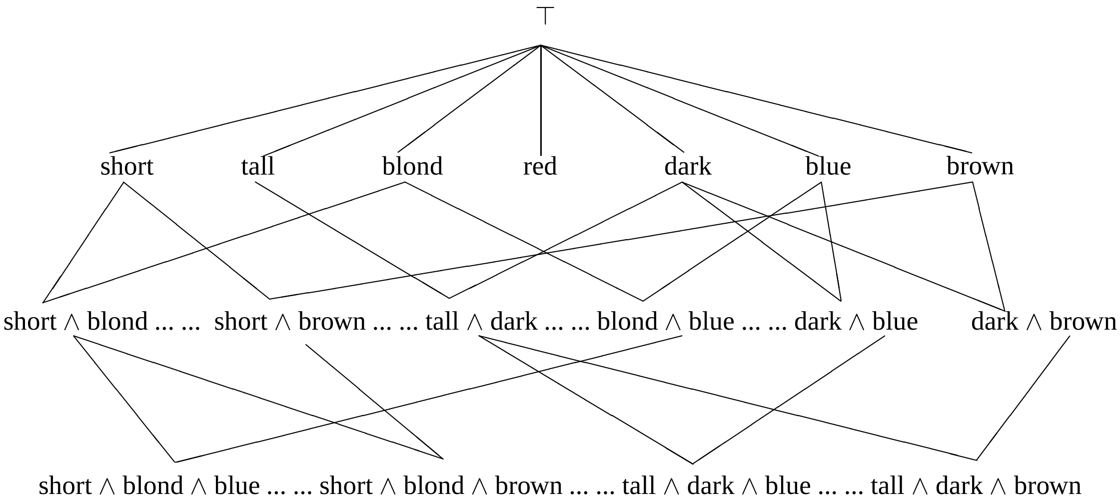

Figure 2. A conjunctive concept space

The *generality* of ϕ is defined by:

$$G(\phi) = \frac{|m(\phi)|}{|U|} = \frac{a+b}{n}, \quad (3)$$

which indicates the relative size of the concept ϕ. Obviously, we have $0 \leq G(\phi) \leq 1$. A concept is more general if it covers more instances of the universe. A sub-concept has a lower generality than its super-concept. The quantity may be viewed as the probability of a randomly selected element satisfying ϕ.

The *absolute support* of ψ provided by ϕ is:

$$AS(\phi \Rightarrow \psi) = AS(\psi|\phi) = \frac{|m(\psi) \cap m(\phi)|}{|m(\phi)|} = \frac{a}{a+b}, \quad (4)$$

The quantity, $0 \leq AS(\psi|\phi) \leq 1$, states the degree to which ϕ supports ψ. It may be viewed as the conditional probability of a randomly selected element satisfying ψ given that the element satisfies ϕ. In set-theoretic terms, it is the degree to which $m(\phi)$ is included in $m(\psi)$. Clearly, $AS(\psi|\phi) = 1$, if and only if $m(\phi) \neq \emptyset$ and $m(\phi) \subseteq m(\psi)$. That is, a rule with the maximum absolute support 1 is a certain rule.

The *mutual support* of ϕ and ψ is:

$$MS(\phi, \psi) = \frac{|m(\phi) \cap m(\psi)|}{|m(\phi) \cup m(\psi)|} = \frac{a}{a+b+c}. \quad (5)$$

One may interpret the mutual support, $0 \leq MS(\phi, \psi) \leq 1$, as a measure of the strength of a pair of rules $\phi \Rightarrow \psi$ and $\psi \Rightarrow \phi$.

The *change of support* of ψ provided by ϕ is defined by:

$$CS(\phi \Rightarrow \psi) = CS(\psi|\phi) = AS(\psi|\phi) - G(\psi) = \frac{a}{a+b} - \frac{a+c}{n}. \quad (6)$$

Unlike the absolute support, the change of support varies from -1 to 1. One may consider $G(\psi)$ to be the prior probability of ψ and $AS(\psi|\phi)$ the posterior probability of ψ after knowing ϕ. The difference of posterior and prior probabilities represents the change of our confidence regarding whether ϕ is actually related to ψ. For a positive value, one may say that ϕ is positively related to ψ; for a negative value, one may say that ϕ is negatively related to ψ.

The generality $G(\psi)$ is related to the satisfiability of ψ by all objects in the database, and $AS(\phi \Rightarrow \psi)$ is related to the satisfiability of ψ in the subset $m(\phi)$. A high $AS(\phi \Rightarrow \psi)$ does not necessarily suggest a strong association between ϕ and ψ, as a concept ψ with a large $G(\psi)$ value tends to have a large $AS(\phi \Rightarrow \psi)$ value. The change of support $CS(\phi \Rightarrow \psi)$ may be more accurate.

4.5. Concept learning as search

In concept learning, it is assumed that the extension of the concept is given through a set of positive and neg-

Input: a training set of examples S and
a partition of the training set Π,
Output: a set of formulas F.
Set $F = \emptyset$;
Set $k = 1$;
While S is not empty
  For each k-conjunction ϕ which is not a specification of a concept in F
    If $m(\phi) \cap S \neq \emptyset$ is a subset of a class in Π
      Add ϕ to F;
  For each k-conjunction ϕ in F
    Delete $m(\phi)$ from S;
  Set $k = k + 1$;
Return (F).

Figure 3. An algorithm for finding all most general concepts

ative examples. One may search for the most general specifications whose extensions are subsets of the extension of the given concept. The conjunctive concept space provides the search space, and the quantitative measures can be used to either direct the search or evaluate the results of learning. Two methods are outlined based on version space method [9], DNF learning [10], PRISM learning method [2], and a granular computing approach for machine learning [28].

We assume that a training set is partitioned into a group of subsets, each represents the extension of a concept. Furthermore, if two objects have the same description, they are in the same class of the partition.

In the first method, we learn all shortest conjunctive formulas that define the sub-concepts of a concept. This can be easily done by searching the conjunctive concept space from general concepts to specific concepts. Figure 3 presents such an algorithm for finding all most general sub-concepts of a family of concepts given by a partition.

For Table 1, the algorithm produces the set of conjunctive sub-concepts of the class $+$:

$$\begin{aligned} \{\ &\text{Hair} = \textit{red}, \\ &\text{Hair} = \textit{blond} \wedge \text{Eyes} = \textit{blue}, \\ &\text{Height} = \textit{short} \wedge \text{Eyes} = \textit{blue}, \\ &\text{Height} = \textit{tall} \wedge \text{Hair} = \textit{blond}\ \}. \end{aligned}$$

and the set of conjunctive sub-concepts of the class $-$:

$$\{\ \text{Hair} = \textit{dark}, \quad \text{Eyes} = \textit{brown}\ \}.$$

From the intension point of view, the two formulas $\text{Height} = \textit{short} \wedge \text{Eyes} = \textit{blue}$ and $\text{Hair} = \textit{blond} \wedge \text{Eyes} = \textit{blue}$ do not have a sub-concept relationship. On the other hand, their extensions with respect to Table 1 are:

$$\begin{aligned} &m(\text{Height} = \textit{short} \wedge \text{Eyes} = \textit{blue}) = \{o_1\}, \\ &m(\text{Hair} = \textit{blond} \wedge \text{Eyes} = \textit{blue}) = \{o_1, o_6\}. \end{aligned}$$

Form only extension point of view, one may choose the second formula as it covers more examples of $+$. Therefore, the algorithm in Figure 3 consider both intensions and extensions.

In practice, one may only be interested in a subset the formulas to characterize a partition. Instead of considering concepts based on the sequence defined by k, one may consider concepts in a sequence defined based on an evaluation function [28]. The algorithm given in Figure 4 finds a set of most general concepts whose extensions cover the training sets.

The set of concepts derived from the algorithm in Figure 4 depends on the evaluation function. For example, one may prefer concept with high generality and high support. The quantitative measures discussed earlier can be used to define various evaluation functions.

Consider the evaluation function defined by:

$$\begin{aligned} \textit{eval}(\phi) \quad = \quad &\max\{AS(\phi \Rightarrow \text{Class} = +), \\ &AS(\phi \Rightarrow \text{Class} = -)\}. \quad (7) \end{aligned}$$

That is, a concept is evaluated based on its maximum absolute support value of the class $+$ and the class $-$. For the information Table 1, the algorithm of Figure 4 produces a set of conjunctive concepts in the following sequence:

$$\begin{aligned} + :& \quad \text{Hair} = \textit{red}, \\ - :& \quad \text{Hair} = \textit{dark}, \\ - :& \quad \text{Eyes} = \textit{brown}, \\ + :& \quad \text{Height} = \textit{short} \wedge \text{Eyes} = \textit{blue}, \\ + :& \quad \text{Height} = \textit{tall} \wedge \text{Hair} = \textit{blond}. \end{aligned}$$

Input: a training set of examples S and
a partition of the training set Π,
an evaluation function $eval$,
Output: a set of formulas F.
Set $F = \emptyset$;
Set $WF =$ the set of all 1-conjunctions ϕ with $m(\phi) \neq \emptyset$;
While S is not empty
Select a best formula ϕ from WF according to the evaluation function $eval$
If $m(\phi)$ is a subset of a class in Π
Add ϕ to F;
Delete ϕ from WF;
Delete $m(\phi)$ from S;
If $m(\phi)$ is not a subset of a class in Π
Replace ϕ by its most general specifications;
Delete from WF concepts that are specifications of concepts in F;
Delete from WF every concept ϕ with $m(\phi) \cap S = \emptyset$;
Return (F).

Figure 4. An algorithm for finding a set of most general concepts

It finds the same atomic concepts as the former algorithm. However, the concepts defined by 2-conjunction are different. When a different evaluation function is used, different results may be obtained.

5. Application Level Study of Concepts

In the application level, one considers the issues related to the correct and effective use of concepts, such as concept definition and characterization, classification, and explanation. One may also explore the relationships between concepts. The application level study of concept may be guided by the purposes of learning, which in turn can be studied within a wide context of scientific research [29].

In a recent paper, Yao and Zhao [29] argued that scientific research and data mining are much in common in terms of their goals, tasks, processes and methodologies. Consequently, data mining and knowledge discovery research can be benefited from the long established studies of scientific research and investigation [8]. Concept learning is a specific topic of data mining and knowledge discovery. The same argument immediately applies.

Scientific research is affected by the perceptions and the purposes of science. Generally speaking, “science is the search for understanding of the world around us. Science is the attempt to find order and lawful relations in the world. It is a method of viewing the world.” [8] The main purposes of science are to describe and predict, to improve or manipulate the world around us, and to explain our world [8]. The results of the scientific research process provide a description of an event or a phenomenon. The knowledge obtained from research helps us to make predictions about what will happen in the future. Research findings are useful for us to make an improvement in the subject matter. Research findings can be used to determine the best or the most effective interventions to bring about desirable changes. Finally, scientists develop models and theories to explain why a phenomenon occurs.

Goals similar to those of scientific research have been discussed by many researchers in data mining. For example, Fayyad *et al.* [1] identified two high-level goals of data mining as prediction and description. Prediction involves the use of some variables to predict the values of some other variables, and description focuses on patterns that describe the data. Some researchers studied the issues of manipulation and action based on the discovered knowledge [6]. Yao, Zhao and Maguire [30] introduced a model of explanation-oriented data mining, which focuses on constructing models for the explanation of data mining results. The ideas may have a significant impact on the understanding of data mining and effective applications of data mining results.

Concepts learning should serve the same purposes, namely, to describe and predict, to improve or manipulate the world around us, and to explain our world.

Consider the example of Table 1. Concept learning enables us to describe and explain the classes of $+$ and $-$ using other concepts defined by attributes. Based on the results of the algorithm in Figure 4, the class $+$ is described and explained by the disjunction of three conjunctive concepts:

$$(\text{Hair} = \mathit{red}) \vee$$
$$(\text{Height} = \mathit{short} \wedge \text{Eyes} = \mathit{blue}) \vee$$
$$(\text{Height} = \mathit{tall} \wedge \text{Hair} = \mathit{blond}).$$

Similarly, the class $-$ is described by the disjunction of two conjunctive concepts:

$$(\text{Hair} = \mathit{dark}) \vee (\text{Eyes} = \mathit{brown}).$$

It should be noted that the same class may also be described and explained by a different set of concepts.

For prediction, results of algorithm in Figure 4, produce the following classification rule:

$$\text{Hair} = \mathit{red} \Rightarrow \text{Class} = +,$$
$$(\text{Height} = \mathit{short} \wedge \text{Eyes} = \mathit{blue}) \Rightarrow \text{Class} = +,$$
$$(\text{Height} = \mathit{tall} \wedge \text{Hair} = \mathit{blond}) \Rightarrow \text{Class} = +,$$
$$\text{Hair} = \mathit{dark} \Rightarrow \text{Class} = -,$$
$$\text{Eyes} = \mathit{brown} \Rightarrow \text{Class} = -.$$

That is, we can predict the class of an object based on its attribute values.

In some situations, the tasks of description and prediction may not be clearly separated. In order to have a good prediction one must have a good description and explanation.

The concept learning methods can be applied to study relationships between attributes. This can be simply done by generating the partition Π using one subset of attributes, and by learning using another subset of attributes. The results can be explained in a similar manner.

6. Conclusion

From the perspective of cognitive informatics, this paper examines concept formation and learning. In the philosophy level study, we focus on the definition, interpretation of concepts, and cognitive process for concept formation and learning. In the technique level, we focus on a specific language for defining concepts and present two algorithms for concept learning. In the application level, we study explanations and uses of the learned results.

The objective of the paper is aimed at a more general framework for concept formation and learning, rather than a more efficient algorithm. Although some of the results are not entirely new, their treatment from the cognitive informatics perspective leads to new insights. The proposed framework may be easily applied to study any topic in cognitive informatics.

Our investigation demonstrates that the introduction of cognitive informatics offers us opportunity and challenges to re-consider many issues in established fields.

References

[1] Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. From data mining to knowledge discovery: an overview, in: *Advances in Knowledge Discovery and Data Mining*, Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds.), AAAI/MIT Press, Menlo Park, CA, 1-34, 1996.

[2] Cendrowska, J. PRISM: an algorithm for inducing modular rules, *International Journal of Man-Machine Studies*, **27**, 349-370, 1987.

[3] Demri, S. and Orlowska, E. Logical analysis of indiscernibility, in: *Incomplete Information: Rough Set Analysis*, Physica-Verlag, Heidelberg, 347-380, 1998.

[4] Feger, H. and de Boeck, P. Categories and concepts: introduction to data analysis, in [22], 203-223.

[5] Lin, T.Y., Yao, Y.Y. and Louie, E., Value added association rules, *Proceedings of PAKDD'02*, 328-333, 2002.

[6] Ling, C.X., Chen, T., Yang, Q. and Cheng, J. Mining optimal actions for profitable CRM, *Proceedings of 2002 IEEE International Conference on Data Mining*, 767-770, 2002.

[7] Michalski, R.S., Carbonell, J.G., and Mitchell, T.M. (Eds.), *Machine Learning, an Artificial Intelligence Approach*, Morgan Kaufmann Publishers, Inc., Los Altos, California, 1983.

[8] Martella, R.C., Nelson, R. and Marchand-Martella, N.E. *Research Methods: Learning to Become a Critical Research Consumer*, Allyn and Bacon, 1999.

[9] Mitchell, T.M. Generalization as search, *Artificial Intelligence*, **18**, 203-226, 1982.

[10] Mooney, R.J. Encouraging experimental results on learning CNF, *Machine Learning*, **19**, 79-92, 1995.

[11] Murphy, G.L. Theories and concept formation, in [22], 173-200.

[12] Ogden, C.K. and Richards I.A. *The Meaning of Meaning: a Study of the Influence of Language upon Thought and of the Science of Symbolism*, 8th edition, Harcourt Brace, New York, 1946.

[13] Patel, D., Patel, S. and Wang, Y. (Eds.) *Proceedings of the Second IEEE International Conference on Cognitive Informatics*, IEEE Computer Society, Los Alamitos, California, 2003.

[14] Pawlak, Z. *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Boston, 1991.

[15] Peikoff, L. *Objectivism: the Philosophy of Ayn Rand*, Dutton, New York, 1991.

[16] Posner, M.I. (Ed.) *Foundations of Cognitive Science*, The MIT Press, Cambridge, Massachusetts, 1989.

[17] Quinlan, J.R. Learning efficient classification procedures and their application to chess endgames, in: *Machine Learning: An Artificial Intelligence Approach*, vol. 1, Michalski, J.S., Carbonell, J.G., and Michell, T.M. (Eds), Morgan Kaufmann, Palo Alto, CA, 463-482, 1983.

[18] Simon, H.A. and Kaplan, C.A. Foundations of cognitive science, in [16], 1-47, 1989.

[19] Simpson, S.G. What is foundations of mathematics? 1996, http://www.math.psu.edu/simpson/hierarchy.html, retrieved November 21, 2003.

[20] Smith, E.E. Concepts and induction, in: *Foundations of Cognitive Science*, Posner, M.I. (ed.), The MIT Press, Cambridge, Massachusetts, 501-526, 1989.

[21] Sowa, J.F. *Conceptual Structures, Information Processing in Mind and Machine,* Addison-Wesley, Reading, Massachusetts, 1984.

[22] Van Mechelen, I., Hampton, J., Michalski, R.S. and Theuns, P. (Eds.), *Categories and Concepts, Theoretical Views and Inductive Data Analysis*, Academic Press, New York, 1993.

[23] Wang, Y. On cognitive informatics, in [24], 34-42, 2002.

[24] Wang, Y., Johnston, R.H. and Smith, M.R. (Eds.) *Proceedings of the First IEEE International Conference on Cognitive Informatics*, IEEE Computer Society, Los Alamitos, California, 2002.

[25] Wang, Y., Patel, S., Patel, P. and Wang, Y. A layered reference model of the brain, in [13], 7-17, 2003.

[26] Yao, Y.Y. Modeling data mining with granular computing, *Proceedings of the 25th Annual International Computer Software and Applications Conference*, 638-643, 2001.

[27] Yao, Y.Y. A step towards the foundations of data mining, in: *Data Mining and Knowledge Discovery: Theory, Tools, and Technology V*, Dasarathy, B.V. (Ed.), The International Society for Optical Engineering, 254-263, 2003.

[28] Yao, J.T. and Yao, Y.Y. A granular computing approach to machine learning, *Proceedings of the 1st International Conference on Fuzzy Systems and Knowledge Discovery*, 732-736, 2002.

[29] Yao, Y.Y. and Zhao, Y. Explanation-oriented data mining, Manuscript, 2004.

[30] Yao, Y.Y., Zhao, Y. and Maguire, R.B. Explanation-oriented association mining using rough set theory, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Proceedings of the 9th International Conference*, 165-172, 2003.

[31] Yao, Y.Y. and Zhong, N. An analysis of quantitative measures associated with rules, *Proceedings of PAKDD'99*, 479-488, 1999.