Abduction in Non-Axiomatic Logic

Pei Wang
Center for Research on Concepts and Cognition, Indiana University
pwang@cogsci.indiana.edu
http://www.cogsci.indiana.edu/farg/peiwang/

Abstract

This paper introduces various inference rules (deduction, abduction, and induction) in Non-Axiomatic Logic. These rules are represented in a term-oriented language, and justified according to a common semantic foundation. The implementation and application of these rules are briefly described. Finally, this approach is compared with the other approaches, specially with respect to abduction.

1 Introduction

In the study of artificial intelligence, many researchers reach the conclusion that the currently dominating logic system, First-Order Predicate Logic (FOPL), is not a proper model of intelligent reasoning, and various alternative logic systems have been proposed. Non-Axiomatic Logic (NAL) [Wang, 1994; Wang, 1995; Wang, 2000] is one of them.

What distinguishes NAL from other logic systems is its assumption of *insufficient knowledge and resources*. When the logic is used to answer questions according to given knowledge, the knowledge may be uncertain and incomplete (with respect to the questions), and the system may not have enough time to consider all relevant knowledge for a given question. Consequently, each piece of knowledge in such a system is only true to a certain degree, which can be revised according to new evidence. Also, the system must use plausible inference rules to derive "best guesses" when no sure or optimum answers can be obtained.

It is easy to see that FOPL cannot be used in the above situation. In the following, it will be shown that a logic designed to work in such a situation needs a new formal language, a new semantic theory, and a new set of inference rules.

Among the various aspects of NAL, this paper is focused on *abduction*. For other relevant issues, please see the publications in the author's webpage. In the following, we start with a review of First-Order Non-Axiomatic Logic (FONAL), which has been mostly covered by previous publications on NAL, but is needed to understand the new progress. Then, Higher-Order Non-Axiomatic Logic (HONAL) is formally specified for the first time. Finally, NAL is compared with the other approaches, specially in its treatment of abduction.

2 First-Order Non-Axiomatic Logic

This section provides an updated summary of FONAL, based on the previous publications [Wang, 1994; Wang, 1995; Wang, 2000].

2.1 Language and semantics

Under the assumption of insufficient knowledge and resources, a statement in NAL cannot be either completely true or completely false. Instead, its truth value must be a matter of degree. Since the only available information about the world is the system's "experience", i.e., stream of input statements (with their truth values), the truth value of a statement should indicates its relationship with available evidence in the experience of the system.

Therefore, NAL needs a formal language in which the (positive and negative) evidence for a given statement can be naturally defined and measured, which in turn defines the truth value of the statement. For this reason, NAL use a *term-oriented language*.

In NAL, an *Inheritance relation* "$\subset$" is a reflexive and transitive relation defined between a *subject term* S and a *predicate term* P, where a *term* is the name of a concept. Intuitively, a statement $S \subset P$ says that S is a *specialization* of P, and P is a *generalization* of S. This roughly corresponds to "S is a kind of P" in English. For example, "Bird is a kind of animal" can be represented as $bird \subset animal$.

In NAL, the *extension* and *intension* of a term T are defined as sets of terms:

$$T^E = \{x \mid x \subset T\} \ ; \ \ T^I = \{x \mid T \subset x\}$$

Intuitively, they include all known specialization (instances) and generalizations (properties) of T, respectively. [1]

Please note that the above definition of extension and intension of a term is different from the common one, where the "extension" of a term is the corresponding "objects" in an outside world, while its "intension" is the corresponding concepts in a Platonic space. NAL cannot accept such a definition, because neither such an outside world nor such a Platonic space can be assumed with insufficient knowledge. Even though, the NAL definition still preserves the intuitive

[1] The notations for extension and intension used in this paper are different from those in previous publications on NAL, where E_T and I_T were used.

meaning of the two words, that is, "extension" is for "instances", which are more specific than the term itself, while "intension" is for "properties", which are more general. Under the NAL definition, "extension" and "intension" is a dual relation among terms. That is, T_1 is in the extension of T_2, if and only if T_2 is in the intension of T_1.

From the reflexivity and transitivity of Inheritance, it can be proven that

$$(S \subset P) \Longleftrightarrow (S^E \subseteq P^E) \Longleftrightarrow (P^I \subseteq S^I)$$

where the first relation is an Inheritance relation between two terms, while the last two are subset relations between two sets (extensions and intensions of terms).

The above theorem identifies $S \subset P$ with "P inherits the extension of S, and S inherits the intension of P", which is a summary of multiple statements, so can be used as evidence for the Inheritance statement. Uncertain statements correspond to the situations where the above inheritance of extension and intension is incomplete.

For a statement $S \subset P$ and a term M, we have

- if M is in the extensions of both S and P, it is positive evidence for the statement (because as far as M is concerned, P indeed inherits the extension of S);
- if M is in the extensions of S but not the extension of P, it is negative evidence (because as far as M is concerned, P fails to inherit the extension of S);
- if M is in the intensions of both P and S, it is positive evidence for the statement (because as far as M is concerned, S indeed inherits the intension of P);
- if M is in the intension of P but not the intension of S, it is negative evidence (because as far as M is concerned, S fails to inherit the intension of P);
- otherwise, M is irrelevant to the statement.

Therefore, when all pieces of evidence are treated as equal, the amount of positive evidence is

$$w^+ = |S^E \cap P^E| + |P^I \cap S^I|$$

the amount of negative evidence is

$$w^- = |S^E - P^E| + |P^I - S^I|$$

and the amount of all evidence is

$$w = w^+ + w^- = |S^E| + |P^I|$$

The *truth value* of a statement in NAL is a pair of numbers in [0, 1], $<f,\ c>$. f is the *frequency* of the statement, defined as

$$f = w^+/w$$

so it indicates the proportion of positive evidence among all evidence. c is the *confidence* of the statement, defined as

$$c = w/(w+1)$$

so it indicates the proportion of current evidence among evidence in the near future (after a unit-weight evidence is collected). When $f = 1$, it means that all known evidence is positive; when $f = 0$, it means that all known evidence is negative; when $c = 0$, it means that the system has no evidence on the statement at all (and f is undefined); when $c = 1$, it means that the system already has all the evidence on the statement, so that it will not be influenced by future experience. Therefore, "absolute truth" has a truth value $<1,\ 1>$, and in NAL $S \subset P <1,\ 1>$ can be written as $S \subset P$, as we did earlier in the discussion. Under the "insufficient knowledge" assumption, such a truth value cannot be reached by empirical knowledge, though it can be used for analytical knowledge (such as theorems in mathematics), as well as serve as idealized situation in semantic discussions. For a more detailed discussion on frequency, confidence, and their relation with probability theory, see [Wang, 2001].

In this way, we get an "experience-grounded semantics", where truth value of a statement indicates its relation with available evidence, rather than its relation with a model or an "outside world".

2.2 Basic inference rules

The above subsection defines truth value in terms of amount of evidence collected in idealized situation (so the evidence itself is certain). In actual situation, the truth value of a statement cannot be determined in this way. Instead, the input knowledge comes into the system with truth value assigned by the user or other knowledge sources (according to the above semantics), and derived knowledge is produced recursively by the built-in inference rules, which have truth value functions that determine the truth value of the conclusion according to those of the premises. The truth value functions are defined according to the above semantics.

Typical inference rules in term logic take a *syllogistic* form, that is, given a pair of statements, if they share a common term, then a conclusion between the other two (not shared) terms can be derived by an inference rule.

Different combinations of premises correspond to different inference rules, as listed in the following table:

$$\textbf{Deduction}: \quad \frac{\begin{array}{l} M \subset P <f_1,\ c_1> \\ S \subset M <f_2,\ c_2> \end{array}}{S \subset P <f,\ c>}$$

$$\textbf{Abduction}: \quad \frac{\begin{array}{l} P \subset M <f_1,\ c_1> \\ S \subset M <f_2,\ c_2> \end{array}}{S \subset P <f,\ c>}$$

$$\textbf{Induction}: \quad \frac{\begin{array}{l} M \subset P <f_1,\ c_1> \\ M \subset S <f_2,\ c_2> \end{array}}{S \subset P <f,\ c>}$$

Defined in this way, the difference among the three types of inference is purely formal:

- in deduction, the shared term is the subject of one premise and the predicate of the other;
- in abduction, the shared term is the predicate of both premises;
- in induction, the shared term is the subject of both premises.

If we only consider combinations of premises with one shared term, these three exhaust all the possibilities.[2]

According to the previously described semantics, in NAL an inference rule is valid as long as the conclusion is based on the evidence provided by the premises. In different rules, the truth value of the conclusion is determined by the truth values of the premises in different ways.

Since frequency and confidence cannot be handled as probability values [Wang, 2001], we treat them as *extended Boolean variables* to get the truth value functions.

Step 1: we treat all relevant variables as binary variables taking 0 or 1 values, and determine what values the conclusion should have for each combination of premises, according to the semantics of NAL.

In the case of deduction, the result comes from the following analysis:

- Inheritance relation is transitive in its ideal form, so f and c are both 1 if and only if f_1, c_1, f_2, and c_2 are all 1.
- Two negative premises produce no conclusion, so c is 1 only when f_1 or f_2 is 1.
- An unfounded premise produces no conclusion, so c is 1 only when c_1 and c_2 are both 1.

In the cases of abduction and induction, the mid-term M is used as possible evidence for the conclusion, so what can be directly determined in the conclusion are not f and c, but w and w^+. Concretely, we have:

- M is positive evidence (i.e., w^+ is 1) if and only if f_1, c_1, f_2, and c_2 are all 1.
- In abduction, M is evidence (i.e., w is 1) if and only if f_1, c_1, and c_2 are 1.
- In induction, M is evidence (i.e., w is 1) if and only if f_2, c_1, and c_2 are 1.

Step 2: The truth values of conclusion obtained above are represented as Boolean functions of those of the premises.

$$\begin{aligned} \textbf{Deduction}: \quad & AND(f,c) = AND(f_1,c_1,f_2,c_2) \\ & c = AND(c_1,c_2,OR(f_1,f_2)) \\[1ex] \textbf{Abduction}: \quad & w^+ = AND(f_1,c_1,f_2,c_2) \\ & w = AND(f_1,c_1,c_2) \\[1ex] \textbf{Induction}: \quad & w^+ = AND(f_1,c_1,f_2,c_2) \\ & w = AND(c_1,f_2,c_2) \end{aligned}$$

Step 3: The Boolean operators are extended into real number functions defined on [0, 1] in the following way:

$$\begin{aligned} NOT(x) &= 1 - x \\ AND(x_1,...,x_n) &= x_1 * ... * x_n \\ OR(x_1,...,x_n) &= 1 - (1-x_1) * ... * (1-x_n) \end{aligned}$$

They are the operators used in probability theory under independent assumptions. For why this set of function is selected in NAL, see [Wang, 1995].

[2]The order of the premises does not matter for our current purpose. The symmetric conclusion $P \subset S$ is omitted to simplify the discussion.

Step 4: Using the extended operators, plus the relationship between truth value and amount of evidence, to rewrite the above functions, so to get the following truth value functions for the rules:

$$\begin{aligned} \textbf{Deduction}: \quad & f = f_1f_2/(f_1+f_2-f_1f_2) \\ & c = c_1c_2(f_1+f_2-f_1f_2) \\[1ex] \textbf{Abduction}: \quad & f = f_2 \\ & c = f_1c_1c_2/(f_1c_1c_2+1) \\[1ex] \textbf{Induction}: \quad & f = f_1 \\ & c = f_2c_1c_2/(f_2c_1c_2+1) \end{aligned}$$

In deduction, when f_1 and f_2 are both 0, we add the convention that $f = 0$, so that the truth value function always returns a valid value.

2.3 Other rules in FONAL

There are other rules in FONAL. In this paper we only briefly describe some of them that are directly related to abduction.

Though the abduction (as well as deduction and induction) rule introduced in the previous section are defined on Inheritance relation, they can be used on other types of relations.

For example, an arbitrary relation R among three terms A, B, and C is usually written as $R(A,B,C)$, which can be equivalently rewritten as one of the following Inheritance statements (i.e., they have the same meaning and truth value):

- $(A,B,C) \subset R$, where the subject term is a compound (A,B,C), an ordered tuple. This statement says "The relation among A, B, C (in that order) is a special case of the relation R."
- $A \subset R(*,B,C)$, where the predicate term is a compound $R(*,B,C)$ with a "wildcard", $*$. This statement says "A is such an x that satisfies $R(x,B,C)$."
- $B \subset R(A,*,C)$. Similarly, "B is such an x that satisfy $R(A,x,C)$."
- $C \subset R(A,B,*)$. Again, "C is such an x that satisfy $R(A,B,x)$."

In this way, all types of relations can be treated as Inheritance by the rules defined previously. For example, from $R(A,B,C)$ and $R(A,B,D)$, $C \subset D$ can be derived by the abduction rule (truth values omitted in this example). This is the case because the two premises can be rewritten as $C \subset R(A,B,*)$ and $D \subset R(A,B,*)$, and from them $C \subset D$ (and $D \subset C$, of course) can be derived by abduction. Intuitively, since C and D share the same relation R with A and B, Inheritance relations between the two are supported to a degree.

Therefore, with **transformation rules** that convert a statement to its equivalent forms, the system can do abduction (as well as other types of inference) on arbitrary types of relations. Please note that these relations have a position different from that of the Inheritance relation in NAL. Inheritance is "built-in" for NAL, in the sense that its meaning is fixed, and directly recognized by the inference rules. On the other hand, the other relations are "user-defined", whose meaning is learned by the system according to relevant experience. They are processed by the inference rules indirectly, as Inheritance relations consisting of compound terms. Whether to let Inheritance play a central role is a major difference between NAL and FOPL, in the latter Inheritance (similar to the often used "is-a") is treated just like other relations.

The **revision** rule merges statements that have the same content (i.e., the same $S \subset P$ form), but based on separated bodies of evidence. Formally, it has the form

$$\textbf{Revision}: \quad \frac{\begin{array}{l} S \subset P <f_1, c_1> \\ S \subset P <f_2, c_2> \end{array}}{S \subset P <f, c>}$$

Given the additivity of evidence during revision, we have

$$w^+ = w_1^+ + w_2^+ \; ; \;\; w = w_1 + w_2$$

which corresponds to the following truth value function:

$$f = \frac{f_1 c_1/(1-c_1) + f_2 c_2/(1-c_2)}{c_1/(1-c_1) + c_2/(1-c_2)}$$
$$c = \frac{c_1/(1-c_1) + c_2/(1-c_2)}{c_1/(1-c_1) + c_2/(1-c_2) + 1}$$

This rule can merge multiple abductive conclusions into a more confident one.

The **choice** rule determines which statement is more likely to be confirmed in the future among a given set of statements. What it does is to pick the one with the highest *expectation* value, defined as

$$e = c(f - 0.5) + 0.5$$

This rule can be used to select the best abductive conclusion, according to both frequency and confidence.

For a more detailed discussion about the revision rule and the choice rule, see [Wang, 1995].

2.4 Implementation of NAL

NAL is a logic, with its formal language, semantics, and inference rules. This logic can be used in a reasoning system, if proper memory and control mechanism are provided. Such a system, Non-Axiomatic Reasoning System (NARS), has been under development for years. The current version, NARS 4.1, is a Java applet available at the author's webpage, which contains several simple examples to show the various aspects of the system.

Like NAL, NARS is also designed under the assumption of insufficient knowledge and resources. What NARS has beside NAL is a resource management mechanism that handles the limited time and space resources, as well as a user interface to communicate with the users and other systems. To describe NARS as a whole is far beyond the scope of this paper. Interested readers can visit the author's webpage.

A customized version of NAL has been integrated into the design of a commercial software, Webmind. Though the project has not been finished yet, NAL has shown its advantage over FOPL in the development so far. For more information about this project, visit *http://www.webmind.com*.

3 Higher-order Non-Axiomatic Logic

This section describes the recent progress which is not covered by previous publications on NAL.

3.1 Extended language and semantics

While the Inheritance relation is introduced as a relation between *terms*, a *higher-order relation* is a relation defined between *statements*. Concretely, we define an *Implication* relation as a reflexive and transitive relation between two statements. Intuitively, an Implication statement $P \rightarrow Q$ means "If P, then Q", where P and Q are statements themselves. For example, $(dove \subset bird) \rightarrow (dove \subset animal)$ is an Implication statement, with (first-order) statements $dove \subset bird$ and $dove \subset animal$ as components. This statement says "If dove is a kind of bird, then dove is a kind of animal".

In NAL, Implication is defined in a way that makes it completely isomorphic to Inheritance.

In the place of extension and intension, here we have *sufficient condition* and *necessary condition*, that is, when $P \rightarrow Q$ is true, we call P a sufficient condition of Q, and Q a necessary condition of P. Formally, the *sufficient conditions* and *necessary conditions* of a statement P are defined as sets of statements:

$$P^S = \{x \mid x \rightarrow P\} \; ; \;\; P^N = \{x \mid P \rightarrow x\}$$

Parallel to the case of Inheritance relation, positive evidence of $P \rightarrow Q$ consists of shared sufficient and necessary conditions of P and Q, while negative evidence consists of the sufficient conditions of P but not of Q, and necessary conditions of Q but not of P. Formally, we have

$$w^+ = |P^S \cap Q^S| + |Q^N \cap P^N|$$
$$w^- = |P^S - Q^S| + |Q^N - P^N|$$
$$w = w^+ + w^- = |P^S| + |Q^N|$$

Truth value of an Implication statement is defined in the same way as that of an Inheritance statement, that is,

$$f = w^+/w \; ; \;\; c = w/(w+1)$$

Defined in this way, Implication is completely isomorphic to Inheritance, so that for each result on Inheritance, there is a corresponding result on Implication. On the other hand, the two are not the same, because "A is a kind of B" and "If A, then B" have different meaning. Inheritance "A is a kind of B" cannot be simply understood as "If x is a kind of A, then x is also a kind of B" where x is a variable, but should be understood as "If x is a kind of A, then it is also a kind of B; if B is a kind of y, then A is also a kind of y". The latter statement is a conjunction of two Implication statements, and cannot be reduced into a single Implication statement without changing its meaning in the context of NAL.

3.2 Basic higher-order inference rules

Given the isomorphism between Inheritance and Implication, we can easily get a set of HONAL inference rules from the FONAL rules developed previously. For example, HONAL has the following higher-order syllogistic rules:

$$\begin{array}{rl}\textbf{Deduction}: & \begin{array}{l} M \rightarrow P \ <f_1,\ c_1> \\ S \rightarrow M \ <f_2,\ c_2> \\ \hline S \rightarrow P \ <f,\ c> \end{array} \\[2ex] \textbf{Abduction}: & \begin{array}{l} P \rightarrow M \ <f_1,\ c_1> \\ S \rightarrow M \ <f_2,\ c_2> \\ \hline S \rightarrow P \ <f,\ c> \end{array} \\[2ex] \textbf{Induction}: & \begin{array}{l} M \rightarrow P \ <f_1,\ c_1> \\ M \rightarrow S \ <f_2,\ c_2> \\ \hline S \rightarrow P \ <f,\ c> \end{array}\end{array}$$

The truth value function for each rule is the same as in the corresponding rule of Inheritance relation, though the rules in these two tables are not identical, but isomorphic.

Similarly, the FONAL rules for revision, choice, and transformation have their corresponding forms in HONAL.

3.3 Treating any statement as Implication

As discussed previously, an Inheritance statement cannot be seen as an Implication statement by taking a term as a corresponding statement. However, there is another way to relate the two types of statements.

By definition, $S \subset P < f,\ c >$ says that "The belief the system has on statement $S \subset P$, according to available evidence, is measured by truth value $< f,\ c >$". Now if we assume the available evidence on $S \subset P$ can be written as a complex statement E, then the same meaning can be represented by $E \rightarrow (S \subset P) <f,\ c>$, that is, "The belief the system has on statement 'If E is true, then $S \subset P$ is true' is measured by truth value $<f,\ c>$". In this way, a statement S is equivalently transformed into an Implication statement $E \rightarrow S$ ("If the available evidence is true, then S is true").

This transformation is a conceptual one, not an actual one in the sense that there is a statement used by NAL corresponding to the above E. This conceptual transformation is introduced to help us to find valid inference rules. Specially, now we can handle the following situations where Inheritance and Implication statements are mixed together (truth values omitted):

- M, when being treated as $E \rightarrow M$, will do **deduction** with $M \rightarrow P$ to get $E \rightarrow P$, which is actually P.
- M, when being treated as $E \rightarrow M$, will do **abduction** with $P \rightarrow M$ to get $E \rightarrow P$, which is actually P.
- P and S, when being treated as $E \rightarrow P$ and $E \rightarrow S$, respectively, will do **induction** together to get $S \rightarrow P$ (and $P \rightarrow S$, of course).

Please note that in induction, the two premises should be indeed derived from the same evidence. It is a difference between NAL and traditional propositional logic: given the truth value of S and P, the truth value of $S \rightarrow P$ is not always derivable, and even when it can be derived, it usually has a much lower confidence value than M and P have.

By hiding the conceptual transformation introduced above and adding truth values back, we get the third table of inference rule:

$$\begin{array}{rl}\textbf{Deduction}: & \begin{array}{l} M \rightarrow P \ <f_1,\ c_1> \\ M \ <f_2,\ c_2> \\ \hline P \ <f,\ c> \end{array} \\[2ex] \textbf{Abduction}: & \begin{array}{l} P \rightarrow M \ <f_1,\ c_1> \\ M \ <f_2,\ c_2> \\ \hline P \ <f,\ c> \end{array} \\[2ex] \textbf{Induction}: & \begin{array}{l} P \ <f_1,\ c_1> \\ S \ <f_2,\ c_2> \\ \hline S \rightarrow P \ <f,\ c> \end{array}\end{array}$$

What we get here is very similar to how the three types of inference are defined in propositional logic [Aliseda, 2000; Flach and Kakas, 2000], except in NAL the statements have truth values attached to indicate their uncertainty. Since this table can be seen as a special case of the second table, the truth value functions remain unchanged.

In summary, HONAL has three types of syllogistic inference rules, that is, three forms of abduction (and the same for deduction and induction). The three types share the same truth value function, though each rule has its own meaning and applicable situation.

4 Comparison and Discussion

As mentioned at the beginning of the paper, NAL is proposed as a logic that assumes the insufficiency of knowledge and resources. In a sense, it challenges the dominating position of FOPL and its variants in the study of Artificial Intelligence. Since most existing approaches toward abduction are developed within the framework of FOPL (see [Flach and Kakas, 2000] for a survey), it is not a surprise that NAL is different from them in several major issues.[3]

4.1 Syllogistic and inferential definition of abduction

Different definitions have been proposed for abduction (as well as for induction, and even deduction). On this issue, I fully agree with Flach and Kakas [2000] that we should not treat "abduction" as a Platonic concept waiting to be found, but should build useful definitions that help the research. For this reason, what Peirce had in mind when he coined the term "abduction", or why he changed his mind in later years, only serves as important source of inspiration, but not as sole standard to judge the validity of a definition of abduction.

Approaches of defining abduction can be classified into two types: syllogistic and inferential. An inferential definition identifies abduction as a type of inference *process* that carries out a certain cognitive function, such as *explanation* or *hypothesis generation*, while a syllogistic definition

[3]This paper will not touch differences between NAL and FOPL that are not directly related to abduction.

specifies it as a type of inference *step* with a specific pattern [Flach and Kakas, 2000; Josephson and Josephson, 1994; Wang, 2000].

As shown by the three rule tables, in NAL the distinction among deduction, abduction, and induction is formally specified at the inference-step level, according to the position of the shared term (or statement) in the premises. Such a formal definition makes discussions about them clear and concrete.

To use a formal definition to distinguish various inference types does not prevent us from attribute them with different cognitive functions. Given the definition used in NAL, it is valid to say that among the three, only deduction produces conclusive results, while the other two only produce tentative results. Both abduction and induction can be seen as "reversed deduction", and the former usually corresponds to explanation, and the latter to generalization. These descriptions are similar to the ones proposed as inferential definitions of the three types. However, in NAL these descriptions are *secondary*, derived from the syllogistic definition. This approach has the advantage of avoiding ambiguity and oversimplification in the definition, and at the same time preserve the intuitive meaning of the terms (i.e., deduction, abduction, and induction) associated with different types of inference.

Though abduction defined in NAL usually can be interpreted as "explanation", to define "abduction" as "explanation" at the inference-process level is a quite different decision. This is the case because what we called "explanation" in everyday thinking may include complex cognitive process where multiple types of inference are involved. Therefore, to abstract such a process into a consistent and non-trivial pattern is not an easy thing to do, if possible [Wang, 2000].

For the same reason, to define abduction as "inference toward the best explanation" [Josephson and Josephson, 1994] makes things even harder, because besides the derivation of explanations, this definition further requires the evaluation of explanations, and the comparison of competing candidates. In this process, many other factors should be taken into account, such as simplicity, surprising to the system, and relevance to the given context [Aliseda, 2000; Psillos, 2000]. If we cover all of these issues under "abduction", it becomes such a complex process that few concrete conclusions can be made. Such a definition is not wrong, but not very useful.

4.2 Multi-valued and binary conclusions

Currently, most theories of abduction use binary logic [Flach and Kakas, 2000], with a few exceptions, such as NAL and the Bayesian approach [Poole, 2000].

In the framework of binary logic, abduction is usually defined formally as "reverse deduction" which starts from a given conclusion and background knowledge to find a premise that is consistent with the background knowledge, and derives the conclusion deductively.

Such a definition is logically sound, and can lead to fruitful results. However, it ignores certain factors that are crucial for a system working with insufficient knowledge and resources.

In empirical science and everyday life, we usually do not throw away theories that have known counter examples and inexplicable phenomena. If we do that, there is hardly anything left. Since we usually have insufficient knowledge in these domains, we have to live with imperfect knowledge, because they are still far better than random guesses. When selecting among competing explanations and hypothesis, measurement of (positive and negative) evidence becomes necessary — if no explanation is perfect, then the one with more positive evidence and less negative evidence is preferred, which is what measured by the *frequency* defined in NAL. Since evidence may come from time to time, incremental revision becomes inevitable, which requires the absolute amount of evidence to be represented in some way, and this is how the *confidence* measurement becomes necessary.

These measurements enrich our understanding of the inference rules. In the truth value functions, we can see that the fundamental difference between deductive inference and non-deductive (such as abductive or inductive) inference is in the confidence (not the frequency) of the conclusion. In deduction, if both premises are completely true, so is the conclusion. However, in abduction and induction, the confidence of the conclusion is much lower in this situation, meaning that the conclusion is tentative even when the premises are certain, and can be revised by new evidence.

In binary logics, such as FOPL, truth value only indicates whether there is negative evidence, without measuring it quantitatively. The concept of positive evidence does not really exist there. As Popper [1959] argued, if knowledge is represented as universal statement in FOPL, it can be falsified by a piece of negative evidence, but cannot be verified by a piece (or even many pieces) of positive evidence.

To ignore quantity of evidence means it will be hard for the system to distinguish hypotheses that have a little of negative evidence from those that have a lot. Even for hypotheses whose all available evidence is positive, the amount of evidence still matters — a hypothesis conformed only once is quite different from a hypothesis conformed a million times. For these reasons, to study abduction in binary logic is not wrong, but not very useful under the assumption of insufficient knowledge and resources.

The difference between NAL and other non-binary approaches is beyond the scope of this paper. For a comparison of NAL and the Bayesian approach (and other probability-based ones), see [Wang, 2001].

4.3 Term logic and predicate logic

As discussed in [Wang, 2000], NAL belongs to the *term logic* tradition, exemplified by Aristotle's logic [Aristotle, 1989]. This kind of logic is different from *predicate logic*, like FOPL, in its subject-predicate form of knowledge representation, and in its syllogistic form of inference rules.

Peirce found the syllogistic form of abduction by switching the conclusion and a premise in the "Barbara" syllogism of Aristotle, though in his later work he moved to an inferential definition of abduction, which is not about a single inference step, but about a complex inference process which consists of many steps, and the process is usually represented in FOPL [Flach and Kakas, 2000; Peirce, 1931; Wang, 2000]. Currently all study of abduction use FOPL or its variations, with NAL as the only exception. For examples, see [Flach and Kakas, 2000; Michalski, 1993].

NAL is designed to be a term logic for several reasons.

For the language, the subject-predicate form is preferred because it provides a natural way to define "evidence" for a statement, as shown in the previous sections. For Inheritance statement, evidence is defined in terms of extensions and intensions of the two terms that form the statement. For Implication statement, evidence is defined in terms of the sufficient conditions and necessary conditions of the two statements that form the statement. In both cases, the transitive nature of Inheritance and Implication, and the subject-predicate form of the statements, are essential. Since these features are not available in predicate logic, it will be very hard, if not impossible, to do similar things in FOPL.

The syllogistic inference rules of term logic also have features not available in predicate logic. As described previously, each inference rule in NAL takes a pair of statements as premises, under the condition that they must share a term (or a statement, for higher-order inference). The conclusion is a statement consisting of the other two (not shared) terms (or statements). This kind of rules has certain properties:

- For a given pair of premises, the conclusions are fully determined, both in contents and in truth values.
- The premises and conclusions are related to each other, both in contents and in truth values.

Though the above properties looks simple and natural, they are not possessed by predicate logic.

In predicate logic, abduction and induction are often defined as "reversed deduction", in the sense that they are inference processes that produce certain hypotheses, which are consistent with background knowledge, and imply the given conclusion (and with some additional properties). Defined in this way, when the background knowledge and conclusion to be implied are given, the hypotheses may not be fully determined, that is, many hypotheses may satisfy the condition. In fact, this is the major reason for some people to separate "hypothesis generation" and "hypothesis evaluation" — in the framework of predicate logic, a deterministic procedure can be found for the latter, but not for the former [Carnap, 1950; Flach and Kakas, 2000; Peirce, 1931; Popper, 1959].

In NAL, since every piece of empirical knowledge has confidence less than 1, it may be revised by future evidence. In this sense, there is no sharp boundary between "hypothesis", "belief", "knowledge", and "fact" — their difference is relative and conventional. For example, we usually use "hypothesis" for a statement with low confidence, and "fact" for one with high confidence, though there is no qualitative difference between the two. Therefore, in NAL all inference rules (including deduction) on such knowledge carry out "hypothesis generation" in the sense that the conclusions may be revised in the future. At the same time, since each rule has a truth-value function that evaluate the conclusion for its evidential support, the rules also perform "hypothesis evaluation".

The second issue, the content relevance among premises and conclusion, will be addressed in the next subsection.

A common criticism of term logic is its "poor expressive power". Current logic textbooks usually refer to term logic as out-of-date, and predicate logic as up-to-date, with the latter cover the former as a special case. Such a judgment is fair for Aristotle's Syllogism, but for a term logic like NAL, it is no longer valid. With the various relations and compounds added into the system, NAL actually has more expressive power than FOPL when the domain is empirical knowledge. Also, NAL has a grammar that is closer to natural languages than FOPL has. However, a detailed discussion on this topic will have to be left for a future publication.

NAL is not generally better than FOPL, but it is better in situations where the system's knowledge and resources are insufficient. In domains where knowledge and resources can be assumed to be sufficient (with respect to the questions to be answered), FOPL may still be better. Mathematics is such a domain. FOPL was originally developed to be used in mathematics. It is used (and often misused) in other domains since no other proper logic has be developed.

NAL can emulate predicate logic (or any other logic) by representing inference rules of that logic as Implication relations, and applying them as a special case of deduction. When a binary logic is emulated in this way, all premises will be given confidence 1 and strength 0 or 1, and only conclusions in such a form are accepted as final result. In this way, abduction and induction cannot be used to derive final results, though they can be used to generate and evaluate hypotheses, so as to guide the deduction processes. Therefore, NAL can be used as the meta-logic of other logics. In other words, NAL can be used as "the logic of operating system", which can support various forms of application programs, each with its own logic, and used in its own domain.

4.4 Relevant and material Implication

Though both NAL and FOPL use Implication to capture the intuitive meaning of the "if ... then ..." structure in natural languages, the definition in the two are quite different.

In NAL, Implication, in its idealized form, is defined to be a reflexive and transitive binary relation between two statements. In its realistic form, it is multi-valued, with its truth value defined according to available evidence. Here evidence is measured by comparing the sufficient and necessary conditions of the two statements.

In FOPL, Implication is defined as a truth-functional structure (called "material implication"). Given arbitrary statements P and Q, $P \rightarrow Q$ is defined as $OR(NOT(P), Q))$.

Though this definition is useful for various purposes, it suffers from the well-known "implication paradox", which says that $P \rightarrow Q$ is true when P is false (Q can be anything) or Q is true (P can be anything). Though logically consistent with FOPL, this result is highly counter-intuitive, and it gives people a feeling that some important thing is missing in the definition of implication in FOPL — P and Q should be somehow *relevant* to each other, which is assumed by the "if ... then ..." structure in natural languages.

A whole branch of logic, relevant logic [Read, 1988], has been developed specially for this issue. Within the framework of predicate logic, it is not easy to find a simple solution. However, in NAL, this problem does not appear in the first place. As mentioned previously, as a term logic, in NAL the premises and conclusion of an inference step must share a common component, otherwise no conclusion can be derived. Specially, in the induction rule of the third inference table, $P \rightarrow Q$ can be derived from P and Q if and only if the two premises are based on the same (implicitly represented) evidence, so they are guaranteed to be relevant to each other in their contents. From an arbitrary pair of statements, nothing can be derived — to know their truth values is not enough. Even when $P \rightarrow Q$ can be derived from P and Q, its confidence is low, because the rule is induction. Only when P and Q have been repeatedly supported by the same evidence for many times (and the evidence is different at each time), can $P \rightarrow Q$ then become more confident (by merging the individual conclusions with the revision rule).

Defined in this way, the Implication relation in NAL is closer to the intuitive meaning of "if ... then ..." than its FOPL counterpart is. Consequently, it works better in causal inference. In NAL, "causal relation" is not defined as a "built-in" relation, as Implication and Inheritance. Instead, it is a "user-defined" relation discussed previously, whose meaning is learned by the system from its experience, and may change from context to context. Even so, various forms of causal relations should all share Implication as their invariable core, though the other properties of the relation may change. This is the case because the essence of causal inference is to predict the future according to the past.

As mentioned before, in NAL abduction often plays the role of explanation. The Implication define above guarantees that the explanation generated in NAL not only satisfies truth value restriction (be consistent with background knowledge, and can derive given conclusion by deduction), but also satisfies content restriction — the explanation is always relevant to the conclusion.

4.5 Further works in NAL

The components of NAL that are currently under development include:

Compound terms and statements: For two given terms, sometimes the system will form their *intersection*, *union*, or *difference*. For two given statements, sometimes the system will form their *conjunction*, *disjunction*, or *negation* of either of them. There are corresponding inference rules, with their truth value functions attached, that derive statements with compound terms or compound statements as components.

Variables: Variables will be introduced to enable more powerful and flexible knowledge representation. Corresponding inference rules should properly handle variable generation, unification, and instantiation.

Procedural interpretation: By given certain relations a procedural interpretation (as in logic programming), the higher-order inference mechanism can be used to do planning. Furthermore, if certain relations are grounded on executable operations in NARS, the system can execute plans and learn skills, all through inference.

After these components are added into NAL, it should be able to provide a consistent and comprehensive logical foundation for Artificial Intelligence.

References

[Aliseda, 2000] A. Aliseda, Abduction as Epistemic Change: A Peircean model in artificial intelligence. In P. Flach and A. Kakas, (eds.), *Abduction and Induction: Essays on their Relation and Integration*, Chapter 3, pages 45–58. Kluwer Academic Publishers, Dordrecht, 2000.

[Aristotle, 1989] Aristotle, *Prior Analytics*, Translated by R. Smith. Hackett Publishing Company, Indianapolis, Indiana, 1989.

[Carnap, 1950] R. Carnap, *Logical Foundations of Probability*. Chicago University Press, Chicago, 1950.

[Flach and Kakas, 2000] P. Flach and A. Kakas, Abductive and Inductive Reasoning: Background and Issues. In P. Flach and A. Kakas, (eds.), *Abduction and Induction: Essays on their Relation and Integration*, Chapter 1, pages 1–27. Kluwer Academic Publishers, Dordrecht, 2000.

[Josephson and Josephson, 1994] J. Josephson and S. Josephson, (eds.), *Abductive Inference: Computation, Philosophy, Technology*. Cambridge University Press, Cambridge, 1994.

[Michalski, 1993] R. Michalski, Inference Theory of Learning as a Conceptual Basis for Multistrategy Learning. *Machine Learning*, Vol. 11, pages 111–151, 1993.

[Peirce, 1931] C. Peirce, *Collected Papers of Charles Sanders Peirce*, Vol. 2. Harvard University Press, Cambridge, Massachusetts, 1931.

[Popper, 1959] K. Popper, *The Logic of Scientific Discovery*. Basic Books, New York, 1959.

[Poole, 2000] D. Poole, Learning, Bayesian Probability, Graphical Models, and Abduction. In P. Flach and A. Kakas, (eds.), *Abduction and Induction: Essays on their Relation and Integration*, Chapter 10, pages 153–168. Kluwer Academic Publishers, Dordrecht, 2000.

[Psillos, 2000] S. Psillos, Abduction: Between Conceptual Richness and Computational Complexity. In P. Flach and A. Kakas, (eds.), *Abduction and Induction: Essays on their Relation and Integration*, Chapter 4, pages 59–74. Kluwer Academic Publishers, Dordrecht, 2000.

[Read, 1988] S. Read, *Relevant Logic*. Blackwell, Oxford, 1988.

[Wang, 1994] P. Wang, From Inheritance Relation to Non-axiomatic Logic. *International Journal of Approximate Reasoning*, Vol. 11, pages 281–319, 1994.

[Wang, 1995] P. Wang, *Non-Axiomatic Reasoning System: Exploring the Essence of Intelligence*. PhD thesis, Indiana University, 1995. Available on the web at the author's homepage.

[Wang, 2000] P. Wang, Unified Inference in Extended Syllogism. In P. Flach, A. Kakas, (eds.), *Abduction and Induction: Essays on their Relation and Integration*, Chapter 8, pages 117–129. Kluwer Academic Publishers, Dordrecht, 2000.

[Wang, 2001] P. Wang, Confidence as Higher-order Uncertainty. to appear in the proceedings of *The Second International Symposium on Imprecise Probabilities and Their Applications*, Ithaca, New York, June 2001.