The Generation and Evaluation of Generic Sentences

Pei Wang
Department of Computer and Information Sciences
Temple University, Philadelphia, PA 19122, USA
http://www.cis.temple.edu/∼pwang/

Abstract

A new logic is introduced, and applied to the generating, evaluating, using, and revising of generic sentences. This logic is designed for an adaptive system working with insufficient knowledge and resources. It is different from existing logics by using a term-oriented language, an experience-grounded semantics, and a set of syllogistic inference rules. It is argued that generic sentences should be seen as summarized experience, which are represented as multi-valued statements, evaluated according to available evidence, and generated by various types of inference rules.

1 Generic sentences and default rules

In the semantic analysis of natural languages, a "generic sentence" refers to a sentence like "Birds fly", which is general, usually accepted as true in spite of the existence of counterexamples. The focus of the research is on the exact meaning and truth value of this type of sentence [Carlson and Pelletier, 1995, Cohen, 1999].

In the research of artificial intelligence (AI), related work has been carried out as part of *commonsense reasoning* and *non-monotonic logic*, and the sentences are usually called "default rules". Beside the semantic properties of this type of sentences, in AI people are also interested in using this kind of rules to derive tentative conclusions [Reiter, 1980, Davis and Morgenstern, 2004].

The consensus reached in the above research is that a generic sentence (or default rule) is *normally*, *usually*, or *typically* true, so it tolerates exceptions, and is valid to be used in reasoning if the current case is not known to be an exception. If a conclusion derived in this way is refuted by a new fact, it should be considered as false, though the default rule remains valid. For example, from a fact "Tweety is a bird" and a default rule "Birds fly", It is valid to derive a conclusion "Tweety fly". If later it turns out that Tweety cannot fly, the conclusion is revised, but the premises remain the same.

Though some properties of generic sentences can be explained in this way, the above approach has problems in handling certain aspects of genetic sentences.

One problem is their generation. At the current time, the default rules in AI systems are usually provided by the designer. If genetic sentences are about normal situations, how can we (and computer systems) judge whether a given sentence is normally true or not? Do genetic sentences represent the primitive properties of kinds, or come eventually from the individuals in the kinds? [Cohen, 1999] Is there an algorithm of this procedure that can be implemented in a computer? If "Birds fly" comes from induction from flying birds, how many observations are enough to support the conclusion?

A related issue is the revision of the default rules. If our judgments about whether a situation is tentative and fallible, when do we realize that what was seen as exceptions in the past should now be taken as the normal situation? For example, if the environment gets polluted to such an extent that some future generations of birds are poised and lose their flying capacity, can we still take "Birds fly" as a true generic sentence?

Furthermore, even if we have got a set of default rules, there is no guarantee that the conclusions derived from them will be consistent. A well known example of this problem is the "Nixon Diamond": from a fact "Nixon is a Quaker" and a generic sentence "Quakers are pacifists", the conclusion is "Nixon is a pacifist"; from a fact "Nixon is a Republican" and a generic sentence "Republican are not pacifists", the conclusion is "Nixon is not a pacifist". To solve this contradiction, the priority of each default rule needs to be specified and compared [Brewka, 1994].

A special case of the above situation is when the related categories, about which the generic sentences are stated, have a subset relationship among them. For example, from a fact "Tweety is a bird" and a generic sentence "Birds fly", the conclusion is "Tweety flies"; from a fact "Tweety is a penguin" and a generic sentence "Penguins do not fly", the conclusion is "Tweety does not fly". Again there is a contradiction. Since "penguin" is a subset of "bird", people usually give the second conclusion a higher priority, according to the "specificity principle" [Touretzky, 1986]. However, even this situation is not as simple as it looks. As argued in [Wang, 1995b], specificity is just one of several factors influencing the handling of conflicting conclusions, and it is not always the dominating factor.

In this paper, we do not plan to review all the solutions proposed for the above problems. Instead, a new logic is introduced, which provides unified solutions to the above problems and many other problems in logic and AI.

The logic to be discussed is NAL (Non-Axiomatic Logic). It is designed for an adaptive reasoning system working with insufficient knowledge and resources [Wang, 1995a]. In the following we will focus on the aspects of the logic that is directly related to genetic sentences. For the other issues, please visit the author's website for publications and demonstrations.

NAL uses a formal language, Narsese, for knowledge representation, and a set of inference rules to derive new knowledge from given knowledge, as well as to answer questions according to available knowledge. The system works with "insufficient knowledge and resources", in the sense that it is open to all possible new knowledge and questions (expressible in Narsese, of course), and the questions must be answered in real time. It means that new knowledge may conflicts with old knowledge, a question may go beyond the current knowledge of the system, and the system cannot afford the time expense of exhaustively using its knowledge on every question.

To work under this restriction, NAL uses a term-oriented language, an experience-grounded semantics, and a set of syllogistic inference rules.

2 Inheritance judgments

The basic form of knowledge in NAL is an *inheritance statement* "$S \rightarrow P$", where S is the *subject term*, P is the *predicate term*, and "$\rightarrow$" is the inheritance relation. In the simplest case, a term is an identifier, like "bird", "animal", and "fly".

The inheritance relation is defined by *being reflexive and transitive*. According to this definition, for any term X, "$X \rightarrow X$" is always true (reflexivity). Also, if "$X \rightarrow Y$" and "$Y \rightarrow Z$" are true, so is "$X \rightarrow Z$" (transitivity). Intuitively, "$S \rightarrow P$" indicates that "S is a specialization of P, and P is a generalization of S". For example, "Birds are animals" can be represented as "$bird \rightarrow animal''$.

Roughly speaking, the system's *experience* is the knowledge it gets from the environment. In an idealized situation, the experience consists of inheritance statements defined above. Given experience K, the system's *beliefs*, K^*, is the transitive closure of K, excluding statements whose subject and predicate are the same term.

With respect to a given K, the *truth value* of a statement is either *true* or *false*. Given experience K, the truth value of a statement is true if it is in K^*, or has the form of $T \rightarrow T$, otherwise it is false.

Given experience K, let the set of all terms appearing in K to be the *vocabulary* of the system, V_K. Then, the *extension* of a term T is the set of terms $T^E = \{x \mid x \in V_K \wedge x \rightarrow T\}$. The *intension* of T is the set of terms $T^I = \{x \mid x \in V_K \wedge T \rightarrow x\}$. Intuitively, the extension of a term includes its specializations, and its intension includes its generalizations. The *meaning* of a term T consists of its extension and intension.

The above definitions set up an idealized situation, where the experience consists of binary statements, and terms have well-defined meanings. In the following, this idealized situation is extended into a practical situation, where statements are multi-values, and terms have fuzzy meanings.

For an inheritance statement "$S \rightarrow P$" and a term M, if both "$M \rightarrow S$" and "$M \rightarrow P$" are true, it is positive evidence for the statement; if "$M \rightarrow S$" is true but "$M \rightarrow P$" is false, it is negative evidence. Symmetrically, if both "$P \rightarrow M$" and "$S \rightarrow M$" are true, it is positive evidence for the statement; if "$P \rightarrow M$" is true but "$S \rightarrow M$" is false, it is negative evidence. Evidence is defined in this way, because as far as a term in positive evidence is concerned, the inheritance statement is correct; as far as a term in negative evidence is concerned, the inheritance statement is incorrect.

Since according to the previous definition, terms in the extension or intension of a given term are equally weighted, the amount of evidence can be simply measured by the size of the corresponding set. Therefore, for "$S \rightarrow P$," the amount of positive, negative, and total evidence is, respectively,

$$\begin{aligned} w^+ &= |S^E \cap P^E| + |P^I \cap S^I| \\ w^- &= |S^E - P^E| + |P^I - S^I| \\ w &= w^+ + w^- \\ &= |S^E| + |P^I| \end{aligned}$$

When comparing competing beliefs and deriving new conclusions, we usually prefer *relative measurements* to *absolute measurements*. Also, it is often more convenient for the measurements to take values from an interval, while the amount of evidence has no upper bound.

A natural relative measurement for uncertainty is the *frequency* of positive evidence among all available evidence. In NAL, the "frequency" of a statement is defined as

$$f = w^+/w$$

When $w = 0$ (and therefore $w^+ = 0$), f is defined to be 0.5.

A *confidence* measurement is used in NAL to indicate the stability of a frequency evaluation, by comparing the amount of available evidence to the future evidence of a unit amount. Therefore, we have

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

That is, confidence is the percent of the amount of evidence the system currently has among the amount of evidence the system will have in the near future.

Together, f and c form the *truth value* of a statement in NAL. A statement plus its truth value is called a "judgment" in NAL.

Given the above definition of frequency, after the coming of evidence of the unit amount, the new f value will be in the interval

$$[w^+/(w+1),\ (w^+ + 1)/(w+1)]$$

This is because the current frequency is w^+/w, so in the "best" case, when all evidence in the near future is positive, the new frequency will be $(w^+ + 1)/(w+1)$; in the "worst" case, when all evidence in the near future is negative, the new frequency will be $w^+/(w+1)$.

The interval representation of uncertainty provides a mapping between the "accurate representation" and the "inaccurate representation" of uncertainty, because "inaccuracy" corresponds to willingness to change a value within a certain range.

Within the system, it is necessary to keep an accurate representation of the uncertainty for statements, but it is unnecessary for communication purposes. To simplify communication, uncertainty can be represented by a verbal label.

In this situation, the truth value corresponds to the relative ranking of the label in the label set.

If in a language there are only N words that can be used to specify the uncertainty (or some kind of degree) of a statement, and all numerical values are equally possible, the most informative way to communicate is to evenly divide the [0, 1] interval into N sections: [0, 1/N], [1/N, 2/N], ..., [(N-1)/N, 1], and use each label for each section.

For example, if the system has to use a language where "false," "ambivalent" and "true" are the only valid words to specify truth value, and it is allowed to say "I don't know," then the most reasonable approach for input is to map the three words into frequency intervals [0, 1/3], [1/3, 2/3], and [2/3, 1] (corresponding to truth values $<0, 2/3>$, $<1/2, 2/3>$, and $<1, 2/3>$), respectively, and ignore all "I don't know". For output, all judgments whose confidence is lower than 1/3 become "I don't know". and for the others, one of the three words is used, according to the section in which the frequency of the judgment falls.

Given experience K (as a finite set of binary inheritance statements), for an inheritance relation "$S \rightarrow P$" derived from it, w is always finite. Also, since the system has no need to keep statements for which there is no evidence, w should be larger than 0. For uncertainty represented in the other two forms, these translate into $0 < c < 1$ and $l < u$, $u - l < 1$.

Beyond the above normal values of uncertainty, there are two limit cases useful for the interpretation of uncertainty and the design of inference rules:

Null evidence: This is represented by $w = 0$, $c = 0$, or $u - l = 1$. It means that the system knows nothing at all about the statement.

Full evidence: This is represented by $w = \infty$, $c = 1$, or $l = u$. It means that the system already knows everything about the statement — no future modification of the uncertainty value is possible.

Though the above values never appear in actual beliefs of the system, they play important role in system design. Especially, when both frequency and confidence are 1, the judgment "$S \rightarrow P < 1, c >$" represents an inheritance statement "$S \rightarrow P$", which is the one used to define the idealized experience. In this way, the semantics of Narsese is defined by a subset of the language. Given the experience of the system K consisting of binary inheritance statements, the truth value of a Narsese judgment, with subject and predicate in V_K, can be determined by comparing the meaning of the two terms. All these judgments form the beliefs of the system, K^*.

Similarly, we extend the concept of "meaning". For a system whose beliefs are represented in Narsese, the meaning of a term still consists of the term's extensional and intensional relations with other terms. The only difference is that the definition of extension and intension is modified as follows: A judgment "$S \rightarrow P < f, c >$" states that S is in the extension of P and that P is in the intension of S, with the truth value of the judgment specifying their degrees of membership.

Consequently, extensions and intensions are no longer ordinary sets with well-defined boundaries as in the idealized situation. They are similar to fuzzy sets [Zadeh, 1965], because terms belong to them to different degrees. What makes them different from fuzzy sets is how the "membership" is measured (in NAL, two numbers are used) and interpreted (in NAL, it is experience-grounded) [Wang, 1996].

3 Inference rules

NAL uses *syllogistic* inference rules. A typical syllogistic rule takes two judgments sharing a common term as premises, and derives a conclusion, which is a judgment between the two unshared terms. For inference among inheritance judgments, there are three possible combinations if the two premises share exactly one term:

$$\begin{array}{l} \{M \rightarrow P <f_1, c_1>,\ S \rightarrow M <f_2, c_2>\} \vdash S \rightarrow P <F_{ded}> \\ \{M \rightarrow P <f_1, c_1>,\ M \rightarrow S <f_2, c_2>\} \vdash S \rightarrow P <F_{ind}> \\ \{P \rightarrow M <f_1, c_1>,\ S \rightarrow M <f_2, c_2>\} \vdash S \rightarrow P <F_{abd}> \end{array}$$

In each of these rules, the two premises come with truth values $<f_1, c_1>$ and $<f_2, c_2>$, and the truth value of the conclusion, $<f, c>$, is a function of them — according to the experience-grounded semantics, the truth value of the conclusion is evaluated with respect to the evidence provided by the premises. Following the work of Peirce [Peirce, 1931], the three rules above are taken as corresponding to *deduction*, *induction*, and *abduction*, respectively, as indicated by the names of the truth-value functions.

These truth-value functions are designed in the following procedure:

1. 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.

2. Represent the variables in the conclusion as Boolean functions of those in the premises, satisfying the above conditions.

3. Extend the Boolean operators into real number functions defined on [0, 1] in the following way:

$$\begin{array}{rcl} not(x) & = & 1 - x \\ and(x_1, ..., x_n) & = & x_1 * ... * x_n \\ or(x_1, ..., x_n) & = & 1 - (1 - x_1) * ... * (1 - x_n) \end{array}$$

4. Use the extended operators, plus the relationship between truth value and amount of evidence, to rewrite the functions as among truth values (if necessary).

The *deduction* rule in NAL extends the transitivity of idealized inheritance. For the frequency of the conclusion, f, it is 1 if and only if both premises have frequency 1. As for the confidence of the conclusion, c, it reaches 1 only when both premises have truth-values $< 1, 1 >$. Therefore, the Boolean function we get for deduction is

$$f = and(f_1, f_2), \ \ c = and(f_1, c_1, f_2, c_2)$$

which leads to truth-value function

$$F_{ded} : \ f = f_1 f_2, \ \ c = f_1 c_1 f_2 c_2$$

The deduction rule is symmetric to the premises, that is, their order does not matter.

In NAL, *abduction* is the inference that, from a shared element M of the *intensions* of S and P, determines the truth value of "$S \rightarrow P$," and *induction* is the inference that, from a shared element M of the *extensions* of S and P, determines the truth value of "$S \rightarrow P$." Therefore, derived from the duality of extension and intension, we have a duality of abduction and induction in NAL.

In both cases, the premises provide a piece of positive evidence with a unit amount if and only if both of them have truth-value $< 1, 1 >$, which can be represented as a Boolean function

$$w^+ = and(f_1, c_1, f_2, c_2)$$

For the total amount of evidence, in abduction we get

$$w = and(f_1, c_1, c_2)$$

and in induction we get

$$w = and(c_1, f_2, c_2)$$

Please note that in the above representation we are mixing two forms of uncertainty measurement: in the premises, the truth values are used, while in the conclusion, the amounts of (positive/total) evidence are used. Also, in these two rules, the two premises play different roles, and their order matters.

After rewriting the result as truth-value functions, we get

$$F_{abd} : \ f = f_2, \ c = f_1 c_1 c_2 / (f_1 c_1 c_2 + k)$$

$$F_{ind} : \ f = f_1, \ c = f_2 c_1 c_2 / (f_2 c_1 c_2 + k)$$

In the process of designing truth-value function for induction (and abduction), a crucial point is to see that when the premises are $\{M \rightarrow P, \ M \rightarrow S\}$, it is the term M (as a whole) that is taken as evidence, and the amount of evidence it can provide is less than 1. A mistake easy to make here is to think of M as a set of evidences for "$S \rightarrow P$," and think of the number of instances in M as the amount of evidence of the conclusion. That interpretation is inconsistent with the semantics of Narsese.

When two premises contain the same statement, but come from different sections of the experience, the *revision* rule is applied to get a summarized conclusion:

$$\{S \rightarrow P <f_1, c_1>,\ S \rightarrow P <f_2, c_2>\} \vdash S \rightarrow P <F_{rev}>$$

From the additivity of the amount of evidence and the relation between truth value and amount of evidence, the following function is derived:

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

The revision rule produces more confident conclusions from less confident conclusions. In this way, general patterns repeatedly appear in the experience can be recognized and learned.

If the two premises containing the same statement come from overlapping sections of the experience, they cannot be used for revision, otherwise some evidence will be repeatedly counted. Instead, the *choice* rule selects the one with a higher confidence value.

The choice rule is also used when two judgments contain different statements, but compete as answers for a question. For example, "Ravens are birds" and "Penguins are birds" compete as answers for question "What are (the best examples of) birds?" In this case, the one with a higher *expectation* is chosen, and expectation is defined as a function of frequency and confidence:

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

In NAL, a question is typically answered in a process consisting of multiple inference steps, with the conclusion of one rule used as the premise of another rule.

4 Compound terms

The inheritance statement "$S \rightarrow P$" is the basic form of statement in Narsese, but it is not the only form. In addition to it, the language contains other types of statements, which are all built upon this basic form.

Derived inheritance relations: Beside the inheritance relation defined previously, Narsese also includes several of its variations:

- The *similarity* relation "$\leftrightarrow$" is symmetric inheritance. For example, "$raven \leftrightarrow crow$" means "Raven is similar to crow".
- The *instance* relation "$\circ\!\!\rightarrow$" is an inheritance relation where the subject term is treated as an atomic instance of the predicate term. For example, "$Tweety \circ\!\!\rightarrow bird$" means "Tweety is a bird".
- The *property* relation "$\rightarrow\!\!\circ$" is an inheritance relation where the predicate term is treated as a primitive property of the subject term. For example, "$bird \rightarrow\!\!\circ fly$" means "Birds fly".

Sets: Though in general a term is not a set, in NAL there are special terms for sets:

- The *extensional set* $\{T\}$ is a term with T as the only instance. Consequently, "Tweety is a bird" can also be represented as "$\{Tweety\} \rightarrow bird$".
- The *intensional set* $[T]$ is a term with T as the only property. Consequently, "Birds fly" can also be represented as "'$bird \rightarrow [fly]$".

Compound terms: In inheritance statements, the terms not only can be simple terms (as in the above examples), but also can be compound terms formed by other terms with a logical operator. For example, if A and B are terms, then

- their *extensional intersection* $(A \cap B)$ is a compound term, initially defined by $(A \cap B)^E = (A^E \cap B^E)$ and $(A \cap B)^I = (A^I \cup B^I)$,
- their *intensional intersection* $(A \cup B)$ is a compound term, initially defined by $(A \cup B)^I = (A^I \cap B^I)$ and $(A \cup B)^E = (A^E \cup B^E)$.

Right after a compound term is created, its meaning is determined according to its logical relation with its components (its "definition"), but as soon as the system begins to get (input or derived) judgments on the compound, they also contribute to its meaning, and such contributions cannot always be reduced to its components. Therefore the *principle of compositionality* in semantics is only partially true in NAL.

Ordinary relation: In Narsese, only the inheritance relation and its variations are defined as logic constants that are directly recognized by the inference rules. All other relations are converted into inheritance relations with compound terms. 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):

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

Higher-order term: In Narsese, a statement can be used as a term, and called a "higher-order" term. For example, "Birds are animals" is represented by statement "$bird \rightarrow animal$", and "People know that birds are animals" is represented by statement "$(bird \rightarrow animal) \circ\!\!\rightarrow (\perp\ know\ people\ \diamond)$", where the subject term is a statement. Compound higher-order terms are also defined: if A and B are higher-order terms, so do their negations ($\neg A$ and $\neg B$), disjunction ($A \vee B$), and conjunction ($A \wedge B$).

Higher-order relation: Higher-order relations are the relations whose subject term and predicate term are both higher-order terms. In Narsese, there are two of them defined as logic constants:

- *implication*, "$\Rightarrow$", which intuitively corresponds to "if-then", and is defined as isomorphic to *inheritance*, "$\rightarrow$";
- *equivalence*, "$\Leftrightarrow$", which intuitively corresponds to "if-and-only-if", and is defined as isomorphic to *similarity*, "$\leftrightarrow$".

For each type of the above terms/statements, its meaning/truth-value is defined similarly to how we define meaning/truth-value for term/statement in inheritance relation. There are inference rules taking these statements as premises or conclusions. Detailed discussion about them is beyond the scope of this paper. They are mentioned here merely to show that Narsese does not only have inheritance statements.

5 Generic sentences in NAL

Generic sentences are usually represented as inheritance statements in Narsese. For example:

- "Birds are animals" can be represented as
"$bird \rightarrow animal$".
- "Birds fly" can be represented as
"$bird \rightarrow [fly]$".
- "Ravens are black birds" can be represented as
"$raven \rightarrow ([black] \cap bird)$".
- "Crows are smaller than ravens" can be represented as
"$crow \rightarrow (\perp\ smaller\ \diamond\ raven)$".

In these sentences, the subjects are atomic Narsese terms, which can also be used as predicates, like common nouns (see [Krifka, 1995] for a comparison between Chinese and English common nouns). They can be seen as "kinds", and are "unbounded" [Cohen, 1999] — they can be applied to new situations.

If the subject is an extension set, the statement is no longer generic:

- "Mei Sheng is a giant panda" can be represented as
"$\{Mei\text{-}Sheng\} \rightarrow giant\text{-}panda$".
- "The giant panda is an endangered species" can be represented as
"$\{giant\text{-}panda\} \rightarrow ([endangered] \cap species)$".

The undesired conclusion "Mei Sheng is an endangered species" cannot be derived from these two statements using the deduction rule of NAL, because the predicate of the first sentence and the subject of the second sentence are not the same term.

As mentioned previously, NAL uses a numerical representation of truth value for internal representation. By default, the above sentences will be translated to the above statements with truth value $<1, 2/3>$.

Represented in Narsese, the "Nixon Diamond" consists of the following judgments:

(1) $\{Nixon\} \rightarrow Quaker <1, 2/3>$

(2) $Quaker \rightarrow pacifist <1, 2/3>$

(3) $\{Nixon\} \rightarrow Republican <1, 2/3>$

(4) $Republican \rightarrow pacifist <0, 2/3>$

From (1) and (2), by deduction, we get

(5) $\{Nixon\} \rightarrow pacifist <1, 4/9>$

Similarly, from (3) and (4), by deduction, there is

(6) $\{Nixon\} \rightarrow pacifist <0, 4/9>$

From the conflicting judgments (5) and (6), by revision, NAL gets

(7) $\{Nixon\} \rightarrow pacifist <1/2, 8/13>$

Therefore, the given knowledge provides the same amounts of positive and negative evidence to the conclusion, so to the question "Is Nixon a pacifist?", the system has no preference between a "Yes" answer and a "No" answer.

However, if the system has other related beliefs before getting the above sentences, the situation will be different. If the system thinks that Nixon is more typically a Republican than a Quaker, the final conclusion will have a frequency less than 1/2. If the system knowns some Republicans that are pacifists, the final conclusion will have a frequency more than 1/2. So in general, the result will show the overall effect of the related factors.

The problem of Tweety has a similar structure: from "Tweety is a bird" and "Birds fly", "Tweety flies" is derived; from "Tweety is a penguin" and "Penguins do not fly", "Tweety cannot fly" is derived. However, here the situation is different from the previous one, because we also assume the system knows that "Penguins are birds". Under the assumption that "Tweety is a bird" is derived partially from it and "Tweety is a penguin", the two conflicting conclusions are based on overlapping evidence, so cannot be summarized through revision. Instead, the choice rule picks the more confident conclusion as the answer. Since "Tweety" has more properties of "penguin" than of "bird", we can expect "Tweety cannot fly" to have a higher confidence.

Furthermore, "Tweety", being a bird that cannot fly, is recognized as a piece of negative evidence for "Birds fly", and as a result, the frequency of the latter statement is decreased, though it is still much larger than 1/2. In this way, a "default rule" will be weakened by counter examples, until the total amount of negative evidence reaches the amount of positive evidence. After that, the negation of the statement (such as "Birds cannot fly") may become a rule, with a truth value corresponding to the available evidence.

To generate generic sentences in the first place, in NAL there are several possibilities. "Robins fly" can be derived from "Robins are birds" and "Birds fly" by deduction; "Birds fly" can be derived from "Tweety is a bird" and "Tweety flies" by induction; "Robins are birds" can be derived from "Birds fly" and "Robins fly" by abduction. Induction and abduction usually produce conclusions in pair, and each may have a different truth value. For example, the above abduction also generate "Birds are robins". Later, however, this sentence will get more negative evidence than "Robins are birds". Of course, all the above derivations include truth value calculations. In general, from premises with the same truth values, deductive conclusions have higher confidence than inductive and abductive conclusions. Judgments with the same content can be summarized through revision to get more confident conclusions.

What is the threshold of truth value for a judgment to be accepted as a default rule? In NAL, there is such a line, and "acceptance" is a matter of degree. All the judgments are evaluated according to available evidence, and every of them may be involved in future inference processes, though some of them may contribute more than the others to the final results.

6 Discussions

The above description shows that the major components of NAL are fundamentally different from that of conventional reasoning systems. To discuss these differences in detail is beyond the scope of this paper (such discussions can be found in the previous publications on this research). In this paper, we only address the aspects of NAL that are directly related to generic sentences and default rules.

A key difference in representation between NAL and the previous approaches on generic sentences and default rules [Carlson and Pelletier, 1995, Cohen, 1999, Reiter, 1980, Davis and Morgenstern, 2004] is that NAL is a multi-valued logic, where the truth value of each statement is a pair of real numbers. This representation is based on the empiricist belief that knowledge is a summary of experience. This is why general statements are valuable for an adaptive system, even if there are counter-examples.

To indicate whether an observed relation can be generalized to similar situations, a measurement is necessary — the system needs to know how often a given relation can be observed. Therefore, "normally" is closely relation to "usually", and without a numerical measurement, it is difficult to decide whether a statement is "usually true".

Though NAL uses numerical truth values internally, it allows binary sentences in its input and output, as approximations of the internal beliefs. A common criticism to the numerical approaches in reasoning is that "The information necessary to assign numerical probabilities is not ordinarily available" [McCarthy and Hayes, 1969]. In NAL, a rough mapping is used between the numerical and verbal representation of truth value, because what is important here is not the *accuracy* of the representation, but a representation supporting the related inference. In everyday activities, even when we attach the same word, such as "likely" to two sentences, we often still be able to say that one is "more likely" than the other. This phenomenon hints that there are details in internal representation (beliefs) that is lost in external representation (sentences). Exactly where each word falls in the numerical range is not that crucial — different people often have different habit in using these words, which is fine as far as the usage is consistent, and not too far away from the convention of the language community.

For a general-purpose system like NAL that is open to all possible new evidences, a numerical measurement of uncertainty is necessary for the internal representation of the system, not because it is accurate, but because it is uniform and simple, especially because the system needs to have rules for induction, abduction, and so on. In these types of inference, uncertainty emerges even if all the premises are certain, and the amount of evidence is a dominant factor for the processing of the uncertainty. For this reason, it is still desired to represent uncertainty numerically in the internal representation, even if verbal labels of uncertainty can be used in the interface language (as explained when discussing the "frequency interval") for the sake of simplicity and naturalness.

According to the experience-grounded semantics of NAL, the truth value of a statement does not measure how close the statement is to the "state of affairs", but how close the statement is to the experience of the system. The truth value of "Birds fly" is determined by how often the birds under consideration fly, as far as the system remembers, not by a comprehensive statistics on the percent of flying birds among all birds in the universe — for an adaptive system, an objective statistics is not as relevant as its past experience. For someone lives in an island with a black swan, the next swan to meet is more likely to be black, though it is known that there are much more white swan in the world.

This semantics provides a justification for non-deductive inference rules (induction, abduction, and so on) — these rules are "truth-preserving" in the sense that the conclusion is supported by the evidence provided by the premises. As a result, general conclusions can be generated through induction, as well as through deduction and abduction, and be revised by new evidence. The previous conclusions about the invalidity of induction [Hume, 1748, Popper, 1959] basically says that "objective truth" cannot be obtained from finite evidence, but the problem disappears after we realize that for an adaptive system, "true" in empirical knowledge can, and only can, be defined with respect to the past experience.

The truth value measurement of NAL is intuitively related to probability theory, but is not derived from it. This is the case because a direct application of probability theory assumes sufficient knowledge about the probability distribution of related items [Wang, 2001, Wang, 2004], so is inconsistent with the assumption of insufficient knowledge and resources.

Almost all of the previous works on generic sentences are based on first-order predicate logic, which was developed originally for the formalizing of mathematical reasoning [Frege, 1970, Whitehead and Russell, 1910]. In mathematics, the knowledge and resources of the system can be assumed to be sufficient (with respect to the problems to be solved). However, this is not the case in empirical reasoning. Since the "generic sentences" under discussion is mainly a issue in empirical reasoning, it is improper to base the work on a mathematical logic or its extensions.

The NAL approach for generic sentences is characterized by term-oriented language, experience-grounded semantics, and syllogistic inference rules. This logic is not designed specially for generic sentences, but for adaptive reasoning under insufficient knowledge and resources. Nevertheless, generic sentences turn out to be a topic handled well by the logic. This is not a surprise, if we see these sentences as generalizations of experience. Under the restriction of knowledge and resources, these generalizations cannot be perfect, but they are still valuable for our life.

Acknowledgment

This research is partially supported by the Project of Key Research Institute of Humanities and Social Sciences in Universities, MOE of China, 02JAZJD720017. Thanks to Zhou Beihai for his comments and suggestions.

References

[Brewka, 1994] Brewka, G. (1994). Reasoning about priorities in default logic. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 940–945, Menlo Park, California. AAAI Press.

[Carlson and Pelletier, 1995] Carlson, G. N. and Pelletier, F. J., editors (1995). *The Generic Book*. The University of Chicago Press, Chicago.

[Cohen, 1999] Cohen, A. (1999). *Think Generic! : The Meaning and Use of Generic Sentences*. CSLI Publications, Stanford, California.

[Davis and Morgenstern, 2004] Davis, E. and Morgenstern, L. (2004). Introduction: Progress in formal commonsense reasoning. *Artificial Intelligence*, 153:1–12.

[Frege, 1970] Frege, G. (1970). Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. In van Heijenoort, J., editor, *Frege and Gödel: Two Fundamental Texts in Mathematical Logic*, pages 1–82. Harvard University Press, Cambridge, Massachusetts.

[Hume, 1748] Hume, D. (1748). *An Enquiry Concerning Human Understanding*. London.

[Krifka, 1995] Krifka, M. (1995). Common nouns: a contrastive analysis of Chinese and English. In Carlson, G. N. and Pelletier, F. J., editors, *The Generic Book*, chapter 11. The University of Chicago Press, Chicago.

[McCarthy and Hayes, 1969] McCarthy, J. and Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In Meltzer, B. and Michie, D., editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, Edinburgh.

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

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

[Reiter, 1980] Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13:81–132.

[Touretzky, 1986] Touretzky, D. (1986). *The Mathematics of Inheritance Systems*. Pitman Publishing, London.

[Wang, 1995a] Wang, P. (1995a). *Non-Axiomatic Reasoning System: Exploring the Essence of Intelligence*. PhD thesis, Indiana University.

[Wang, 1995b] Wang, P. (1995b). Reference classes and multiple inheritances. *International Journal of Uncertainty, Fuzziness and and Knowledge-based Systems*, 3(1):79–91.

[Wang, 1996] Wang, P. (1996). The interpretation of fuzziness. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 26(4):321–326.

[Wang, 2001] Wang, P. (2001). Confidence as higher-order uncertainty. In *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, pages 352–361, Ithaca, New York.

[Wang, 2004] Wang, P. (2004). The limitation of Bayesianism. *Artificial Intelligence*, 158(1):97–106.

[Whitehead and Russell, 1910] Whitehead, A. and Russell, B. (1910). *Principia mathematica*. Cambridge University Press, Cambridge.

[Zadeh, 1965] Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8:338–353.