

## KNOWLEDGE BASED REFINEMENT

Dao Nam Anh

Institute of Information Technology

### 1. Introduction and motivation

The ability of Man beings to classify and appraise objects, situations in order to make resolutions and their aptitude of generating from observations in different spaces, at various moments to have fairly sufficient knowledge about the world in which they live are interesting research problems. In recent years, a number of generation methods utilize a background knowledge. Relying on this, objects and situations are characterized and new concepts are produced [5].

The background knowledge has different abstract levels for these methods. In data-intensive methods of machine learning it exists typically as the knowledge about mathematical logic, and is used in all other methods. In the conceptual clustering [6] the background knowledge includes problems' constrains, properties of attributes, and criterions for evaluating the quality of constructed classifications. The common characteristic of traditional methods is the logically rightful knowledge, which usually must not have reasoning mistake and which cannot correct concepts and rules immediately.

Human knowledge is great, it has a lot of inaccuracies and man must analyze an actually large number of data in restricted time. Therefore he might argue approximately and his knowledge ought to be refined incessantly in order to enlarge the inference's speed: detect and rectify inconsistencies, remove redundancies, cover holes, simplify expert-derived decision rules. The work provides a method of knowledge based refinement in close relation with inference. It proposes a new approach based on the notion of conflicts. A mathematical formulation for resolving conceptual conflicts among rules is developed. The conceptual conflicts are begot in probabilistic reasoning and then they are eliminated in machine learning. Knowledge based refinement is processed as learning from the conflicts' discovery. This

VIỆN CÔNG NGHỆ THÔNG TIN

Typeset by  $\text{\LaTeX}$

can be one of the approaches to represent neural networks [3,8]. The method will be put in the knowledge based systems tools TESOR 2.0 [2] in near time.

## 2. Concept and conceptual conflict

Knowledge is usually represented by concepts and relations between them [4]. The meaning of a concept is given by its relations with other concepts and can be modified by editing these relations. A new concept will be constructed by describing its relations with existing ones. In reliant knowledge bases all concepts are related closely, these can be displayed by a rule or a sequence of rules. When a concept has changed its semantics the meaning of other concepts will be also changed.

In inference a concept can play different roles: data, goal, solution, criterion for estimating the quality of produced categorization. In this work concepts are symbolized by predicates and the relations among concepts are represented by rules.

*Definition 1:* Knowledge base is  $KB = (\mathbf{P}, \mathbf{R})$ , where  $\mathbf{P}$  is a set of predicates, and  $\mathbf{R}$  - a set of rules of the following form

$$p_1 \& p_2 \& \dots \& p_n \rightarrow p'; \quad p_i, p' \in \mathbf{P}, \quad \forall i = \overline{1, n}. \quad (1)$$

Let's have the following aided denotations for definition 1:

$Left(k)$	= $\{p/k \in \mathbf{R}, \forall p \in \mathbf{P}, p \text{ is placed in left part of } k\text{-rule}\};$
$Right(k)$	-Predicate $p'$ - right predicate of $k$ -rule;
$Weight(p, k)$	-Weight of predicate $p$ in $k$ -rule, $Weight(p, k) \in [0, 1];$
$Ver(k)$	-Belief of $k$ -rule or Weight of right predicate, $Ver(k) \in [0, 1];$
$Pro(p)$	-Probability of predicate $p$ in inference, $Pro(p) \in [0, 1];$
$Ref(p)$	= $\{k/\forall k \in \mathbf{R}, p \in Left(k)\};$
$ S $	-Members' number of the set $S$ .

$$RuleSuggestion(k, \lambda, \delta_\lambda) = \begin{cases} True & \text{if } Ver(k) \in [\lambda + \delta_{\lambda/2}, 1] \\ Undefined & \text{if } Ver(k) \in (\lambda - \delta_{\lambda/2}, \lambda + \delta_{\lambda/2}) \\ False & \text{if } Ver(k) \in [0, \lambda - \delta_{\lambda/2}]; \lambda, \delta_\lambda, \lambda \pm \delta_{\lambda/2} \in [0, 1] \end{cases} \quad (2a)$$

$$PredSuggestion(p, k, \lambda, \delta_\lambda) = \begin{cases} True & \text{if } Weight(p, k) \in [\lambda + \delta_{\lambda/2}, 1] \\ Undefined & \text{if } Weight(p, k) \in (\lambda - \delta_{\lambda/2}, \lambda + \delta_{\lambda/2}) \\ False & \text{if } Weight(p, k) \in [0, \lambda - \delta_{\lambda/2}]; \lambda, \delta_\lambda, \lambda \pm \delta_{\lambda/2} \in [0, 1] \end{cases} \quad (2b)$$

$$PredConfirmation(p, \tau, \delta_\tau) = \begin{cases} True & \text{if } Pro(p) \in [\tau + \delta_{\tau/2}, 1] \\ Undefined & \text{if } Pro(p) \in (\tau - \delta_{\tau/2}, \tau + \delta_{\tau/2}) \\ False & \text{if } Pro(p) \in [0, \tau - \delta_{\tau/2}]; \tau, \delta_\tau, \tau \pm \delta_{\tau/2} \in [0, 1] \end{cases} \quad (2c)$$

$D \Rightarrow (p, Pro)$  -Inference's tree of predicate  $p$  with probability  $Pro$  from predicates of  $D$ ;

$(P_1, R_1) \Rightarrow (p, s)$  -Inference's tree of the predicate  $p$  from predicates of  $P_1$  by rules of  $R_1$ :  $P_1 \subseteq \mathbf{P}$ ,  $R_1 \subseteq \mathbf{R}$ ,  $p \in P, \exists P_1 \Rightarrow (p, s)$ ,  $s = PredConfirmation(Right(p))$ ,  $s \in \{True, False, Undefined\}$ ;  $R_1 = \{k/P_1 \Rightarrow (p, Pro(p))\}$ .

*Definition 2:* Knowledge  $KB = (\mathbf{P}, \mathbf{R})$  has a conceptual conflict on a set of predicates  $P_1 \subseteq \mathbf{P}$ , if  $\exists R_1, R_2 \subseteq \mathbf{R}$ ,  $R_1, R_2 \neq \emptyset, R_1 \neq R_2, \exists p \in \mathbf{P}$ ,  $(P_1, R_1) \Rightarrow (p, True)$ ,  $(P_1, R_2) \Rightarrow (p, False)$ ;

Inference is an purposeful process. It aims to find certain solutions in the shortest time possible. Beginning by forward chaining from data concepts or backward chaining from objective concepts, the inference is carried out in order to find the solution's concepts, from that and data concepts the objective concepts can be reduced.

Conceptual recognition is an important part of the inference, whose first act is transforming the received data to the concepts of the existing knowledge base. That means representing and classifying the given data by a private comprehension, from an individual point of view. The representation and classification of the given data are to further referred to find conclusions. Now we introduce a model of conceptual recognition.

### Conceptual recognition model

Given:  $KB = (\mathbf{P}, \mathbf{R})$ ;

a set of data predicates  $D$ ,  $D \subseteq \mathbf{P}$ ,  $PredConfirmation(p, \tau, \delta_\tau) \neq Undefined, \forall p \in D$ ;

$\forall p \in \mathbf{P} \setminus D, PredConfirmation(p, \tau, \delta_\tau) = Undefined$ ;

Find:  $S$ ,  $S \subseteq \mathbf{P}$  -a set of solution predicates, where  $\forall p \in S, \exists D \Rightarrow (p, Pro(p))$ ,  $PredConfirmation(p, \tau, \delta_\tau) \neq Undefined$ .

Machine learning is a process closely united with reasoning and has a purpose of knowledge bases' development and refinement. In order to raise the rapidity of reasoning it is necessary to find continuously more general and exact rules from the existing ones. The traditional method discover common and original characters from a number of objects, situations and give new rules. Let's consider machine learning and inference as paralleled processes.

When the space of knowledge base is very wide and the reasoning time is arranged, learning and inference might have an influential mechanism of concepts' analysis to easily change the depth and the width of inference. Inference is an inductive process and must select the best alternative among several of ways to obtain objectives. The preference is done by estimating concepts which are dependent on inductive levels, times of reference from other concepts to the evaluated concept in the inductive

process, probability and role of the evaluated concept at the moment of reasoning: objective or data concepts.

The probabilistic reasoning can be expressed as follows: In the analysis of a rule, the concept in the right part of the rule may be assured by suggesting all concepts in the left part or only some of them, under the control of evaluation's concepts, which exist as ordinary concepts and which are used in the estimation. Since concepts can be reformed, updated, the evaluation usually does not give the same solutions for the same set of input data. The weights of predicates are used in the representation of reasoning. In the left part of rule each predicate owns defined weight and the rule examines only the predicates that hold a great enough weight. If one predicate modifies its meaning but has not enough weight in left part of some rule, then the right predicate of the rule must convert its semantics but it does not do that because of the insufficient weight of the left one. This is the main reason of the conceptual conflicts. When the knowledge base has conceptual conflicts the inference has the chance to confirm and deny simultaneously the same concept because of the existence of the discord between the rules.

### 3. Discovering and solving conceptual conflicts

Algorithm 1 attempts to represent probabilistic reasoning in the conceptual recognition, where conceptual conflicts can be begot because of the approximation of the reasoning. The algorithm represents the predicates' weights by attribute *Weight* ( $p.k$ ) with the bounds  $\lambda \pm \delta_{\tau/2}$  which determines predicate's analysis in rules. Weights and probabilities are represented as values in interval  $[0,1]$ , divided into three parts to represent meanings: *True*, *False*, *Undefined*. Eq. (2a,2b,2c). The approximation of reasoning is depicted by conditionally selecting left predicates in calculating probability of right predicate, Eq. (4,5).

Conceptual conflicts are discovered by definition 2 that is illustrated by the conditions Eq. (6a,6b). Then they are solved by finding the predicate, that have the biggest belief at the moment of inference. Weights of the predicate in relative rules will be changed in Eq. (9,10). When one predicate is suggested more than once Eq. (7), the weight of the predicate in relative rules will be changed too, Eq. (11,12).

The probabilities' limits  $\tau \pm \delta_{\tau/2}$  decide the continuation of predicate's analysis in inference. Parameters  $\lambda, \tau, \delta_{\lambda}, \delta_{\tau}, o(\delta_{\tau}), \Delta_{\tau}$  just are predicates. Having relations with other predicates they can change values by corresponding rules.

#### Algorithm 1: Conceptual recognition()

```

Conceptual Recognition() {
    while (Exit Active Concepts())
        Update Active Concepts(); }

```

Exit Active Concepts () {

Take a set of predicates  $I, I \subseteq R$ :

$$S = \{k/k \in R, \forall k \in Ref(p), RuleSuggestion(k, \lambda, \delta_\lambda) \neq Undefined, p \in P, PredConfirmation(p, \tau, \delta_\tau) \neq Undefined\} \quad (3)$$

Return  $(!(I \equiv \emptyset));$

Update Active Concepts() {

Take the set of predicates

$$S = \{p/p = Right(k), \forall k \in I, PredConfirmation(p, \tau, \delta_\tau) \neq Undefined\} \quad (4)$$

with probabilities that are determined by the following function:

$$Pro(p) = Ver(k) \sum_{q \in Left(k), PredSuggestion(q, k, \lambda, \delta_\lambda) \neq Undefined} Weight(q) Pro(q) / |Left(k)| \quad (5)$$

if (Exit Conceptual Conflicts()) then Eliminate Conflicts();

if (Exit Many Predicate's Suggestions()) then Increase Predicate's Weight();

Exit Conceptual Conflicts () {

$\exists$  predicate  $p^* \in S, Pro(p^*)$  is calculated more then once and exit at least two defined values of  $Pro(p^*)$ , remarked as  $\{Pro(p^*), \{Pro'(p^*)\}$ , for that

$$\begin{aligned} &PredConfirmation(Pro(p^*), \tau, \delta_\tau) = True \ \& \\ &PredConfirmation(Pro'(p^*), \tau, \delta_\tau) = False \\ &||Pro(p^*) - \tau| - |Pro'(p^*) - \tau|| \leq \Delta_\tau. \end{aligned} \quad (6a)$$

If only condition Eq. (6a) is satisfied then final value of  $Pro(p^*)$  is the value, for that  $|Pro(p^*) - \tau|$  is maximum from all values of  $Pro(p^*)$ . }

Exit Many Predicate's Suggestions() {

$\exists$  predicate  $p^{**} \in S, Pro(p^{**})$  is calculated more then once and its values give the same value by

$$PredConfirmation(Pro(p^{**}), \tau, \delta_\tau) \quad (7)$$

Final value of  $Pro(p^{**})$  is the value, for that  $|Pro(p^{**}) - \tau|$  is maximum from all values of  $Pro(p^{**})$ . }

Eliminate Conflicts () {

Let  $Q$  is a set of nodes - predicates contained in trees:

$$Q = \{p/\forall p \in D \rightarrow (p^*, Pro(p^*)) \cup D \Rightarrow (p^*, Pro'(p^*)), p^* \text{ from Eq.(6)}\} \quad (8)$$

Every note  $p$  is a predicate, reduced by a rule remarked as  $r_p$ , with probability  $Pro(p)$  and joined in left part of other rule  $k$  with weight:  $Weight(p,k)$ . For node  $j \in Q$ , if  $|Pro(p) - \tau| = \max_{p' \in Q \setminus \{p\}} |Pro'(p) - \tau|$  than decrease weight:

$$Weight(p,k) = \begin{cases} \max(Weight(p,k) - O(\delta_\lambda), 0), & \text{if } Weight(p,k) \geq \lambda \\ \min(Weight(p,k) + O(\delta_\lambda), 1), & \text{if } Weight(p,k) < \lambda \end{cases} \quad (9)$$

and lower the belief  $Ver(r_p)$

$$Ver(r_p) = \begin{cases} \max(Ver(r_p) - O(\delta_\lambda), 0), & \text{if } Ver(r_p) \geq \lambda \\ \min(Ver(r_p) + O(\delta_\lambda), 1), & \text{if } Ver(r_p) < \lambda \end{cases} \quad (10)$$

Increase Predicate's Weight() {

For node  $p \in Q$ , if  $|Pro(p) - \tau| = \max_{p' \in Q \setminus \{p\}} |Pro'(p) - \tau|$  than increase weight:

$$Weight(p,k) = \begin{cases} \min(Weight(p,k) + O(\delta_\lambda), 1), & \text{if } Weight(p,k) \geq \lambda \\ \max(Weight(p,k) - O(\delta_\lambda), 0), & \text{if } Weight(p,k) < \lambda \end{cases} \quad (11)$$

and increase the belief  $Ver(r_p)$

$$Ver(r_p) = \begin{cases} \min(Ver(r_p) + O(\delta_\lambda), 1), & \text{if } Ver(r_p) \geq \lambda \\ \max(Ver(r_p) - O(\delta_\lambda), 0), & \text{if } Ver(r_p) < \lambda \end{cases} \quad (12)$$

The algorithm uses a rule's belief (the weight of right predicate of rule) depended on the origin of the rules. The belief of the rule given from other knowledge base is the estimation of the knowledge base for itself. The belief of the rule made from another depends on the beliefs of those partial rules.

Weights, probabilities of predicates are used in the algorithm to represent influential reasoning. That is a method to find conclusions with optimistic conditions concerning exactness and time.

### Example 1.

Now the "Penguin" example [8,9] will be considered in conceptual concepts' approach. Superscript numbers in a rule are weights of predicates and belief of the rule.

$$KB = (\mathbf{P}, \mathbf{R}), \quad \mathbf{P} = \{ Penguin, Bird, Fly, Tweety \}, \\ \mathbf{R} = \{ Penguin^9 \rightarrow Fly^2(1) \rightarrow Bird^9 \rightarrow Fly^9(2); Penguin^1 \rightarrow Bird^9(3);$$

$$Tweety^9 \rightarrow Penguin^1(5); Tweety^8 \rightarrow^1 Bird(6); \} \quad (13)$$

$$\lambda = 0.5, \delta_\lambda = 0.2, \tau = 0.5, \delta_\tau = 0.2, \Delta_\tau = 0.2, o(\delta_\tau) = 0.1. \quad (14)$$

Let's have at the beginning one active predicate:  $Pro(Tweety) = 1$ .

Rules (4), (5) and Eq. (5) give:

$$Pro(Penguin) = 1.0 \times 0.9 \times 1.0 = 0.9$$

$$Pro(Bird) = 1.0 \times 0.8 \times 1.0 = 0.8 \quad (15)$$

$$\text{Rule (3) and Eq. (5) give } Pro(Bird) = 0.9 \times 0.9 = 0.81 \quad (16)$$

Predicate *Bird* satisfies condition Eq. (7). Therefor from (15), (16) the  $Ver(3)$  is increased by Eq. (12)

$$Ver(3) = \min(0.9 + 0.1, 1.0) = 1.0 \quad (17)$$

After applying Eq. (5) for rule (1), rule (2) we have

$$rule(1) : Pro(Fly) = 0.2 \times 0.9 \times 0.9 = 0.162 \quad (18)$$

$$rule(2) : Pro(Fly) = 0.9 \times 0.9 \times 0.81 = 0.658 \quad (19)$$

and conceptual conflict, controled by Eq. (6a,6b) with values from (14). (18). (19):

$$PredConfirmation(0.162, 0.5, 0.2) = False \ \&$$

$$PredConfirmation(0.658, 0.5, 0.2) = True$$

$$\|0.162 - 0.5\| - \|0.658 - 0.5\| = 0.180 < 0.2 \quad (20)$$

As predicate *Fly* satisfies condition Eq. (6a,6b), from (18), (19), (20) the  $Weight(Bird,2)$  and  $Ver(3)$  is decreased by Eq. (9,10) correspondingly:

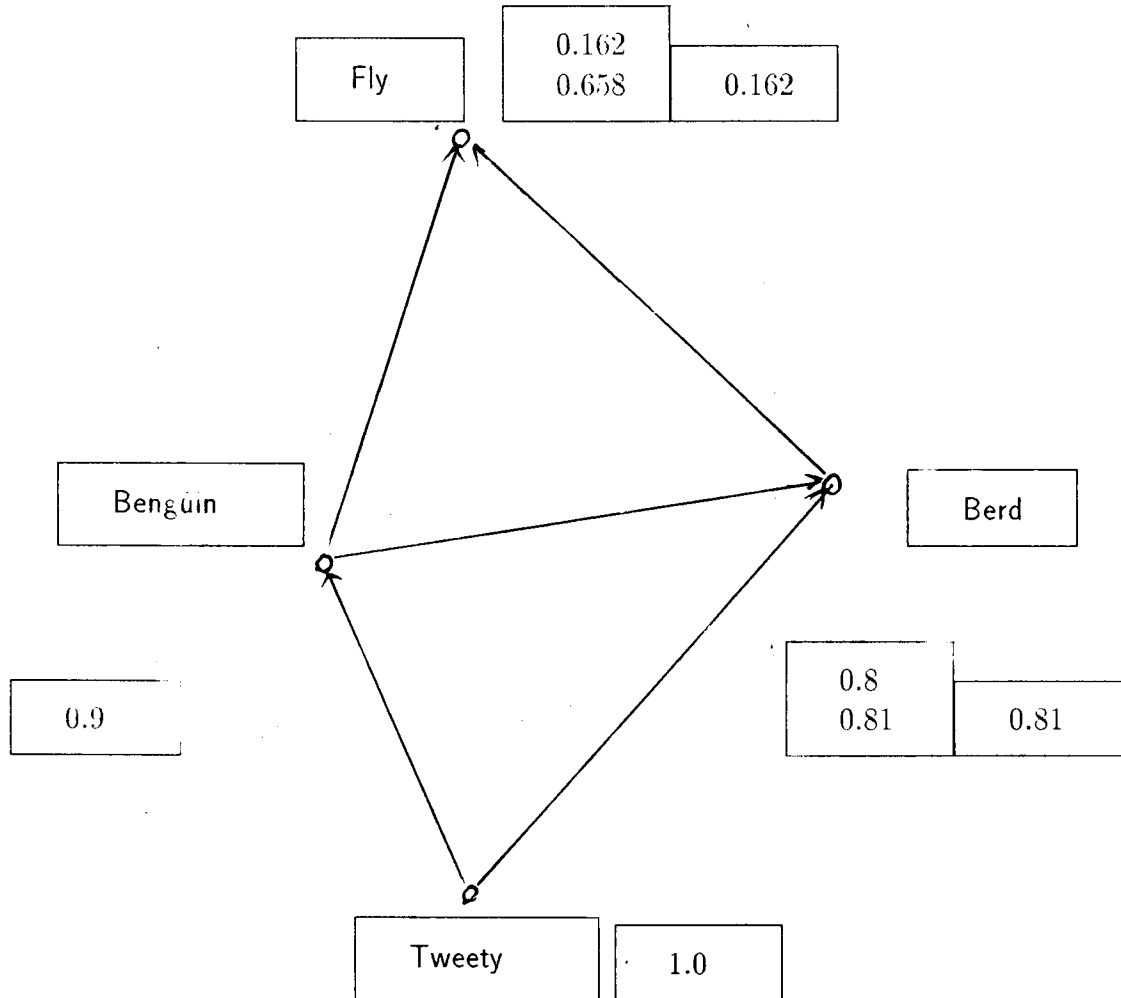
$$Weight(Bird, 2) = \max(0.9 - 0.1, 0.0) = 0.8$$

$$Ver(3) = \max(1.0 - 0.1, 0.0) = 0.9 \quad (21)$$

So that, after correction the weight of predicate *Bird* in rule (2) is lowed. The changed KB now is:

$$R = \{ Penguin^9 - Fly^2 (1); Bird^8 - Fly^9 (2); Penguin^1 - Bird^9 (3);$$

$$Tweety^9 \rightarrow Penguin^1 (5); Tweety^8 - Bird^1 (6) \quad (22)$$



Next inference is processed as follows:

$$\text{Rule (3) and Eq. (5) give } Pro(Bird) = 0.9 \times 0.9 = 0.81 \quad (23)$$

After applying Eq. (5) for rule (1), rule (2) we have:

$$\text{rule (1): } Pro(Fly) = 0.2 \times 0.9 \times 0.9 = 0.162 \quad (24)$$

$$\text{rule (2): } Pro(Fly) = 0.8 \times 0.9 \times 0.81 = 0.583 \quad (25)$$

Put values from (24), (25), (14) in Eq. (6a,6b):

$$PredConfirmation(0.162, 0.5, 0.2) = False \ \&$$



$$PredConfirmation (0.583, 0.5, 0.2) = True$$

$$||0.162 - 0.5| - |0.583 - 0.5|| = 0.255 > 0.2 \quad (26)$$

As

$$|0.162 - 0.5| > |0.583 - 0.5| \text{ so that } Pro(Fly) = 0.162 \quad (27)$$

Conceptual conflict did not occur.

#### 4. Summary and some suggested extensions of the method

The paper develops a mathematical formulation based on probabilistic reasoning and weights for eliminating conceptual conflicts among rules. It uses a combination of weights (to possibly represent deaf of importance of the predicate) and probabilities (to represent possibly occurrence of the predicate). Relying on weights and probabilities of predicates the probabilistic reasoning begets conceptual conflicts. Knowledge based refinement is considered as a process of discovering and eliminating the conceptual conflicts.

The mathematical formulation used in the work can be developed for machine learning, where differences among examples are considered as conceptual conflicts. The examples of learning are referred in some background knowledge base. After changing predicates' weights in rules of the knowledge base when there are many the same confirmations of some predicate or conceptual conflicts, the knowledge base will represent successfully all common and original characteristics of the examples.

This work presents early results on the subject of conceptual conflicts, and naturally, many problems remain to be solved. Here are some interesting topics for further research:

- In this method, the limits  $\lambda, \tau, \delta_\lambda, \delta_\tau, o(\delta_\lambda), \Delta_\tau$  used in algorithm 1 are represented as values. A desirable extension of the algorithm would describe them as predicates with their relative rules.
- The conceptual recognition is a partial problem of inference. The algorithm should express the activities of conceptual conflicts in inference.
- Giving examples of machine learning from a large number of training examples could lead to picture more clear the role of conceptual conflicts in the problem.

#### References

1. Fagin Ronald, Halpern J.Y., Vardi M.Y., *A model-theoretic analysis of knowledge*, Journal of Association for Computing Machinery, Vol. 38, No.2, April 1991, 82-128.

2. Ho Tu Bao, Dao Nam Anh, *Coupling of knowledge and data in TESOR system*, Informatics and Cybernetics, Vol.1, 1992, Hanoi, Vietnam, 1992, pp.21-31.
3. Lippmann R.P., *An Introduction to Computing with Neural Nets*, IEEE ASSP magazine, April 1987, pp.4-21.
4. Mitchell T. M., *Generalization as Search*, Artificial intelligence, March 1982, 18(2), 203-226.
5. Mitchell T.M., Keller R.M., Kedar-Cabekki S.T, *Explanation-based generalization: a unifying view*, Machine Learning, Kluwer academic publishers, Vol.1, 1986, 47-76.
6. Michalski R.S., Stepp R.E., *Learning from observation: conceptual clustering*, Machine Learning: An Artificial Intelligence Approach, 1983, 331-363.
7. Michalski R. S., *A theory and methodology of inductive learning*, Artificial Intelligence, Vol. 20, 1983, 111-161.
8. Judea Pearl, *Probabilistic reasoning in intelligent systems: Networks of Plausible Inference*, Morgan Kaufmann publishers, San Mateo, California, 1991, 552.
9. Roberts R. B., Goldstein I., *The FHT manual*, AI Memo No. 409, Massachusetts Institute of Technology, 1978.

## Abstract

*In the work knowledge based refinement in close relation with inference is considered. It proposes a new approach based on the notion of conflicts. A mathematical formulation for resolving conceptual conflicts among rules is developed. The conceptual conflicts are begot in probabilistic reasoning and then they are eliminated in machine learning. Knowledge based refinement is processed as learning from the conflicts' discovery.*

Institute of Informatics  
National Center for Natural Sciences and Technology of Vietnam  
Nghia Do, Tu Liem, Ha Noi, Viet Nam