J. Indian Inst. Sci., Nov.–Dec. 1996, 76, 739–750.

Cross-entropy measure for precedent case selection in case-based reasoning (CBR)

M. C. MATHUR
Electrical Engineering Department, Indian Institute of Technology, New Delhi 110 016.
email:mcmathur@ee.iitd.ernet.in

Received on January 31, 1995; Revised on August 8, 1995.

Abstract

In case-based reasoning the most important phase is related to the use of precedent cases. In some domains of inquiry such as the legal domain, the selection of a precedent case assumes critical importance. Researchers in case-based reasoning (CBR) appear to have not paid sufficient attention to the process of precedent case selection from among a large set, perhaps due to the fact that humans are poor at it as subjectiveness comes in. This paper proposes a selection criterion based on cross-entropy, a fundamental information theoretic measure and is almost directly applicable to the problem of case selection in CBR. Besides, it is a general measure applicable to a large class of domains.

Keywords: Cross-entropy, prior cases, CBR, AI.

1. Introduction

In artificial intelligence (AI), case-based reasoning (CBR) has attracted the attention of many researchers as it appears promising by being capable of exploiting the faculty of human being. This faculty helps humans in learning by doing things themselves and/or seeing things happening. For example, a pulley operator learns through examples demonstrated to him before he operates the pulley system.

The above approach leads to the basic idea of CBR in which case a reasoner solves a new problem by adapting solutions that were used to solve old problems. In this sense, a case-based reasoner tries to find those cases that solved problems similar to the one on hand, and adapts previous solutions to fit the current problem, taking into account any difference between the current and the previous situation[1].

However, a question arises as to what form should a case acquire so that it qualifies to be called a 'relevant' case. The answer to this point lies in finding 'relevant' cases involved in characterising the problem on hand by assigning appropriate features to it. The cases from the 'relevant' set may be selected which match best the input case. These may also be called precedent cases.

But the implementation of this aspect requires a classification or differentiation technique to classify various cases each of different characteristic feature. In this sense, this approach may be termed 'case-based reasoning' for classification problem-solving in which the fundamental problem-solving method revolves around the indexing and matching of past cases to a current problem.

Incidentally, this approach contrasts with the traditional expert system approach wherein past cases are compiled, usually into rules, by inductive or example-based generalisations from a set of one or more past cases, so that the problem revolves around reasoning with these rules. On the other hand, much of the power and intelligence of a case-based system resides in its indexing of past cases and in its creation of exemplar case.

To implement CBR, we need a strict discriminator which can distinguish cases based on minutest differences in information content. Our reckoning is that such a discriminator is possible by cross-entropy minimisation process[2]. The idea is that several cases with a variety of minor or major features show that 'appearances' of all these features put together are a random phenomenon whatever may be the situation. In that eventuality we take each case as consisting of any set of features. These features are assumed to have certain frequency of occurrence over a lar e period of time. Hence, each case may be represented as a probability assigned to each feature of the case. These values should be the averages of frequencies of occurrences of features in similar situations. For example, a dominant feature shows up quite often with relatively larger value of frequency, hence the higher average value of probability in comparison to that of other features. Anyone who observes this phenomenon of occurrence over a long period will be able to supply the value of probabilities of each feature of the cases. To our mind this collection or computation of average probabilities is quite possible in the legal domain. Perhaps in this very domain there is a strong need to select cases based on past experience about various features.

Once this characterisation of cases with probabilities is carried out, the Kullback–Liebler cross-entropy measure formulation[2] may be applied, which on minimisation gives rise to a probabilistic distance closeness among cases[3-6].

2. Case-based reasoning

Case-based reasoning for inferencing is obtained by relating the situation on hand to a precedent case which has been used previously to solve problems similar to the given problem. But in the method of case-based reasoning, the answer to the problem is approximate. Moreover, if the given problem is rather unusual in comparison to the previous one which has been solved, then difficulties arise in obtaining the desired solution. A case-based system is restricted to variations in known situations and delivers approximate, but fast results which are well ground in experience. In comparison to this, rule-based systems are flexible and are therefore capable of providing near-optimal solutions, but are slow and prone to errors as undesirable combinations of rules may get chained to produce incorrect answers. Case-based reasoning is a more suitable approach for domains which are realistically more complex. In complex situations, the rule-based method will be futile as it requires a large number of rules, many of them quite subtle unverifiable and the chain of reasoning becomes long and tedious[1]. Against this, in CBR one can always find a short relation or connection between input case and the retrieved solution[1].

There are two major advantages of CBR. First, the experience is more like a library of past experience than a set of rules; hence, cases better support knowledge transfer from domain expert to system, and therefore the justification of solution from system to domain experts. Secondly, many real-world domains are so complex that it is either impossible or impractical to specify all the rules involved; on the other hand, cases or solutions can always be given[1].

So to employ case-based reasoning we need relevant precedent cases which match the input case best. This search for cases has to be based on the features of the input case. It is important to note that the method of CBR is simple but critically hinged to the appropriate selection of precedent case(s). We consider it interesting to see how certain systems such as HYPO select or infer cases for CBR.

2.1. *CBR case studies*

2.1.1. *Legal domain, HYPO*

The logic of CBR is that once some experience has been obtained on any human activity, it is tackled easily when similar situation arises in future. The experience obtained is useful in dealing with very complex but soft situations even if it results in multiple answers. Instances of this kind abound in the legal domain. Considerable work has been done in this area by Ashley and Rissland who have developed a program called HYPO[7,8]. Literature reports many other programs such as JUDGE, CHEF, etc. The common characteristics, particularly the one about taking decisions, is based on past real-life experiences of similar kind. However, each program produced also has its specialty depending on the domain to which it belongs to. For example, HYPO does case-based legal reasoning in the area of patent law. Given a description of a case involving some claimed violation, HYPO uses its base of precedent cases to generate plausible arguments for prosecution or defence. Similarly, given a case description such as the release of trade secret to a competitor, and the goal of arguing for the defence, HYPO looks for those cases in memory most similar to the given case that were decided in favour of defence. HYPO then looks for ways in which to reduce the apparent differences, if any, between the given case and the retrieved successful case. For doing this, a large number of similar cases to the associated case on hand are required. A related issue is creating a library of similar cases.

2.1.2. *Design domain, CHEF*

From its title CHEF appears to be a program for preparing recipes. It generates new recipes by adapting old ones (Chinese soup recipes). This is essentially a design domain where old object has to be constructed to satisfy several goals simultaneously. This approach has several domains such as 'architecture', 'programming' and 'plan generation'.

CHEF has a library of about 20 recipes. The input to CHEF is a list of goals such as 'prepare hot stir fry dish' with chicken and noodle. The output is the desired recipe, which the user of CHEF evaluates. If the recipe is unsatisfactory, the user submits a failure report to CHEF which repairs it, and modifies its case library to avoid similar errors in future. CHEF learns from failures and its adaptation process is very complex.

We notice that CHEF has a set of recipes used as library or past cases. Similarly, the library or case base on failures is updated so that an acceptable recipe is prepared in later attempt. This example also highlights the need of case base to successfully obtain the desired results[7].

2.2. *Case comparisons*

The basic premise of CBR is related to contrasting cases among themselves. The key elements involve how prior cases are used for (i) credit assignment of factual features, (ii) justification, and (iii) arguments in domain that do not have strong causal theories. But compare and contrast is a difficult task as it requires specifying why they are the same or different. In other words, pointing crucial differences is an important component of carrying out explanations, arguing and planning. This also means that if we know why the cases are the same or what the crucial differences are, then it becomes an important component of explaining, arguing and planning. In fact, one cannot reason analogically without it only by focusing on important differences, as well as similarities, nor can one choose the best case, avoid the worst or extrapolate from unrelated cases. Despite the importance of this crucial intellectual skill, most expert systems do not represent cases or have the control structure to facilitate comparing cases. Research in CBR focuses on that deficit and how to correct it[8].

We notice from the above that there is a strong need for a technique which decides the closeness of precedent cases with respect to the case on hand. We will talk about a cross entropy-based method for comparing case features. Cross-entropy is a fundamental information theory concept which obtains, when used appropriately, probabilistic distances between cases using the factual features of the cases.

2.3. *Case dimension*

Case dimensions[8] represent valid relationships between various clusters of operative facts and valid conclusion they support or undermine. In case-based reasoning, generally a line of argument is pursued. Dimensions provide not only indices into lines of arguments of cases and their attendant analysis and argumentation but also a mechanism by which to judge the strength or weakness of a fact/situation with respect to the line of reasoning. Dimensions are also the basis for indexing prior cases for organising cases in case knowledge base.

The later step taken usually is carrying out indexing and relevant assessment of past cases by: (i) analysing how prior case can be reviewed from the point of view of the cfs (case fact situation), and (ii) determining which aspect of the prior cases apply, and how strongly, to the cfs.

This kind of analysis, when accomplished through dimensions, case analysis record and claim lattice mechanism, allows the program to promote some prior cases over others as precedents for interpreting and arguing the cfs. The program of Ashley and Risslands[10] compares and contrasts the cfs and prior cases at the level of facts, justification and arguments to come up with the best case pros and cons decision and to pose instructive hypothetical variants of the cfs[8]. This step is possible only when close precedent has been selected.

3. Cross-entropy measure

In Section 2.2, there is a comment on compare and contrast among prior cases in the context of case selection. It highlights the fact that despite the importance of the crucial intellectual skill which humans possess, most experts do not represent cases objectively or have no control structure to facilitate comparing and contrasting cases. Therefore, all such attempts on the part of human beings will always be constrained by the inherent weakness of being imperfect or less objective. The comment further suggests to remove the difficulty regarding selecting cases objectively. But it is difficult to say whether this would be possible in the near future using purely symbolic approach. Therefore, we should make efforts to develop or use methods or techniques already existing to help evaluate prior cases objectively. Out of the methods available, we reckon the information-theoretic cross-entropy formulation would be highly suitable for prior case selection.

Cross-entropy formulation is a strong theory based on highly incisive discriminator so that minutest differences in information or in description of cases would be noticed while classifying them. The classification result is obtained in probabilistic distance between two probability distributions representing two separate cases. We will use this method in the next section.

In the above context it becomes essential or mandatory to select a case which should be as close to the input case as feasible technically based on the factual features of the case. The argument is that if the selected case is the closest technically, further analysis would bring in supportive components more strongly in terms of additional strength in each component such as case dimension, claimed lattice, case analysis, etc., which should help generate strong arguments in domains such as legal practice and design.

Our belief is that appropriate selection of prior cases, as per the above suggestions, helps the method gain strength. Also, it avoids analysis of each component, whether prospective or not, for the case-based reasoning. As an extension of that argument, if each case is to be analysed thoroughly before selection, it results in loss of time and money, sometimes without any success and consequently weakens the process of selection. So, based on the argument that closer the precedent case is to the input case, stronger the argument will be leads to the cardinal principle "analyse only those prior cases which have been provided closest based on a criterion".

3.1. *Cross-entropy as a measure*

We have so far discussed the importance of closely similar cases in CBR. We now consider an information-theoretic cross-entropy measure for selecting the precedent cases from among a general, large set. Such a measure will be of immense help to the case-selection process in choosing prior case objectively.

Kulback–Leibler minimum cross-entropy principle (MinXEnt) is an entropy-optimisation principle. It is also known as the minimum discrimination information principle. It emphasises the minimum cross-entropy of a probability distribution P from another probability distribution Q. This cross-entropy between the two distributions is also called the probabilitic distance between P and Q, or cross-entropy of P with respect to Q.

The Kulback–Leibler entropy between two probability distributions, p and q, is defined as follows:

$$(P{:}Q) = \Sigma\, p_i \log (p_i/q_i)$$

where $P = (p_1, p_2, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_n)$.

We assume that whenever $q_i = 0$, the corresponding p_i is also zero. Beside, there are several properties which one can use when dealing with two probability distributions. Since we are concerned with prior case selection objectively, we adapt the above formulation of cross-entropy to help evaluate precedent cases for selection to be used in CBR.

We have mentioned in the introduction that appearance of certain features of the prior cases is a random phenomenon. We consider each case consisting of certain features. On this basis, we see that each case is a probability distribution with usual probabilities assigned to each feature.

Example: Let us consider how probability distributions are compared using cross-entropy formulation.

We assume that we are given *a priori* probability distribution $Q(0.05, 0.10, 0.15, 0.20, 0.22, 0.28)$. We also assume that the mean of second distribution P is 4.5. Then we are required to find the minimum cross-entropy distribution P using the information given above. It is also desirable to determine the entropy of P.

Solution: As per Kullock formulation, we minimise

$$\Sigma\, p_i \log \frac{p_i}{q_i} \tag{1}$$

$$\text{subject to } \Sigma_i = 1 \tag{2}$$

$$\text{and } \Sigma_i p_i = 4.5. \tag{3}$$

Equation (2) is a natural constraint, whereas (3) pertains to given information regarding probability distribution P. Here the probability distribution P is regarded as a dice of six faces with values 1, 2, 3, 4, 5, 6.

From the above, we may write a lagrangian as follows:

$M = \sum_{i=1}^{6} p_i \log (p_i/q_i) + \lambda_0(1-\sum_{i=1}^{6}) + \lambda_1(4.5-\Sigma_i p_i)$ or,

$M = p_1 \log (p_1/q_1) + p_2 \log (p_2/q_2) + p_3 \log (p_3/q_3) + p_4 \log (p_4/q_4) + ..+$

$p_6 \log (p_6/q_6) + \lambda_1 (4.5-p_1-2 + p_2-3 + p_3-4 + p_4-5 + p_5-6 + p_6)$.

Then $M/dp_i = 0$ given

$p_i = q_i\, a\, b^i$

where $\Sigma_{i=1} q_i a\, b^i = 1$

$\Sigma_{i=1} ab^i = 4, 5$.

Substituting for $q_1, q_2, ..., q_6$ and solving for a,b we get distribution

$$P = \{0.03, 0.078, 0.131, 0.192, 0.234, 0.330\}.$$

We see that the entropy $H(p) = 1.605$. In this case, we have found a probability distribution P corresponding to a six-face dice which is closest to a given distribution Q.

It is quite possible that we are not given Q. In that case we use uniform distribution U instead of Q, and we are required to proceed as follows:

minimise $\sum_{i=1}^{6} p_i \log (p_i/1/16)$ or maximise $- \Sigma\, p_i \log p_i$

subject to the same constraint as earlier. When we carry out this optimisation, we get

$$\sum_{i=1}^{6} c\, d^i = 1,\ \sum_{i=1}^{6} i\, c\, d^i = 4.5$$

and the distribution

$$P' = \{0.0543, 0.0788, 0.1142, 0.1654, 0.2378, 0.3475\}.$$

We see that the entropy of P', $H(p') = 1.613$ is greater than $H(p)$. This is because out of all distributions with mean 4.5, P' has the maximum entropy while P is the only other distribution with this mean, and as such the entropy of P has to be less than or equal to the entropy of P'.

These examples indicate the way cross-entropy formulation may be utilised to determine the relative standing of distribution with respect to others. The uniform distribution can also be taken as reference for comparison among distributions.

3.2. *Precedent case selection*

(a) The main difference between other AI systems built to solve problems and CBR is that CBR uses a large case library as against the set of first principles[11]. Therefore, the strength of CBR lies in the case library. To achieve success, CBR organises cases in memory and a rich indexing mechanism is used to use it effectively. Therefore, whenever a problem is presented, one is reminded of relevant past cases/experiences but is not encumbered by a lot of unwanted memory. This aspect of indexing emphasises indexing past cases by their factual features. However. we need to distinguish important indices from unimportant ones to avoid situations where everything seems related to every other thing, making it difficult to focus on relevant memories.

There are some features that are important only in a certain context, and others vary from domain to domain. Therefore, a general CBR system must be capable of selecting proper sense of indices from experience. Though this aspect is not the subject matter of this paper, our method of case selection can take care of this problem by assigning different probabilities to the features that can reflect on their importance or occurrence in changed situations. A case is usually stored as a monolithic structure, although in some variations cases can be stored in piecemeal. This will mean that in one situation all the features are the basis of case selection with some assigned probabilities, while in the other the case may be split in correspondence with a new input case feature. This will also entail change in case size feature, assuming that relevant statistical data is available on the new features.

(b) The basic assumption of the following development is that each case be considered to be a probability distribution with each feature of the case having some probability of occurrence. This allows the case representation of the form :

$$\text{case 1: } (x_1, x_2, \ldots, x_n)$$

where $x_1, x_2, \ldots, x_n$ are the probabilities associated, respectively, with feature 1, 2,..., n. The basis for this association is that occurrence of a particular feature is a random event.

It is important to realise that all the probabilities considered above are, in fact, proportions of overall features of the case. In other words, the features possessed by a case are finite in number so that their sum adds up to a fixed entity.

This implies that $p_i = x_i/(x_1 + x_2 +, \ldots, x_n)$, $i = 1, 2, \ldots, n$ are non-negative proportions whose sum is unity. In the literature Kullback–Leiber entropy relating to distributions P and Q has been defined as follows

$$D(P{:}Q) = > p_i \log p_i/q_i,\ i = 1, 2\ .\ .\ .n$$

where $P = (p_1, p_2, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_n)$. We assume that wherever $q_i = 0$, the corresponding p_i is also zero.

One basic property of the cross-entropy measure is that $D(P{:}Q) >= 0$, which vanishes when $P = Q$. We can deduce from this property that the minimum value of $D(P{:}Q)$ is zero. We can show that $D(P{:}Q)$ is not a symmetric measure, *i.e.*, $D(P{:}Q) =/= D(Q{:}P)$. But we can create a symmetric measure, that is,

$$J(P{:}Q) = D(P{:}Q) + D(Q{:}P) = \textstyle\sum_{i=1}^{n} p_i \log (p_i/q_i) + \sum_{i=1}^{n} q_i \log (q_i/p_i)$$

is symmetric, since $J(P{:}Q) = J(Q{:}P)$. We call $J(P{:}Q)$ is a measure of symmetric cross-entropy, or symmetric divergence. It can be viewed as a special case of cross-entropy or the discrimination, a measure which defines the information-theoretic similarity between two probabilitic distributions. In this sense, it is a well-defined measure of dissimilarity between *a priori* and posterior beliefs about a case. We, therefore, can call $J(P{:}Q)$ j measure. It also implies that j measure is a special case of cross-entropy and helps in expressing our confidence in using it as a measure for discriminating one case *vis-a-vis* another.

3.3. *An example from legal domain*

We now consider an example to show how the process of case comparison is carried out using cross-entropy as a measure for selection of cases.

In literature on CBR, several case studies have been mentioned. Among these the one which generated great interest in using AI technique in the legal domain is the work of Ashley and Rissland[7,8], which is called HYPO. It works in the area of trade secret law which lays down provisions for restricting a company from accessing trade secret of another company.

In CBR, dimensions or features are important factors based on factual information. These factors usually influence the outcome of the case. Features provide a framework for describing, comparing and generating arguments about a case. Using these features, one can assess the strength and weakness of one's own case and that of the adversary. This further requires identifying the applicable features as per the claim sought. HYPO has 13 dimensions such as competitive advantage—gained, disclosed—secrets brought, tools.

A legal dispute contains facts which make some dimensions applicable and others inapplicable. Some dimensions favour the plaintiff and others the defendant. Hence, there may be conflicting situations that are generally not resolved, and are used in argument by respective sides. In this sense, a precedent case represents the decision factors that resolved the case. This becomes possible because the precedent case carries the past judgement in a similar situation.

The use of precedent cases for resolving competing factors is essential because legal domain does not provide any quantitative formulation or authoritative weights. On the other hand, attorneys may agree that a particular factor is more important than the other in the given situation. But they also accept that the same may not be true in another fact situation. Thus, it is not possible to assign weights to the factors. The issues are resolved by citing a similar precedent case that assists the factors in the present case so that it can be resolved the way the precedent has been resolved. Thus the importance of similar precedent cases is highlighted once again.

At the optimal level the problem of the precedent case selection is stated as follows:

Given *a priori* input case, represented as probability distribution Q, choose that precedent case, represented as another probability distribution P, which satisfies the given constraints, and is also closest to Q such that cross-entropy is minimised.

We have seen earlier in Section 3.1 how the distributions can be compared. In particular, the comparison with the uniform or normal distributions appear to be very useful to determine the probabilistic distance between the two distributions.

In every legal case there are at least two parties involved—the plaintiff and the defendant. The complaint lodged by the plaintiff in the form of a written legal document is used to identify the features or dimensions of *a priori* case. Let us assume that the plaintiff is the owner of a company dealing with manufacturing/trading a product. One day an employee switches to another disclosing trade secrets, knowledge of the product and tools. As a result, the plaintiff's business is adversely affected. The plaintiff files a suit, resulting in legal case/dispute.

This situation has been analysed by Ashley and Rissland who identified 13 factual features pertaining to the case. Some features favour the plaintiff and others the defendant.

The plaintiff argues that the product trade was established by him through research and development. The defendant says that the man who switched over was not adequately compensated for his contribution and for other issues involved. From the CBR point of view this appears to be a fairly general case/problem. Combining features favouring plaintiff will help to create *a priori* case. Similarly the defendant's case may be created.

Based on common features we must select precedent cases which should be closest to *a priori* case. To select cases we must use the following method:

(i) Let five features be arranged as per their significance.

(ii) Features are assigned appropriate probabilities, giving rise to probability distribution: Case 1: (0.4, 0.25, 0.2, 0.10, 0.05).

(iii) Compare available precedent cases with *a priori* case. Each case is represented as a probability distribution. The precedent cases we have are represented as follows: Case 2:(0.43, 0.31, 0.16, 0.03, 0.07); Case 3:(0.23, 0.012, 0.03, 0.50, 0.12).

In order to compare Case 2 with Case 1, we adopt an approach in which both the cases are compared with uniform distribution. We also compare *a priori* case with uniform distribution, so that we get information about all the three with uniform distribution as the reference.

(iv) Let us minimise

$\Sigma\, p_i \log (p_i/q_i)$ $i = 1, 2, 3, 4, 5.$

subject to the conditions $\Sigma p = 1$,

$\Sigma p_i c_i = c$.

We get

$p_i = q_i \exp (\lambda_1, \lambda_2, \ldots, \lambda_{mc})$.

This means that each $p_i > 0$. We notice the exponential functional form of p_i. It appears to be q_i times the MaxEnt Shannon entropy maximisation value of the very same constraints. The MaxEnt value is given by

$T_i = K\, T \exp(-\beta c\, c_i)$.

Substituting appropriately, we get

$p_i = q_i\, K\, T \exp(-\beta\, c_i)$

where K is internal constant.

$K = 1/(\exp(-\beta c_1) + \exp(-\beta c_2) + \ldots\exp(-\beta c_5))$

and T = the sum of probabilities = 1.

Hence, finally we get

$p_i = q_i K \exp(-\beta c_i)$, where β is a constant which; we assume $\beta = 0.01$.

Using this formula and the known values of q_i we get the following three distributions closest to the uniform distributions corresponding to cases 1, 2, 3.

Case 4: (0.392, 0.245, 0.116, 0.196, 0.098, 0.049)
Case 5: (0.421, 0.304, 0.157, 0.029, 0.067)
Case 6: (0.156, 0.034, 0.05, 0.55, 0.336)

From the above, we can determine the value of entropy of all the distributions by using the Shannon's formula $\Sigma\, p_i \log pi$, where $i = 1, 2, \ldots 5$. Carrying out this computation shows that the entropy of the three distributions is (4):1.25, (5):1.68, (6):1.13. This shows that *a priori* distribution is closer to the uniform distribution, but the first precedent case is closest to the uniform distribution. The second precedent case is the farthest from the uniform distribution. So, if we have to choose one out of the two cases we must choose the distribution which is close to uniform distribution, that is the one with entropy = 1.68. This way the precedent cases may be classified so as to create a set of cases which may be arranged one after the other according to the probablistic distance between them.

Further, using certain properties of principle(MinxEnt), it is possible to combine the case in parts. This possibility may be useful in dealing with some special situations.

4. Conclusions

We have emphasised the importance of case-based selection aspect of CBR. Perhaps the lack of suitable methods for case selection is the reason that motivated us to pursue this work. Our present knowledge of minimum cross-entropy principle(MinxEnt) has helped us in conceptualising its direct use in the process of precedent case selection. Basic premise of MinxEnt is based on the concept of probabilistic distance between probability distributions. This leads to the assumption that legal precedent cases or for that matter any system of prior cases can be represented as probability distributions. Then the formulation of cross-entropy, an optimisation process provides information pertaining to probabilistic distance among distributions. Cross-entropy concept is a fundamental information-theoretic concept and is highly sensitive to variations in information content of any distribution. This measure ensures that any minute change in case features and probability distribution will be taken care of by the formulation and is reflected in the characteristic representations of cases and conclusions drawn based on them.

Further efforts are required to make the cross-entropy-based technique of case selection widely acceptable as it is objective in nature and is quite general. It is also desirable to provide justification as to why and how the precedent cases be represented as probability distributions. If such efforts succeed it will bring in objectivity in approaching precedent cases.

References

1. Riesbeck, C. K. and Schank, R. C. — *Case-based reasoning: an overview.* In *Inside case-based reasoning*, 1989, Earlbaum Associates.
2. Kapur, J. N. and Kesavan, H. K. — *Entropy optimization principles with applications*, 1992, Academic Press.
3. Ashley, D. K. — *Modeling legal argument: Reasoning with cases and hypothetical*, 1990, MIT Press.
4. Ashley, D. K. — Arguing by analogy in law: Case-based model. In *Analogical reasoning—Perspectives of artificial intelligence, cognitive science and philosophy* (Hemand, D.H., ed.), 1988, Kluwer.
5. Kolodner, J. L. — Expanding problem solving capability through case based inference, *Proc. 4th Annual Int. Conf.* 1987, Morgan Kaufman.
6. Carbonell, J. G. — Derivational analogy, a theory of reconstructive problem solving and expertise acquisition. In *Machine learning, artificial intelligence approach*, Morgan Kaufman.
7. Rissland, E. and Ashley, K. D. — Hypothetical as heuristic device, *Proc. AAAI*-86, pp. 289–297, Morgan Kaufman.
8. Ashley, K. D. — Distinguishing—a reasoner's wedge, *Proc. 9th Conf. on Cognitive Science Soc.*, pp. 737–747, 1987, Erlbaum Associates.
9. Hammond, K. J. — Case-based planning: viewing planning as a memory task. In *Perspectives in AI*, 1989, Academic Press.
10. Ashley, K. D. and Rissland, E. — Compare and contrast, a test expertise, *Proc. AAAI*-87, Morgan Kaufman.
11. Rich, E. and Knight, K. — *Artificial intelligence*, 1991, Tata McGraw-Hill.