人 工 知 能 71－5
（1990. 7. 13）

Analogy by Simulation - a Weak Justification Method

(PRELIMINARY REPORT)

Jun ARIMA

ICOT Research Center
Mita Kokusai Bldg. 21F 4-28 Mita 1-chome, Minato-ku, Tokyo, 108, Japan
Phone: +81 3 456 2514, C.Mail Address: arima%icot.jp@relay.cs.net

Abstract: This paper is a preliminary report on a novel method for analogical reasoning. Davies points out a problem, called *the justification problem*, claiming that we should find a criterion which justifies the conclusion obtained by analogy. This paper takes a certain type of analogy, and discusses an answer of the problem. The central idea of this method is inspired by *simulation*, that is, the unknown system is simulated by a well-known system and phenomena which would occur in the unknown system are predicted by projecting phenomena which occur in the well-known system. A projected conclusion is *indirectly justified* in the sence that the conclusion is justified in another similar system.

シミュレーションによる類推: 一つの弱正当化法

(PRELIMINARY REPORT)

有馬 淳 (I C O T)

概要: 本稿では論理に基づいた類推に関する研究の途中報告を行なう. T.R. Davies は類推に関して正当化問題 (justification problem) と呼ばれる問題を提起し，類推によって得られる結論が十分に正当化されるべき基準を明らかにすべきであると指摘した. 本稿はある種の類推を取り上げ，正当化問題を議論する. 新しく提案する手法はシミュレーションに発想を得ており，未知のシステムを類似の既知のシステムでシミュレートすることにより，未知のシステムに生ずる現象を予報するというものである. 既知システムより転写される結論は，既知システムでは正当化されるという意味で，弱く正当化されており，この手法は非演えき的な類推を表している.

1 Introduction

When we explain a process of reasoning by analogy, we may say that "An object T is similar to another object S in that T shares a property P with S. S satisfies another property Q. Therefore, T satisfies Q, too", or it may be expressed more formally by a schema "If $P(T) \wedge P(S) \wedge Q(S)$ holds, then $Q(T)$ holds". Here, T will be called the *target*, S be the *source*, P be the *similarity* or *shared property* between T and S, and Q be the *projected property.*

Nevertheless, the above description of the process of analogy is insufficient. Researchers studying analogy have come to recognize the neccesity to reveal some implicit knowledge which influences the process but does not appear in the above schema. T.R.Davies *et al.* [6] gives an intuitive example which shows the existence of such implicit knowledge: Bob's car and Sue's car are both the same type, but we could not infer that Bob's car is painted red just because Sue's car is painted red. From the fact that Sue's car is worth about $3500, however, we may infer that Bob's car is worth about $3500. It clearly suggests that the plausibility of the conclusion depends on some implicit knowledge that is not provided in the premiss and that is on the relation between the similarity and the projected property. To reveal such implicit knowledge which justifies some analogical inference is very important, because it prevents an syntactical application of an analogical schema from yielding useless conclusions.

This paper takes a certain type of analogical inference and proposes a novel method based on a logical criterion which weakly justifies the conclusion obtaibed by the analogy. The central idea of this method is inspired by *simulation*, that is, the unknown system is simulated by a well-known system and phenomena which would occur in the unknown system are predicted by projecting phenomena which occur in the well-known system. A projected conclusion is *indirectly justified* in the sence that the conclusion is justified in another similar system.

2 What is projected and how?

Justification of projecting a certain property has recently aroused the interest of reserchers studying analogy [5, 6]. According to their approach, once we give the implicit knowledge for justification, analogy is collapsed just into deduction. This will be against our intuition ("analogy" is not deductive!). Here, we seek a weaker criterion for the justification which leaves analogy non-deductive.

2.1 The importance of precondition of causality in analogy

Importance of causality in analogy has been emphasized repeatedly [1, 2], for instance, Winston proposed a theory for analogy, where the causal structure of the source situation is assumed to map onto the target situation. When we consider causality more precicely, we may have a different standpoint, that is, any causalities govern every situations, however, whether the causalities may work actually depend on the hold of their precondition which are necessary when causalities influence some situations. Such view of causality has been taken in recent studies on reasoning about action [3, 4]. If we see analogy from this standpoint, we will notice that it is the precondition of causality rather than the causality itself that is mapped by an analogical process. One of the difficulties of clarifying such analogy is that the precondition is often implicit, that is, it may not appear explicitely in the discription of the similality.

2.2 Extracting implicit precondition from simulation

When we observe a phenomenon, we often reason why the phenomenon occurs by making use of some rules about the domain, causalities and, sometimes, knowledge about another well-known domain. For instance, when we need to infer something about another person, by putting ourselves in a certain situation which he occupies, we sometimes find a explanation of how and why he is what he is, and can conjecture unknown properties which he would satisfy, his present state, character, movements of his mind, the purpose of his action and his next action.

This type of inference is very common in human reasoning, sympathy will belong to this type, experiments by simulation may be considered as this case in the technological

field and the inference has been seen in many papers on causal reasoning in the cognitive science fields [8, 9].

Such inference can be considered as a certain type of analogy, where the source is just each of ourselves, the target is he, the similarity is the fact that we have the possibility to cause a same phenomenon (otherwise, we could not explain why he do so), and projected properties are, essentially, preconditions which are needed in the explanation of the phenomenon, for instance, a precondition that he has the same type of mind as we have and that it governs all the process in the conjecture. What we conjecture above is deduced from the projected preconditions and known facts. Thus, we will easily extract possibly related implicit preconditions on the observed phenomenon from simulation in the well-known domain.

We describe our assumption more clearly: *the preconditions of the causality which actually works in one situation are generally projected into another similar situation in which the causality may actually work.* According to our idea, two situations are *similar* when a certain common causality seems to work actually in both situations. The shared property and the projected property are *relevant* because both properties follow from a common causality and some specified preconditions of it. If a specified precondition of the causal relation is satisfied in one situation, the precondition is assumed to hold in another similar situation. Therefore, "the causes in similar situations generally lead to similar effects [1]".

3 A Simulation Method

Using the following example, this section introduces a method for analogy by simulation.

Example

Hector feels pain when he is injured or burnt. Also, Brutus feels pain when he is injured. This may be represented as follows:

$$Has(Hector, Injury) \supset Feels(Hector, Pain)$$

$$\wedge Has(Hector, Burn) \supset Feels(Hector, Pain)$$

$$\wedge Has(Brutus, Injury) \supset Feels(Brutus, Pain)$$

In this case, we would conclude that Brutus feels pain when he is burnt by analogy, however, we would not conclude that Brutus is powerful even if Hector is poweful.

In this method, it is assumed that domain knowledge and knowledge about the source are given. Let knowledge be a set of first order sentences. It is devided into two subsets, a set D of sentences free from a paticular object S and a set $F(S)$ sentences in which S occurs.

In this paper, an ***explanation*** means a minimal deduction path from a certain premise to a paticular conclusion, we say a certain knowledge *directly relates* to an explanation if the knowledge occurs (or, is used) in the explanation, and we say α ***explains*** β (written by "$\alpha \vdash_{exp} \beta$") if there exists an explanation from α to β and if all of the premise directly relate to the explanation, (that is, if we remove a sentence from the premise α, we can not find another deduction path in making use of the remainder premise).

The follwing is a detail of this method. It can be devided into four steps.

(1) Understanding Step:

From the domain knowledge $d \subseteq D$ and the source knowledge $f_1(S) \subseteq F(S)$, a explanation how the source satisfies the shared property is made $(f_1(S), d \vdash_{exp} P(S))$ (that is, $f_1(S)$ and d are minimal sets used in the explanation). $f_1(S)$ will be called the *implicit premise* w.r.t. P, and d be the *implicit causal knowledge.*

$$d: \quad \forall x, i.(Animal(x) \wedge Has(x, i) \wedge Destructive(x, i) \supset Feel(x, Pain))$$
$$f_1(S): \quad Animal(Hector) \wedge Destructive(Hector, Injury)$$
$$P(S): \quad Has(Hector, Injury) \supset Feels(Hector, Pain)$$

(2) Mapping Step:

The implicit premise $f_1(S)$ used in the above explanation is mapped into the target (that is, it is assumed that $f_1(T)$ holds) if it does not cause inconsistency with knowledge about the target. That is, if it mapped, the target can be explained to satisfy the shared property by the same way in Understanding Step $(f_1(T), d \vdash_{exp} P(T))$. If this mapping is proceeded successfully, both domains are considered to be similar w.r.t. a ceratin causal knowledge d as the causality may actually work in the both.

$f_1(T)$: $Animal(Brutus) \wedge Destructive(Brutus, Injury)$

(3) Justifying Relevance Step:

From the implicit causal knowledge d and the source knowledge $f_1(S)$, and additionaly $f_2(S) \subseteq F(S)$, this step tries to *infer* (not necessarily to *explain*) how the source satisfies the projected property Q $(f_1(S), f_2(S), d \vdash Q(S))$. If possible, Q will be called *relevant* to P w.r.t. d. Also, $f_2(S)$ will be called the implicit premise w.r.t. Q.

$f_2(S)$: $Destructive(Hector, Burnt)$
$Q(S)$: $Has(Hector, Burnt) \supset Feels(Hector, Pain)$

(4) Projecting Step:

The implicit premise $f_1(S)$ and $f_2(S)$ are mapped into the target if it does not cause inconsistency with the target knowledge. If it mapped, the target can be explained to satisfy the projected property by the same way in Justifying Relevance Step $(f_1(T), f_2(T), d \vdash Q(T))$, that is, it is conjectured that the target may have a property Q.

$f_2(T)$: $Destructive(Brutus, Burnt)$
$Q(T)$: $Has(Brutus, Burnt) \supset Feels(Brutus, Pain)$

Here, note that other possible projected properties are

$Animal(Brutus)$, $Destructive(Brutus, Injury)$, $Destructive(Brutus, Burnt)$, ...

however, a property which does not relate the explanation in the source domain, like $Powerful(Hector)$ or $TwoEyes(Hector)$, is profibited from being projected, even assuming that $Powerful(Hector) \in F(Brutus)$ or $\forall x.(Animal(x) \supset TwoEyes(x)) \in D$.

4 Conclusion and Remarks

This paper proposes a novel logical method for analogical reasoning.

This method also gives an answer to *the non-dedundancy problem* pointed out by Davies *et al.* [6], the source instance should provide new information about the conclusion. If a

conclusion obtained by this method is not deductive (for instance, $Has(Brutus, Burnt) \supset Feels(Brutus, Pain)$, $Animal(Brutus)$, $Destructive(Brutus, Injury)$, ... in the above example), it is a projected property shared by the source ($Hector$), which is represented in the source knowledge ($F(S)$) or obtained from the domain knowledge additionaly to the source knowledge. That is, the source information is actually used.

This method is general in that it is a logical approach independent from any paticular system. In fact, it seems not to cause inconsistency to studies which have been reported so far, but to make their conclusions more selected. However, this method will not yield a certain type of analogy like the example of cars reported in the introduction, which is called *functional analogy* [7].

Acknowledgements

I am grateful to Natsuki Oka for his useful comments.

References

[1] Winston,P.H.: Leaning Principles from Precedents and exercises, *Artifficial Intelligence*, Vol. 19, No. 3 (1982).

[2] Gentner,D.: Structure-mapping: Theoretical Framework for Analogy, *Cognitive Science*, Vol. 7, No. 2 (1983) 155-170.

[3] Lifschitz,V.: Formal Theories of Action, in *Proc. of The Frame Problem Workshop 87* (1987) 35-57.

[4] Georgeff,M.P.: Many Agents Are Better than One, in *Proc. of The Frame Problem Workshop 87* (1987) 59-75.

[5] Keder-Cabelli,S.: Purpose-directed analogy, in *the 7th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates (1985) 150-159.

[6] Davies,T. & Russel,S.J.: A logical approach to reasoning by analogy, in *IJCAI-87* (1987) 264-270.

[7] Collins,A., Warnock,E.H.,Aiello,N., & Miller,M.L.: Reasoning from incomplete knowledge, In D.G.Bobrow and A.Collins (Eds.), *Representation and understanding: studies in cognitive science*, New York: Academic Press (1975).

[8] Kanouse,D.E.: Language, labeling, and attribution, In E.E. Jones, D.E.Kanouse, H.H.Kelley, R.E.Nisbett, S.Valins & B.Weiner (Eds.), *Attribution: Perceiving the causes of behavior*, General Learning Press (1972) 121-135.

[9] Nisbett,R.E. & Ross, L.: *Human inference: Strategies and shortcomings of social judgement*, Prentice-Hall (1980).