TR-591

Analogy by Simulation-
a Weak Justificaiton Method
(Preliminary Report)

by
J. Arima

September, 1990

ICOT
Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03)3456-3191 ~ 5
Telex ICOT J32964

Institute for New Generation Computer Technology

Analogy by Simulation
- a Weak Justification Method

(PRELIMINARY REPORT)

Jun ARIMA

ICOT Research Center
21F, Mita Kokusai Bldg., 1-4-28 Mita, 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 to 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 *illustratively justified* in the sense that the conclusion is justified in another existing similar system.

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* or the *base*, P be the *similarity* or the *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 necessity of revealing some implicit knowledge which influences the process but does not appear in the above schema. T.R.Davies *et al.* [2] give intuitive examples which show the existence of such implicit knowledge. Here, we take another example and review the problem.

Example[1]: Brutus feels pain when he is cut or burnt. Also, Tacitus feels pain when he is cut. Therefore, when Tacitus is burnt, he would...

[1]This example originates from a famous science-fiction story[7].

Both Brutus and Tacitus feel pain when they are cut (a shared property), but we could not infer that Tacitus is strong just because Brutus is strong (as a projected property). However, we may infer that Tacitus feels pain when he is burnt just because Brutus feels pain when he is burnt (as another projected property). The point is the fact that we prefer the former reason than the latter though neither has any difference in applying the above schema. It clearly suggests that the plausibility of the conclusion depends on some implicit condition that is not provided in the premise of the schema and that relates the similarity and the projected property. To reveal such an implicit condition which justifies some analogical inference is very important, because it prevents an unrestricted superficial application of an analogical schema from yielding useless conclusions.

Many works on analogy have defined similarity independently of what property is projected, or have assumed similarity to be given without clarifying the relation between the similarity and the projected property [13, 3, 5, 6]. For instance, Winston's program works based on a similarity measure which depends on counting equivalent corresponding attributes in a frame, which means that the similarity is decided a priori independently of the projected property. We should consider similarity in the context of the projected property when the projected property is given. There is no doubt that what is similarity depends on what property is projected. The work proposed by T.D.Davies *et al* [2] is also done from such a motive. Their solution does not completely satisfy us, however, in that, according to their approach, once we give the implicit knowledge for justification, the analogy collapses 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 Extracting Implicit Knowledge by Simulation

Consider the traditional philosophical problem of other minds. While I know that I have a mind because of my experiences, I do not have any experience of your experiences. Then, how do I guess that you also have a mind? Apparently, though the only evidence we have on which to base our inference is the external observation of others, we seem to be able to reach the proper conclusion. Let us look at this in a more common situation and see why and how we reach such a plausible conclusion. When we need to infer something about someone, by imaging ourselves in their shoes and by simulating him, we sometimes find a explanation of how and why they are what they are, and can conjecture unknown properties which they would satisfy, their present state, character, movements of mind, the purpose of their past and future actions. A feature of such reasoning may be divided into the following four steps: to think about an observation of an unknown domain as if it were a case of a well-known domain, to extract implicit properties from the well-known domain which are necessary to explain and understand the observation, to map the extracted knowledge into the unknown domain, and then to deduce various plausible conclusions about the unknown domain from the mapped knowlege. Such reasoning can be considered as a certain type of analogy, where the

base case is ourselves, the target is the other person, the similarity is the fact that we can cause the same phenomenon, and the projected properties are, essentially, implicit facts about the base case which are needed as some premises in the explanation of the phenomenon, for instance, the fact that we felt sad in such a case. What we conjecture above is deduced from the projected implicit facts and known facts. Thus, we will easily extract implicit facts possibly relating to the observed phenomenon from simulation in the well-known domain and we can get more conclusions from projected implicit facts than from explicit facts.

This type of inference is very common in human reasoning; we often conjecture in just that way when we try to understand others. Sympathy belongs to this type, experiments by simulation may be considered as this case in the technological field, and such inference has been seen in many papers on causal reasoning in the cognitive science fields [8, 12]. Also, this type of inference might be useful in a more sophisticated treatment of causality in analogy.

The importance of causality in analogy has been emphasized repeatedly [13, 3]. For instance, Winston proposed a theory for analogy, in which the causal structure of the base situation is assumed to map onto the target situation. When we consider causality more precisely, we may have a different standpoint: any causality may affect any situation. However, whether the causalities will work actually depends on the hold of their preconditions which are necessary for causalities to influence certain situations. Such a view of causality has been taken in recent studies on reasoning about action [10, 4]. If we look at 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. Whether an causality may actually work in a target situation is essentially irrelevant to how many equivalent corresponding attributes among two situations, but definitely depends on whether the precondition is satisfied in the target. One of the difficulties in implementing such analogy is that whether the precondition is satisfied by the target is often implicit, that is, it may not appear at all in the description of the target situation. For instance, whether a TV set starts to work when we turn on a switch depends on the hold of a precondition that the switch is the power switch of it and the TV set is not out of order. However, we do not know the way to check it and we often do not understand its mechanism, and, consequently, we do not know whether the precondition holds or not. Analogy by simulation might sometimes help even in such a situation by extracting necessary implicit knowledge from understanding the observed phenomena in the base domain, where we have abundant knowledge.

This paper takes such a type of analogical inference and proposes a novel method based on a logical criterion which weakly justifies the conclusion obtained by the analogy. The analogy by simulation proposed here takes analogy, intuitively, as deductive inference from hypotheses extracted by an *illustrated explanation*, by which we mean an explanation (justification) of a certain phenomenon (corresponding to a shared property) based on existing instances (corresponding to bases). A property is projected only when the property is deduced from facts used in an illustrated explanation and the background knowledge. To put it another way, mapping facts, which are used in the process of explanation based on other existing instances (bases), is essential in analogy by simu-

lation, and makes their conclusions nondeductive. A projected conclusion is *illustratively justified* in that its corresponding original conclusion is justified by an existing instance.

The key point which differentiates this method from other works on analogy is that simulation by analogy is done not based on explicit and superficial facts that two objects share a property, but based on hypotheses that these objects both share implicit properties which cause them to share the explicit property. Therefore, if we find that one of the two satisfies such implicit properties, we could conclude that the other would also satisfy these implicit properties.

Here, we clarify analogy by simulation by making use of a formal logic.

A general objective of the research of computational analogy will be to make a system which, given general knowledge including various cases, can answer a query, "Does T satisfy Q ($Q(T)$?)" by making use of proper analogy. For this goal, we need to solve some crucial problems: how we should retrieve a proper and *similar* base case and how we should focus on a *relevant* property among so many similarities. Unfortunately, researchers have not yet found a *good* answer for these. Moreover, before getting down to such problems, we must answer the questions, "What is *similar* ?" and "What is *relevant* ?", with which this paper is concerned. Here, we give a criterion which should be satisfied by the target, T, the base, B, the shared property, P, and the projected property, Q.

In this method, it is assumed that domain knowledge and knowledge about the bases are given abundantly. Let knowledge be a set of first order sentences A, which set is assumed to be divided into three subsets: a set Σ of sentences free from a particular object B and T, a set F_B of sentences in which B occurs but T never occurs, and a set F_T of sentences in which T occurs but B never occurs. Σ is called the *background knowledge*, F_B called the *base knowledge* and F_T called the *target knowledge*.

In this paper, an *explanation* means a minimal deduction path from a certain premise to a particular 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 is an explanation from α to β and if all of the premises 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).

Illustrative Criterion

Given T (for the target), B (for the base), P (for the shared property), Q (for the projected property), and $A(= \Sigma \cup F_B \cup F_T)$ (for knowledge), The *illustrative criterion* is, for some knowledge $f_B(B) \subseteq F_B$ and $S, S' \subseteq \Sigma$,

i) *nontrivial explicability*: $f_B(B), S \vdash_{exp} P(B)$, where no predicate (or corresponding one, ex: *fluent* introduced in the next section) in $P(B)$ occurs in $f_B(B)$,

ii) *relevancy*: $f_B(B), S' \vdash_{exp} Q(B)$, and

iii) *consistency*: $f_B(T) \cup A$ is consistent, where $f_B(T)$ is a result obtained by replacing every occurrence of B with T in a set of sentences $f_B(B)$.

We say $f_B(T)$ is *illustratively justified* by B w.r.t. P, Q when these satisfy the illustrative criterion.

The illustrative criterion requests the three conditions. The first condition, *nontrivial explicability*, is that, from the domain knowledge $S \subseteq \Sigma$ and the base knowledge $f_B(B) \subseteq F_B$, an explanation as to how the base satisfies the shared property must be made $(f_B(B), S \vdash_{exp} P(B))$ (that is, $f_B(B)$ and S are minimal sets used in the explanation) and the explanation must, intuitively, be made *in other words* than the words in $P(B)$. $f_B(B)$ will be called the *implicit precondition* w.r.t. P, and S be the *implicit causal knowledge* w.r.t. P. This condition would play the two roles: one is to select (or *extract*) knowledge which importantly relates to the cause that the base satisfies the explicit similarity. The other is to minimize hypotheses, which helps to make conclusions as plausible as possible.

The second condition, *relevancy*, is that, from an implicit causal knowledge S' and the implicit precondition $f_B(B)$, the conclusion that the base satisfies the projected property Q $(f_B(B), S' \vdash_{exp} Q(B))$ can be *explained*. If it satisfied, Q might be said *relevant* to P, because both can be considered as conclusions from the same reason (precondition) intuitively.

The final condition, *consistency*, is that the essentially projected property $f_B(T)$ is consistent with the original knowledge.

.en the illustrative criterion is satisfied, the following propositions apparently hold.

- $f_B(T), S \vdash_{exp} P(T)$.
- $f_B(T), S' \vdash_{exp} Q(T)$.

The first proposition means that if $f_B(B), S$ used in the explanation is mapped onto the target (that is, it is assumed that $f_B(T)$ holds), the target, also, can be explained to satisfy the explicit shared property in exactly the same way in the base domain. If this mapping proceeds successfully, both domains are considered to be *similar* in that a certain implicit precondition $(f_B(B))$ might hold in both.

The second proposition means that if the implicit precondition $f_B(B)$ is mapped onto the target, the target can be justified as satisfying the projected property, that is, it can be conjectured that the target has a property Q, too.

3 Example

This section applies our method, analogy by simulation, to the example, using representation based on *situation calculus* [11, 10]. In this representation, we use the following *terms* except usual terms corresponding to individual objects.

- *situations*: S_1,... (for constants), s (for variables), corresponding to the complete states of the universe at instants of time.

- *truth-valued fluents*: F, $Strong(x)$,... (for constants), f,... (for variables), corresponding to certain propositions about a situation. Here we use only truth-valued fluents, so we refer to them simply as fluents in the remainder of this paper. Their values are represented by using the situation and a predicate $Holds$, for instance, $Holds(Strong(Brutus), S_1)$ is the value of $Strong(Brutus)$ is $True$, which expresses that $Brutus$ is strong in the situation S_1.

  In this paper, a tuple of terms $t_1, ..., tn$ will be simply written by $\mathbf{t}$. Also, if $\mathbf{f}$ is a tuple of fluents $f_1, ..., f_n$ then, by $Holds(\mathbf{f}, s)$, we mean $Holds(f_1, s) \wedge ... \wedge Holds(f_n, s)$.

- *actions*: A, $Suffers(x, y)$,... (for constants), a,... (for variables). If s is a situation and a is an action then $Result(a, s)$ stands for the situation that results when the action a is performed in the situation s.

Finally, there are two predicate constants involving a predicate introduced above: $Holds$ and $\lambda a, \mathbf{c_f}, \mathbf{r_f} < a; \mathbf{c_f}; \mathbf{r_f} >$, where the last predicate means that $\mathbf{c_f}$ is a tuple of fluents for the precondition of an action a and that, if the precondition holds in some situation s, every fluent in a tuple of fluents $\mathbf{r_f}$ holds in the situation $Result(a, s)$. We call $\mathbf{c_f}$ *prerequisite fluents* and $\mathbf{r_f}$ *resultant fluents*.

Now, we are ready to write our example in this manner.

Axioms of the example can be classified into three groups. The first group consists of a general axiom on causality. This axiom describes that, if $\mathbf{c_f}$ is the prerequisite fluents of an action a and $\mathbf{r_f}$ is the resultant fluents of it, then the resultant fluents hold in a situation $Result(a, s)$ when the prerequisite fluents hold in a situation s:

$$< a; \mathbf{c_f}; \mathbf{r_f} > \supset \forall s.(Holds(\mathbf{c_f}, s) \supset Holds(\mathbf{r_f}, Result(a, s))). \qquad [\Sigma 1.1]$$

The second group describes tuples of <action;precondition;result>.

$\langle$ **action**	; **prerequisite fluents**	; **resultant fluents**	$\rangle$	
$\langle Suffers(x, y)$	; $Has(x, TheSenseOfTouch)$, $PhysicallyDestructive(y)$	; $FeelingPain(x)$, $PhysicallyDamaged(x)$	$\rangle$	[Σ2.1]
$\langle Tickles(y, x)$	; $Has(x, TheSenseOfTouch)$	; $Laughing(x)$	$\rangle$	[Σ2.2]

•
•
•

The last grope describes phenomena which happened to each individual. It is further divided into the three subsets. The first subset belongs to the background knowledge:

$$\forall s.Holds(PhysicallyDestructive(Cut), s), \qquad [\Sigma\ 3.1]$$
$$\forall s.Holds(PhysicallyDestructive(Burn), s), \qquad [\Sigma\ 3.2]$$

•
•

•

[Σ 3.1] and [Σ 3.2] represent that a cut and a burn both are physically destructive.

The second subset of the last group belongs to the base knowledge. Let us know well what poperties Brutus (the base) satisfies ($F_B(Brutus)$):

$$\forall s.Holds(Has(Brutus, TheSenseOfTouch), s)), \quad [F_B3.1]$$
$$\forall s.Holds(Strong(Brutus), s), \quad [F_B3.2]$$
$$\forall s.Holds(Has(Brutus, TwoEyes), s), \quad [F_B3.3]$$

•
•
•

The above sentences mean that Brutus is able to feel pain, strong, and has the two eyes.

The last subset of the last group belongs to the target knowledge. We do not know much about Tacitus (the target) ($F_T(Tacitus)$):

$$\forall s.(Holds(FeelingPain(Tacitus), Result(Suffers(Tacitus, Cut), s)), \quad [F_T3.1]$$
$$\forall s.Holds(Has(Tacitus, TwoEyes), s), \quad [F_T3.2]$$

[F_T3.1] describes that Tacitus feels pain as the result of being cut. [F_T3.2] describes that Tcacitus has the two eyes.

Now, assuming theses sentences, what we want to predict is whether
$\forall s.Holds(FeelingPain(Tacitus), Result(Suffers(Tacitus, Cut), s))$, hspace10mm
which expresses that Tacitus would feel pain when he is cut.

The following is a detail of a simulation method. It can be divided into two parts: the first step is to find $f_B(Brutus)$ in the case of the satisfied illustrative criterion, and the last step is to map the projected property satisfying the criterion. Further, the first step might be divided into the following three steps according to the three conditions in the illustrative criterion. Here, we assume that the illustrative ground similarity P and the projected property Q are given. That is,

$P(Brutus)$: $\forall s.(Holds(FeelingPain(Brutus), Result(Suffers(Brutus, Cut), s))$
$Q(Brutus)$: $\forall s.(Holds(FeelingPain(Brutus), Result(Suffers(Brutus, Burn), s))$

(1) Understanding Step (nontrivial explainability):
From the domain knowledge Σ and the base knowledge F_B, an explanation of how the base satisfies the shared property is made ($f_B(Brutus), S \vdash_{exp} P(Brutus)$) ($f_B(Brutus) \subseteq F_B$ and $S \subseteq \Sigma$ are minimal sets used in the explanation).

S: [Σ1.1], [Σ2.1], [Σ3.1]
$f_B(Brutus)$: [F_B3.1]

(2) Justifying Relevance Step (relevancy):

From the background knowledge Σ and the implicit precondition $f_B(Brutus)$, this step tries to *explain* how the base satisfies the projected property Q. If it fails, it returns to the Understanding Step and seeks other $f_B(Brutus)$ (that is, selects other explanation).

S': $[\Sigma 1.1]$, $[\Sigma 2.1]$, $[\Sigma 3.2]$
$f_B(Brutus)$: $[F_B 3.1]$

(3) Consistency Checking Step (consistency):

The extracted implicit precondition $f_B(Brutus)$ used in the above explanation is mapped onto the target (that is, it is assumed that $f_B(Tacitus)$ holds) and is tried to check not to cause inconsistency with the original whole knowledge. If it fails, it returns to the Understanding Step and seeks other $f_B(Brutus)$, and otherwise, $f_B(Brutus)$ is output and $Q(Tacitus)$ is illustratively justified.

After all, $Q(Tacitus)$ is illustratively justified by *Brutus* w.r.t. P, so we can obtain $Q(Tacitus)$ as the result of analogy by simulation in the last step.

Here, note that other possible properties which are illustratively justified are

$$\forall s.Holds(Has(Tacitus, TheSenseOfTouch), s), Holds(PhisicallyDamaged(Tacitus), s), ...$$

and the following sentence is also predictable:

$$\forall s.Holds(Laughing(Tacitus), Result(Tickle(Someone, Tacitus), s)).$$

That is, from the fact that Tacitus feels pain when he is cut, it is predicted by our simulation method that Tacitus would satisfy the above properties.

However, a property which does not relate the explanation in the base domain, like $Strong(Brutus)$, is prohibited from being projected. Moreover, in this case, another shared property that both Brutus and Tacitus have the own two eyes can be said *irrelevant* to the projected property, that one feels pain when he is burnt. Because the pair of the both properties will not satisfy the illustrative criterion (We will not be able to find an explanation which relates the fact that Brutus has the two eyes with the fact that Brutus feels pain when he is burnt).

Another interesting feature of this method is that simulation analogy is based on a hypothesis that the base and the target both share implicit properties which cause them to share an explicit property. So, if we come to know that the target does not satisfy such an implicit property, then we lose the reason and conclusions ever obtained by our method are invalidated, even when we do not find inconsistency of the conclusions.

4 Conclusion and Remarks

This paper proposes a novel logical method for analogical reasoning.

This method also gives an answer to *the non-redundancy problem* pointed out by Davies *et al.* [2], the base instance should provide new information about the conclusion. A nondeductive conclusion obtained by this method (for instance,

$Has(Tacitus, TheSenseOfTouch)$, $Damaged(Tacitus)$, ... in the above example) is a projected property shared by the base (*Brutus*), which is obtained from the base knowledge (F_B) and the background knowledge. That is, the base information is actually used.

This method is general in that it is a logical approach independent of any particular 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 preferable certain type of analogy like the example of cars shown by T.R.Davies *et al.* [2], which is called *functional analogy* [1], in which, using the similarity that the type of both cars is *Mustang*, the value of one is conjectured from the value of the other. The reason this method could not be applied to such analogy is that we would not be able to find a *explanation* of how the base satisfies the similarity. (What effective explanation can we find of how a car is a mustang?)

Acknowledgements

I am grateful to Natsuki Oka, Katsumi Inoue, Dr. Makoto Haraguchi and Dr. Koichi Furukawa for their useful comments.

References

[1] 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).

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

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

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

[5] Grainer,R.: Learning by understanding analogy, *Artificial Intelligence*, Vol. 35 (1988) 81-125.

[6] Haraguchi,M.: Analogical reasoning using transformation of rules, *Bull. Inform. Cybernetics*, Vol. 21 (1985).

[7] Horgan,J.P.: The two faces of tomorrow, Tokyo: Tokyo Sogensha (New York: Ballantine Books INC.) (1979)

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

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

[11] McCarthy,J. and Hayes,P.: Some philosophical problems from the standpoint of artificial intelligent, in Meltzer,B and Michi,D.(Eds),*Machine Intelligence 4* Edinburgh: Edinburgh University Press (1969) 463-502.

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

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