In Proceedings of the Sixth International Conference on Cognitive Modeling, 202-207. Mahwah, NJ: Lawrence Earlbaum.

A Computational Analysis Model for Complex Open-ended Analogical Retrieval

Junya Morita (morita@cog.human.nagoya-u.ac.jp)
Graduate School of Human Informatics
Nagoya University, Furo-cho, Chigusa-ku, Nagoya City, Japan

Kazuhisa Miwa (miwa@cog.human.nagoya-u.ac.jp)
Graduate School of Information Science
Nagoya University, Furo-cho, Chigusa-ku, Nagoya City, Japan

Abstract

Although many computational models have successfully simulated the results of controlled psychological experiments, few researchers have attempted to apply their models to complex, realistic phenomena. In this study, MAC/FAC ("many are called, but few are chosen"), which models two stages of analogical reasoning (Forbus, Gentner, & Law, 1995), was applied to our experimental data. In our experiment, subjects were presented a cue story and asked to retrieve cases learned from everyday life. Next they rated the inferential soundness (goodness as an analogy) of each retrieved case. For each retrieved case, we used the algorithms of the MAC/FAC to compute two kinds of similarity scores: content vectors and structural evaluation scores. As a result, the computed content vectors explained the overall retrieval of cases well, whereas the structural evaluation scores had a strong relation to the rated scores. These results support the MAC/FAC's theoretical assumption – different similarities are involved in the two stages of analogical reasoning.

Introduction

In the past, many cognitive psychologists have conducted controlled experiments that have lead to scientific understanding of analogical reasoning. The impetus behind these studies might have been the development of computational models (Falkenhainer, Forbus, & Gentner, 1989; Forbus, Gentner, & Law, 1994). Such models have been used to simulate the results of psychological experiments, and they have also guided subsequent studies.

However, we think the above studies are limited because most of the simulated psychological data has been obtained in closed laboratory situations. Thus it is unclear whether the models are suitable for simulating realistic analogical reasoning. Our aim here is to apply a computational model of analogy to more complex and realistic psychological data. Prior to presenting our study, we review previous studies on analogical reasoning.

Framework for analogy research

Analogical reasoning involves two representations: the base and the target. The base is a past case that one is familiar with. The target is a novel case that one is usually less familiar with. The process of analogical reasoning is comprised of two main components: the retrieval of the base and the mapping from the base to the target; the analogy process is guided by similarities between the base and the target. Using propositional representations (predicate-argument formalism), Gentner (1983) distinguished three types of correspondence between the base and the target.

Correspondence of *attributes*: e.g., The sun is round and yellow → The orange is round and yellow [sun (round) sun (yellow) → orange (round) orange (yellow)].

Correspondence of *first-order relations*: e.g., The planets revolve around the sun. → The electrons revolve around the atom [revolve-around (planet, solar) → revolve-around (electron, atom)].

Correspondence of *higher-order relations*: e.g., Because the sun attracts the planets, the planets revolve around the sun. → Because the atom attracts the electrons, the electrons revolve around the atom [cause (attract (solar, planet), revolve-around (planet, solar)) → cause (attract (atom, electron), revolve-around (electron, atom))].

The above discrimination was based on the types of predicates. The attribute is a predicate type that takes a single argument. On the other hand, the first-order and higher-order relations are predicate types that take multiple arguments. There is no depth in the former, but there is in the latter (Gentner, 1983). Therefore, in the studies on analogy, the former is often called surface similarity, while the latter is called structural similarity.

Computational Model of Analogy

Forbus, Gentner, & Law (1995) developed a computational model called MAC/FAC ("many are called, but few are chosen") that aims to simulate the two stages of analogical reasoning: retrieval and evaluation.

In the first stage of the MAC/FAC model called the MAC stage, several candidate bases are recalled from the memory pool. For quick access to the large number of cases in memory, the MAC stage implements a computationally cheap process. Using content vectors (CVectors), which are simple lists of predicates, each memory item is matched with the target. The outputs of the MAC stage are memory items that have a high dot product of CVectors.

CVectors indicate which predicates are contained and how frequently those predicates appear in the representation. Because CVectors do not distinguish between attribute and relation, the dot product is influenced by the commonalities of the attribute as well as by the commonalities of the relation. Also, since the CVectors have nonstructural features, the dot products tend to overesti-

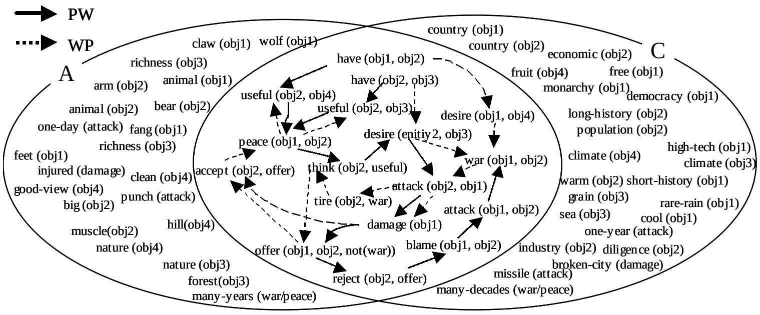

Figure 1: Propositions contained in the target stories.

Table 1: Predicate types shared by target stories.

	A/PW	A/WP	C/PW	C/WP
A/PW	OA+FOR+HOR	OA+FOR	FOR+HOR	FOR
A/WP		OA+FOR+HOR	FOR	FOR+HOR
C/PW			OA+FOR+HOR	OA+FOR
C/WP				OA+FOR+HOR

Note. OA=object attributes; FOR=first-order relation; HOR=higher-order relation.

Table 2: CVectors/SESs between target stories.

	A/PW	A/WP	C/PW	C/WP
A/PW	1.00/28.8	1.00/12.2	0.39/28.8	0.39/12.2
A/WP		1.00/27.9	0.39/12.2	0.39/27.9
C/PW			1.00/28.8	1.00/12.2
C/WP				1.00/27.9

mate structurally dissimilar representations. For example, two propositions ["cause (attract, revolve-around)" and "cause (revolve-around, attract)"] have the same CVector [cause = 1, attract = 1, revolve-around = 1]. Also, the proposition "war (dog1, dog2)" [war = 1, animal = 2] is more similar to a proposition "war (wolf1, wolf2, wolf3)" [war = 1, animal =3] than "war (wolf1, wolf2)" [war = 1, animal =2].

Thus, the candidate bases are further evaluated at the next step called the FAC stage. The FAC stage is modeled by using SME, the *structure-mapping engine* (Falkenhainer, Forbus, & Gentner, 1989). The SME includes the following two sub processes.

Construction of *a local match*, which is a pair of predicates shared by the base and the target. A numerical weight is assigned to each local match according to its relational consistency.

Construction of a *global match*, which is an overall mapping from the base to the target. The global match is constructed by connecting structurally consistent combinations of local matches, satisfying the constraint of *one to one mapping*, which means that the same elements in the base can not match multiple items in the target, or vice versa. For the constructed global match, the structural evaluation score (SES) was computed by summing the numerical weights of the local matches that are included in the global match. The outputs of the FAC stage are the cases that have a high SES.

To summarize, the MAC/FAC assumed a distinction between the two stages of analogical reasoning guided by the different similarities.

Psychological study of analogy

Gentner, Ratterman, & Forbus (1993) provided psychological evidence for the MAC/FAC's assumption. They used story sets manipulated by types of similarity (experiment 2). Each story set was comprised of a base story and four target stories. The target stories shared various predicates with the base story: a surface similarity (SS) that shared attribute and first-order relations, an analogy (AN) that shared first-order and higher-order relations, a literal similarity (LS) that shared all types of predicates, and a story that shared the first-order relations (FOR).

In their experiment, subjects first learned the base story. About a week later, they were presented with the target stories as retrieval cues to help them recall the base stories. Finally they rated the inferential soundness (goodness as an analogy) for each pair of stories. The results implied that different similarities are involved in the process of analogical reasoning. The subjects more often retrieved the base story when they read surface similar stories (LS, SS), rather than when they read structurally similar stories (AN, FOR). However, when evaluating soundness, they rated structurally similar stories (LS, AN) higher than surface similar stories (SS, FOR).

Gentner, Ratterman, & Forbus applied their computational model to the above result. First, they input the base and the targets into the SME and computed the

SES between stories. The results showed that the SME is a good predictor of subjective soundness, suggesting that the story pairs that shared higher-order relations have a higher SES than the other story pairs that did not share the higher-order relations (LS, AN > SS, FOR). In addition, they constructed a memory pool that contained the four target stories (LS, AN, SS, and FOR) and input the base story into the MAC/FAC model as a retrieval cue. As a result, the order of the retrieval rate by the MAC/FAC became LS>SS>AN>FOR. This pattern of retrieval is consistent with the result of human retrieval.

The results of their simulation support the assumed algorithms of the MAC/FAC. However, we think there are limitations to their investigation because they have only studied cases created by the researchers themselves. In real-world situations, individuals make analogy from everyday experience. To apply the model to realistic problems, it is necessary to investigate analogical reasoning using cases that the subjects experienced in everyday life. Therefore, we tested the validity of the algorithms of the MAC/FAC model by applying the model to cases that subjects learned from their everyday life.

Experiment

Materials

Unlike Gentner, Ratterman, & Forbus's study, we did not prepare base stories. Without learning any stories, subjects were presented with a target story as a retrieval cue. The subjects were asked to report remembered cases that came to mind while reading the target story, which consisted of about 600 Japanese characters.

The surface and structural features of the target stories were manipulated. As surface features, a set of attributes related to *animals* (A) and a set of attributes related to *countries* (C) were chosen. As structural features, a story whose plot involved a transition from *peace to war* (PW) and another whose plot was a transition from *war to peace* (WP) were created. Combining the surface and structural features, four types of target stories were prepared: A/PW, A/WP, C/PW, and C/WP.

Figure 1 illustrates the propositions converted from the texts in the target stories. Each of them was included in either set A or set C. Two complements $[(A \cap \bar{C})$ and $(\bar{A} \cap C)]$ contain object attributes, and an intersection $(A \cap C)$ includes first-order relations of two objects. Each of the first-order relations was connected by two types of higher-order relations (PW/WP), represented by two types of arrows (solid/dotted). That is, as the materials in Gentner, Ratterman, & Forbus's study, the four target stories shared first-order relations, varying the attributes and the higher-order relations. Table 1 summarizes the interrelations between the target stories.

Additionally, the similarity scores between the target stories, presented in Table 2, were calculated based on the algorithms of the MAC/FAC. The CVectors were computed as a dot product of the predicate lists, which includes the attributes and the first-order relations as

Subjects' descriptions	Converted propositions
The story about two tigers. In a forest, two tigers lived. Each of them has a turf. And they battled each other for the turf. One day, an animal that lived in the forest persuaded one of the tigers to stop fighting. After this persuasion, the relationship between the two tigers became peaceful.	((animal tiger1) :name prop3) ((animal tiger1) :name animal2) ((animal animal1) :name animal3) ((have tiger1 turf1) :name have1) ((have tiger2 turf2) :name have2) ((desire tiger1 turf2) :name desire1) ((desire tiger2 turf1) :name desire2) ((war tiger1 tiger2) :name war1) ((and desire1 desire2) :name and1) ((cause and1 war1) :name cause1) ((not war1) :name not1) ((offer animal1 tiger1 not1) :name off) ((accept tiger1 off) :name accept1) ((cause offer1 accept1) :name cause2) ((cause war1 off) :name cause3) ((peace tiger1 tiger2) :name peace1) ((cause accept1 peace1) :name cause4) ((many-tree turf1) :name prop1) ((many-tree turf2) :name prop2)

Figure 2: Examples of coding.

components[1]. The SES was computed by using the SME model with an analogy rule that constructs structurally consistent mapping without matching attributes[2]. The scores shown in Table 2 are consistent with the manipulation of the target stories. The CVectors of the pairs sharing attributes are greater than those of the pairs sharing no attributes, and the SESs of the pairs sharing higher-order relations are greater than those of the pairs sharing no higher-order relations.

Furthermore, we conducted a pilot experiment to test whether the above SESs are consistent with human feelings of soundness. After the subjects (n = 8) received the four target stories, they rated the inferential soundness of each pair on a 1 ("low") – 5 ("high") scale. As in Gentner, Rattermann, & Forbus's study, soundness was explained as "the degree to which inferences from one story would hold for the other." The results supported the validity of the SME's evaluation scores. Seven of eight subjects judged the structurally similar pairs (A/PW vs. C/PW and A/WP vs. C/WP) as having higher inferential soundness than the other pairs (A/PW vs. A/WP, CPW vs. C/WP, A/PW vs. C/WP, and A/WP vs. CPW).

Participants

Thirty-three undergraduate and graduate students participated in the experiment. They were divided into four groups varying target stories: a group presented with A/PW (n = 8), a group with A/WP (n = 9), a group with C/PW (n = 8), and a group with C/WP (n = 8).

[1] The CVectors were normalized to unit vectors, as Forbus, Gentner, & Law did.

[2] Forbus, Gentner, & Law (1995) computed the SES using a literal-similarity rule that mapped all types of predicates. However, their simulation showed that the CVector predicted results of human retrieval relatively well, and the SME with the analogy rule predicted the soundness evaluation well. Thus, we treated the CVector as a matcher for the initial retrieval, and the SME with the analogy rule as a matcher for the evaluation of soundness.

Procedure

The subjects participated in the experiment individually or in groups of two to four. The experiment was divided into the following three phases.

Retrieval phase: The subjects were presented with one of the four target stories. Then they were told that "while reading the presented story, you should write out any cases from your everyday life that come to mind." After the instructions, they wrote down for twenty minutes any remembered cases.

Evaluation phase: Following completion of the retrieval phase, the subjects rated the soundness of the match between each retrieved case and the presented target story on a 1 – 5 scale.

Explanation phase: Finally, the subjects were asked to explain the retrieved cases in as much detail as possible.

Applying MAC/FAC to the Data

We obtained two types of data from the experiment: cases retrieved by the subject, and the soundness scores rated by the subjects themselves. These two types of data correspond to the two stages of the MAC/FAC model. We directly applied the model to the data to test the correspondence between the data types and the MAC/FAC stages.

First, all of the retrieved cases were converted to propositional representations. The subjects' descriptions were segmented by the appearance of a predicate, and then a coder judged whether each segmented sentence could be represented as a proposition by using predicates contained in the target stories. When possible, a proposition was constructed by complementing proper arguments. While working, the coder used the list that defines the correspondence between the predicates and each word included in the subjects' description. Figure 2 illustrates examples of the coding.

For the converted representations, two scores of similarity (CVector/SES) were computed using the MAC/FAC model. We also computed the two scores of similarity of the retrieved cases with the *target story that was presented to each subject*, as well as the two scores of similarity with the *target story that was not presented to each subject.* The MAC/FAC predicts that the overall average CVector with the presented target story will be higher than the overall average CVector with the target story that was not presented, and the SES with the presented target story will be closely related to the rated scores of soundness, rather than the SES with the target story that was not presented.

Results and Discussion

This section shows the results of applying MAC/FAC to our data[3]. First, we illustrate the overall similarity scores for the experimental groups to test the assumption at the MAC stage (an initial retrieval stage using the CVectors). Next we examine the relationship between the similarity scores and the soundness scores to test the assumption of the FAC stage, which is an evaluation stage using the SME model.

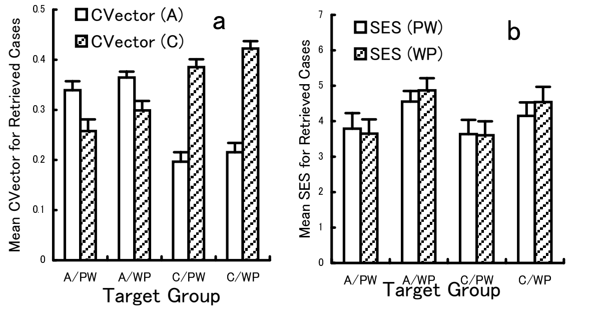

Figure 3: (a) Mean CVector for four groups. (b) Mean SES for four groups. *Note.* Error bars represent one standard error of mean.

1. Test for the MAC stage

To test the assumption at the MAC stage, we computed four scores of similarity for each retrieved case:

CVector (A): The dot product of CVector of the retrieved case and the CVector of the propositional set A (represented as [animal = 2, wolf =1, bear = 1, forest =1 ... have = 2, desire = 2]).

CVector (C): The dot product of CVector of the retrieved case and the CVector of the propositional set C (represented as [country = 2, democracy = 1, monarchy = 1 ... have = 2, desire = 2]).

SES (PW): The SES computed by inputting the retrieved case and the structural feature PW (solid arrows in Figure 1) into the SME.

SES (WP): The SES computed by inputting the retrieved case and the structural feature WP (dotted arrows in Figure 1) into the SME.

We conducted two analyses of variance to investigate the interaction between the above similarity scores and the experimental groups. If the assumption of the MAC stage is reasonable, the CVector (A) should be higher than the CVector (C) in the subjects who were presented with the surface feature A, and the CVector (C) should be higher than the CVector (A) in the subjects who were presented with the surface feature C. On the other hand, it is predicted that the ANOVA using the SES will not reveal clear difference between the similarity scores.

The CVector as a matching algorithm at the MAC stage Figure 3a shows the mean CVector for each group. A 2 × 2 × 2 surface features of target stories (between) × structural features of target stories (between) × types of CVector (within) ANOVA revealed a significant interaction between the surface features of the target stories and the types of CVector [$F(1, 262) = 206.77, p < .05$]. This indicates that the CVector (A) was higher than the CVector (C) in group A [$F(1, 262) = 30.26, p < .05$], and that the CVector (C) was higher than the CVector (A) in group C

[3]The total number of cases retrieved by the subjects was 266. There was no significant difference on the number of cases among the four experimental groups [$\chi^2(3) = 6.15, ns.$]. Thus, we treated each retrieved case as an individual datum for statistical tests.

Table 3: Types of correspondence to soundness scores.

	OA	FOR	HOR	1-1	r
CVector (presented)	yes	yes	no	no	0.23 ($p < .01$)
CVector (not presented)	no	yes	no	no	0.32 ($p < .01$)
SES (presented)	no	yes	yes	yes	0.44 ($p < .01$)
SES (not presented)	no	yes	no	yes	0.38 ($p < .01$)

Note. OA = object attributes; FOR = first-order relation; HOR = higher-order relation; 1-1 means constraint of one to one mapping.

[$F(1, 262) = 220.07, p < .05$]. The retrieved cases were estimated to be more similar to the target story that was presented to the subject than the target story that was not presented to the subject. This result is consistent with the prediction, supporting the assumption of the MAC stage.

The SME as a matching algorithm at the MAC stage Figure 3b shows the mean SES for each group. A 2 × 2 × 2 surface features of target stories (between) × structural features of target stories (between) × types of SES (within) ANOVA revealed a significant interaction between the structural features of the target and the types of SES [$F(1, 262) = 8.01, p < .05$]. However, a simple effect was significant only in group WP [$F(1, 262) = 7.50, p < .05$]. There was no significant difference of types of SES in group PW [$F(1, 262) = 1.60, ns.$]. Those results are not distinctive when compared with the results of the ANOVA using the CVectors. Therefore, it is suggested that the SME, which is assumed as a matching algorithm at the FAC stage, is insufficient to predict the initial retrieval.

2. Test for the FAC stage

To investigate the evaluation of the soundness as the psychological data that corresponds to the FAC stage, we treated the subject groups as counterbalance conditions and reduced the number of factors for the statistical test. Therefore, we computed the following four scores.

CVector (presented) was computed by combining the CVector (A) in group A and the CVector (C) in group C. This score indicates a rough estimate of the overlap between the retrieved cases and the target story that was presented to the subject.

CVector (not presented) was computed by combining the CVector (C) in group A and the CVector (A) in group C. This score indicates a rough estimate of the overlap between the retrieved cases and the target story that was not presented to the subject.

SES (presented) was computed by combining the SES (PW) in the group PW and the SES (WP) in the group WP. This score indicates the depth and breadth of the structural mapping from the retrieved cases to the target story that was presented to the subject.

SES (not presented) was computed by combining the SES (WP) in the group PW and the SES (PW) in the group WP. This score indicates the depth and breadth of the structural mapping from the retrieved cases to the target story that the subject did not receive.

Each score was interpreted as shown in Table 3. The CVector (presented) reflects the commonality of the object attributes (see the second column of Table 3). The SES (presented) reflects the commonality of the higher-order relations (see the fourth column of Table 3). All scores reflect the commonality of the first-order relations (see the third column of Table 3). In addition to the discrimination of predicate types, there is a difference concerning the 1-1 constraint (see the fifth column of Table 3). The CVectors are rough estimations of overlap, which may overestimate the one-to-many and many-to-one mappings. On the other hand, the SES is strictly calculated to satisfy the 1-1 constraint.

The sixth column of Table 3 shows the correlation coefficients between the soundness scores and the similarity scores. There are significant positive correlations in all of the scores, which might indicate that the soundness scores are influenced by the correspondence of the first-order relations, which all of the scores reflect.

To investigate the relation in more detail, we conducted two 5 × 2 ANOVA (soundness scores 1 − 5 × presented/not presented), using the CVectors and the SESs as dependent measures.

The CVector as a matching algorithm at the FAC stage Figure 4a shows the mean CVector for each score of soundness (1 − 5). A 5 × 2 soundness scores (between) × CVector types (within) ANOVA detected a main effect of soundness scores [$F(4, 261) = 9.25, p < .05$.], and a main effect of CVector types [$F(1, 261) = 181.64, p < .05$]. There was no significant interaction between the soundness scores and the CVector types [$F(4, 261) = 1.59, ns.$].

The main effect of the CVector types indicates an advantage for the CVector (presented) over the CVector (not presented), regardless of the soundness scores. Therefore, it is suggested that the soundness scores have no relation to the commonality of attributes, which the CVector (presented) reflects (see the second column of Table 3).

Furthermore, multiple comparisons of the soundness scores revealed significant advantages of scores 2, 3, 4, and 5 over score 1, and significant advantages of score 4 over the scores 2 and 3 ($p < .05$), indicating that the CVector increased with the soundness scores. Thus, as shown in the correlation coefficients (see the sixth column of Table 3), the results support a positive relation

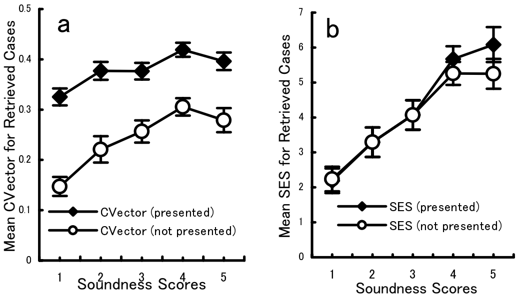

Figure 4: (a) Mean CVector for soundness scores. (b) Mean SES for soundness scores. *Note.* Error bars represent one standard error of mean.

between the soundness scores and the correspondence of the first-order relations (see the third column of Table 3).

The SME as a matching algorithm at the FAC stage Figure 4b shows the mean SES for each score of soundness (1 – 5). A 5 × 2 soundness scores (between) × SES types (within) ANOVA revealed a significant interaction between the soundness scores and the SES types $[F(4, 261) = 7.52, p < .05]$.

The simple effects of the soundness scores were significant for both the SES (presented) $[F(4, 261) = 15.46, p < .05]$ and the SES (not presented) $[F(4, 261) = 11.09, p < .05]$. Furthermore, multiple comparisons for both of the two scores revealed differences between score 1 and scores 3, 4, and 5, between the score 2 and the scores 4 and 5, and between the score 3 and the scores 4 and 5. Compared with the results of the analysis that used the CVectors, there are many significant differences between the soundness scores. Since the SES was calculated to satisfy the 1-1 constraint, unlike the CVector (see the fifth column of Table 3), this result supports the assumption of the FAC stage, which computes accurate structural mapping.

Finally, it was confirmed that the SES (presented) was higher than the SES (not presented) in the cases in which the subjects rated high soundness (scores 4 $[F(1, 261) = 8.76, p < .05]$ and 5 $[F(1, 261) = 35.73, p < .05]$), suggesting the SES (presented) is more strongly related to soundness than the SES (not presented). Since the difference between the two scores is in the higher-order relations (see the fourth column of Table 3), this suggests that computing a higher-order relation is needed for models of evaluation in analogical reasoning.

General Discussion

In summary, our results support the model's assumption. CVectors were good indicators of the overall difference in the initial retrieval between the experimental groups. On the other hand, the SES reflected a strong relation to the soundness scores. Importantly, the model's algorithm predicted the complex open-ended process of analogical reasoning.

Additionally, we would like to assert the psychological importance of our study. Our results are consistent with those of Gentner, Ratterman, & Forbus's study, which showed the different similarities involved in the two stages of analogical reasoning. However, recently, Blanchette & Dunbar (2000) showed that surface similarity has little effect on analogical retrieval in situations where subjects retrieve the cases that they have learned in everyday life. Their experiment asked subjects to generate analogies to the zero-deficit problem encountered by the Canadian government. As a result, the subjects generated few analogies that had surface features in common with the zero-deficit problem, but they generated many analogies that shared deep structures with the target.

Although there are many differences between our experiment and Blanchette & Dunbar's experiment, the difference in results might be explained by a difference in analysis. Blanchette & Dunbar analyzed generated analogies by categorizing their surface/structural features and comparing the frequencies of the categories. On the other hand, we computed the degree of similarity using a computational model. There is a clear difference of resolution between our analysis and Blanchette & Dunbar's analysis. Our detailed, theory-based analysis might detect the effect of surface similarity on the retrieval.

Thus, this paper defends not only the validity of the MAC/FAC model but also proposes a novel and useful method of investigating complex cognition. In the past, many researchers have analyzed complex, open-ended data using such methods as categorizations. There have been few attempts to apply a computational model directly to psychological data. However, as demonstrated in this paper, using a computational model for analysis could open a new way of firmly validating the connection between a theory and data. Without sufficient models, it is impossible to instantiate complex cognitive conceptual products such as surface similarity or structural similarity.

References

Blanchette, I., & Dunbar, K., (2000). How analogies are generated: The role of structural and superficial similarity. *Memory & Cognition*, *28*, 108–124.

Falkenhainer, B., Forbus, K. D., & Gentner, D., (1989). The Structure-Mapping Engine: Algorithm and Example. *Artificial Intelligence*, *41*, 1–63.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*, 141–205.

Gentner, D. (1983). Structure-Mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.

Gentner, D., Rattermann, M., & Forbus, K. D. (1993). The role of similarity in transfer: Separating retrievability for inferential soundness. *Cognitive Psychology*, *25*, 524–575.

Thagard, P., Holyoak, K. J., Nelson, G., & Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence*, *46*, 259-310.