Spatial Representation and Reasoning about Fold Strata: A Qualitative Approach

Yuta Taniuchi[1] and Kazuko Takahashi[1]

Kwansei Gakuin University, 1, Gakuen, Uegahara, Sanda, 669-1330, Japan
{hcs57846,ktaka}@kwansei.ac.jp

Abstract. This paper proposes the method that handles strata based on qualitative spatial reasoning. We make a model for typical fold structure projected onto a two-dimensional plane, extracted by a rectangle. We give a symbolic representation to the model with the features of qualitative configuration and qualitative shape, and propose a reasoning method on this representation. First, we define the validity required of the representation and show the correspondence between the model and its representation. Next, we define operations such as rotation and symmetric transitions on the representation and show that they preserve the validity. Finally, we define the rules of connecting the models, and show reasoning on construction of global data by applying them. When multiple local data collected in distant locations are given, we can find global data by inserting missing parts. The approach based on qualitative spatial reasoning provides a logical explanation of the processes involved in strata-generation prediction, which in the field of structural geology have been examined manually to date, and enables to find results that manual analysis may overlook.

Keywords: Qualitative spatial reasoning · Knowledge representation · Logical reasoning · Shape information.

1 Introduction

Natural disasters and topographical changes are closely related. Landslides and floods caused by heavy rain can be predicted from the lanscapes, and it is required to investigate the topography for safety when doing civil engineering and construction work. The structure of strata and its formation process are important factors to know the temporal change of topography.

In structural-geology research [9], the shapes and structures of strata are analyzed using data at various scales, from the micro level, such as collected small sample data measured in tens of centimeters, or slices that can be observed by microscopy, to the macro level at the out-crop scale of several-hundred meters, or aerial photos of larger regions. Regardless of scale, the entire shape of a stratum is estimated by integrating local data collected from multiple locations, since it is rarely exposed in a real landscape. There is no systematic approach in this estimation process, and human error may exist.

In this study, we propose a systematic approach using qualitative spatial reasoning (QSR), which is a subfield of artificial intelligence. QSR represents spatial entities symbolically without using concrete numerical data, and enables reasoning on the representation [4, 3, 13, 16]. Representation focuses on specific aspects or properties of an object or the relation of objects, depending on the user's purpose, such as mereological relations, the relative positions or directions of objects, rough shapes, and on on. Avoiding the need for precise values enables a small computational burden, and declarative representation suits human recognition. So far, lots of works have been done depending on the focused aspects of spatial data. Here, we focus on the shapes and configuration of objects.

Although it is rather difficult to consider shape in QSR, several researchers have proposed handling the shape of an object by projecting it onto a two-dimensional plane [5, 6, 8, 10–12, 2, 1, 15, 18]. In most of these works, a set of primitives was introduced and the shape of the object was represented by arranging these primitives in the order of their occurrence when tracing the outline of the object. This process indicates that the target is essentially one-dimensional spatial data.

On the other hand, for our application, we have to consider representation based on local data extracted from a stratum, since the entire data do not comprise a closed curve. Moreover, we have to represent not only the shapes of layers that become regions of a two-dimensional plane but also their interconnections. Therefore, we cannot apply existing methods.

In this study, we propose representation and reasoning for a fold as a relatively simple strata structure. To apply QSR to the shapes of strata, there are two primary requirements: one layer continues in one direction if there is no fault, and the relations of inter-layer connections remain unchanged even if a stratum rotates or bends.

First, we define a model for local fold data and the language to describe it. Next, we define the validity required of the representation, and show that the model representation is valid and that a figure can be drawn on a two-dimensional plane for the valid representation. Moreover, we define operations on the representation corresponding to rotation and symmetric transitions, and show that the validity is preserved. Finally, we discuss the interconnection of models that have the same strata configurations. We show the process in which global data is obtained by connecting multiple local data, which was not sufficiently discussed in our previous work [17]. There, we mainly described the horizontal connection because of the page limit. In this paper, we formalize vertical connection precisely and discuss a sequence of connections as the reasoning process on this representation. We also show an application of predicting global data from multiple local data collected at different locations to solve the questions such as "Are they parts of the same large global strata?" or "Can we infer an ancient global stratum by connecting these local data?" It leads to the derivation of spatial relations among multiple local data collected in different places or at different times.

This study provides a mechanical treatment of strata using symbolic representation that focuses on their features. The approach based on qualitative spatial reasoning can provide logical explanations of processes that may be involved in future morphological changes that manual analysis may overlook.

This paper is organized as follows. In Section 2, we identify our target fold and the model. In Section 3, we define a description language. In Section 4, we provide a procedure to generate a representation for a model, and show that the representation and the model have a one-to-one relation. In Section 5 we define operations on this representation. In Section 6, we discuss reasoning on this representation. In Section 7, we compare our study with related works. Finally, in Section 8, we show our conclusions and future works.

2 Model

We describe a typical form of fold strata such as that shown in Figure 1(a) [14]. We assume that there is no fault or hole, and that the curvature of all the layers is the same. We model a vertical cross section of the fold projected onto a two-dimensional plane. We derive the local data extracted from the global data by a rectangle that satisfies the following conditions **[COND]**. Based on these conditions, the fold is divided into regions using multiple smooth continuous curves (called *layer-borderlines*). Pairs of layer-borderlines do not intersect and there is no self-intersection. We treat this figure as our model.

(a) fold form (b) model on a 2D plane

Fig. 1. A model for a fold [17].

[COND]

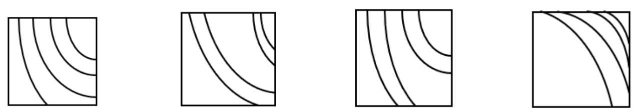
1. All layers and any space (a region containing no layer) in the global data appear to be connected regions in the local data.
2. The end-points of each layer-borderline are not located on a corner of the rectangle.
3. Each layer-borderline is a smooth curve with neither an extremum nor an inflection point.

For example, part of the fold shown in Figure 1(a) is modeled as the figure in Figure 1(b). In the model, the bottom-left point is regarded as the origin and the inclination of the curve is determined to be either increasing or decreasing.

We refer to the borderlines between layers as *layer-borderlines* to discriminate them from the borderline of the rectangle.

Note that since this is a qualitative model, we focus only on the side on which end-points of layer-borderlines occur and the order of the locations, ignoring their precise positions. As for the shape of a layer-borderline, we focus only on its inclination and convexity, ignoring its precise shape. As a result, several figures are regarded as the same model.

Example 1. In Figure 2, (b) is regarded as same as (a), whereas (c) and (d) are not.



(a) (b) (c) (d)

Fig. 2. Qualitative treatment of models [17].

3 Description Language

3.1 Language

We define two kinds of description languages Lang1 and Lang2 to represent the model for local data.

Lang1 is used to describe the configuration of a stratum. This is defined as Lang1 $= \{A_1, \ldots, A_n\} \cup \{\theta\}$ where $A_1, \ldots, A_n$ are the names of the layers and θ denotes the outside of the stratum. $A_1, \ldots, A_n$ and θ are called *layer-symbols*.

Lang2 is used to describe the shape of a layer-borderline. This is defined as Lang2 $=$ {↱, ↴, ⬏, ↳} where ↱, ↴, ⬏ and ↳ indicate convex upward and increasing, convex upward and decreasing, convex downward and increasing, and convex downward and decreasing, respectively. ↱, ↴, ⬏ and ↳ are called *shape-symbols*. We define the sets $Up =$ {↱, ⬏} and $Dn =$ {↴, ↳}.

Let $\sigma = e_1 \ldots e_k$ be either a sequence of symbols in Lang1 or that of those in Lang2. If σ is the null sequence, then we denote it as ϵ. For each i $(1 \leq i \leq k)$, we denote $e_i \in \sigma$, and also denote $first(\sigma) = e_1$, $last(\sigma) = e_k$, $tail(\sigma) = e_2 \ldots e_k$ and $\sigma^{-1} = e_k \ldots e_1$. If $k = 1$, $tail(\sigma) = \epsilon$.

Definition 1 (local data description, layer-sequence) *Local data description is defined as a pair (L, C), where L and C are finite sequences that include symbols in* Lang1 *and* Lang2*, respectively. L consists of four segments in the form $(\sigma^1)(\sigma^2)(\sigma^3)(\sigma^4)$ with auxiliary symbols '(' and ')'. The sequence of symbols without the auxiliary symbols '(' and ')' is called* a layer-sequence *of L.*

A layer-sequence is considered as cyclic data, that is, for a layer-sequence $e_1 \dots e_k$, e_k is considered as e_0, and $e_i \dots e_k e_1 \dots e_{i-1}$ are considered to be equivalent for all i $(1 \le i \le k)$.

Definition 2 (sequence-of-transitions) *For a local data description* (L, C), *let* $I = e_1 \dots e_k$ *be a layer-sequence of* L, *where* $k \neq 1$. *Then the sequence* $c_1 \dots c_k$ *where for each* i $(1 \le i \le k)$, $c_i = e_{i-1}/e_i$, $e_i \in \sigma_i$, $\sigma_i \in \{\sigma^1, \sigma^2, \sigma^3, \sigma^4\}$ *is said to be a* sequence-of-transitions *of* L. *And* $chgpt(c_i, \sigma_i)$ *shows that* c_i *appears in* σ_i.

Example 2. For $L = (A\theta)()(ABC)(B)$, the layer-sequence of L is $I = A\theta ABCB$, the sequence-of-transitions of L is $B/A\ A/\theta\ \theta/A\ A/B\ B/C\ C/B$, and $chgpt(A/\theta, \sigma^1)$ and $chgpt(\theta/A, \sigma^3)$ hold.

3.2 Validity

For a local data description, we introduce the term 'inclination of a layer-borderline' that relates L and C.

Definition 3 (inclination of a layer-borderline) *Let* (L, C) *be a local data description where* $L = (\sigma^1)(\sigma^2)(\sigma^3)(\sigma^4)$. *For each pair of layer-symbols* X *and* Y, *for which* $chgpt(X/Y, \sigma)$ *and* $chgpt(Y/X, \sigma')$ *where* $\sigma \neq \sigma'$ *hold,* the inclination of the layer-borderline C_{XY} *is defined depending on the pair of* σ *and* σ' *as follows:*

- *if* (σ, σ') *is either* $(\sigma^1, \sigma^2), (\sigma^2, \sigma^1), (\sigma^3, \sigma^4)$ *or* (σ^4, σ^3), *then* $C_{XY} = dn$
- *if* (σ, σ') *is either* $(\sigma^1, \sigma^4), (\sigma^2, \sigma^3), (\sigma^3, \sigma^2)$ *or* (σ^4, σ^1), *then* $C_{XY} = up$
- *otherwise,* $C_{XY} = any$.

Definition 4 (validity) *If a local data description* (L, C) *satisfies the following conditions, then it is said to be* a valid representation.

Let $L = (\sigma^1)(\sigma^2)(\sigma^3)(\sigma^4)$ *and its layer-sequence* $I = e_1 \dots e_k$.

v1 *For any pair* X *and* Y *of layer-symbols, if* $chgpt(X/Y, \sigma)$ *holds, then the sequence-of-transitions of* L *includes exactly one* Y/X, *and* $chgpt(Y/X, \sigma')$ *where* $\sigma \neq \sigma'$, $\sigma, \sigma' \in \{\sigma^1, \sigma^2, \sigma^3, \sigma^4\}$ *holds.*

v2 I *is* $X_n\theta$ *or in the form of* $X_1 \dots X_{n-1} X_n X_{n-1} \dots X_1 \theta$ *where* $X_i \neq X_j$ $(1 \le i < j \le n)$.

v3 $|C| = 1$.

v4
- *If for all* C_{XY}, $C_{XY} = up$ or any, *then* $C \in Up$.
- *If for all* C_{XY}, $C_{XY} = dn$ or any, *then* $C \in Dn$.
- *otherwise,* $C \in Up \cup Dn$.

From [v2], the following proposition holds.

Proposition 1 *Let* (L, C) *be a valid representation and* I *be the layer-sequence of* L. *Then* I *is equivalent to* I^{-1}.

4 Representation for a model

We provide the representation for a model. We show that it is valid; and that conversely there exists a model of a valid representation and we can draw a figure satisfying **[COND]**.

4.1 Representation for a model

When a model M of local data is given, starting from the top-left of M, trace the borderline of M in a clockwise manner to obtain a sequence of the layer-symbols that are encountered, and place parentheses around each side of the rectangle. Then we set $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$, where $\sigma_t, \sigma_r, \sigma_b$ and σ_l are the sequence of upper side, right side, lower side and left side of the rectangle, respectively. We set C to correspond to the shape of the layer-borderline. (Note that the shape of all the layer-borderlines is the same.) Then (L, C) is said to be *the representation for M*.

Example 3. The representation for the model shown in Figure 3 is ($(A\theta)()(ABC)(B)$, ↴). The sequence of layer-symbols starts not from layer-symbol B, but from A, although this may seem unnatural. If the sequence were to start from B, the layer-symbol occupying the top-left corner would appear in both σ_t and σ_l. To avoid such a situation and to treat the sequence cyclically, the sequence starts from A, the layer-symbol that is encountered first on tracing.

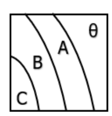


Fig. 3. Representation for a model [17].

Let (L, C) be the representation for a model M. The sequence-of-transitions $c_1 \dots c_k$ of L shows the order of occurrence of the end-points of each layer-borderline on tracing the borderline of M. And for each i $(1 \le i \le k)$, $chgpt(c_i, \sigma_i)$ indicates that the end-point c_i of a layer-borderline is on the side corresponding to σ_i.

4.2 Validity and drawability

Theorem 1 (validity of the model) *The representation for a model is valid.*

Proof. For any pair X and Y of layer-symbols in L, $chgpt(X/Y, \sigma)$ and $chgpt(Y/X, \sigma')$ show that the two end-points of the layer-borderline of X and Y are in σ and σ', respectively. From the first condition of **[COND]**, each layer-borderline of

M does not intersect with itself or another layer-borderline. It has exactly two end-points on the borderlines, which are not on the same side of M, in accordance with the third condition of **[COND]**. Therefore, $\sigma \neq \sigma'$. Thus, validity [v1] holds.

The length of each layer-sequence is even, since each layer-borderline has exactly two end-points. Let $e_1 \ldots e_{2k}$ be the layer-sequence of L. If there is only one layer-borderline, then the layer-sequence of L is $X\theta$ where X is a layer-symbol. If there is more than one layer-borderline, then let $e_0 = e_{2k} = \theta$, $e_1 = X_1, \ldots, e_k = X_k$, where $X_i \in \mathrm{Lang1}$ $(1 \leq i \leq k)$. For each i, j $(0 \leq i < j \leq k-1)$, if the end-points X_i/X_{i+1} and X_j/X_{j+1} occur in this order in the sequence-of-transitions of L, then X_{j+1}/X_j and X_{i+1}/X_i occur in this order in the sequence-of-transitions of L, since layer-borderline pairs should not intersect. Moreover, if we assume that $X_i = X_j$ $(i \neq j)$ holds, then the layer X_i should appear more than twice in L, indicating that it is a disconnected region; this contradicts the first condition of **[COND]**. Therefore, $X_i \neq X_j$. Thus, validity [v2] holds.

Validity [v3] holds from the assumption of the model. Therefore, C_{XY} are defined uniquely and consistently for all pairs of X and Y. Thus, validity [v4] holds.

Theorem 2 (drawability of the representation) *There exists the model of a valid representation.*

Proof. Let (L, C) be a valid representation and $L = (\sigma^1)(\sigma^2)(\sigma^3)(\sigma^4)$.

Let $c_1 \ldots c_{2k}$ be a sequence-of-transitions of L, since the lengths of sequence-of-transitions of L are even from validity [v2]. We locate each $c_i \in \sigma_i$ $(\sigma_i \in \{\sigma^1, \sigma^2, \sigma^3, \sigma^4\})$ $(1 \leq i \leq 2k)$ on the borderline of the rectangle in the clockwise direction: locate the elements σ^1, σ^2, σ^3 and σ^4 on the upper side, right side, lower side and left side, respectively, in accordance with the order of occurrence in the sequence-of-transitions. Then, we can draw each layer-borderline so that its two end-points are not on the same side, and not on a corner, for the following reason.

For any pair X and Y of layer-symbols, we can draw a line between the end-points corresponding to X/Y and Y/X in the sequence-of-transitions. Validity [v1] indicates that a line connecting the two points exists; and validity [v2] indicates that lines do not intersect, and have no extremum or inflection point. Therefore, a region encircled by layer-borderlines and the borderlines of the rectangle is a connected region.

The inclination of all layer-borderlines is the same, which can be determined from validities [v3] and [v4]. Therefore, we can draw a smooth curve according to C.

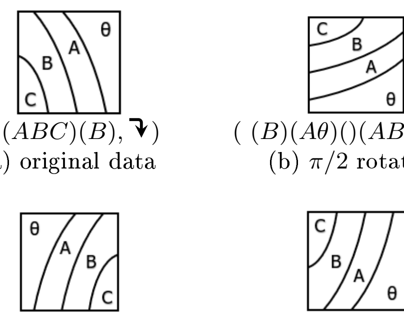
Therefore, the model for a valid representation exists, which means that we can draw a figure corresponding to the model.

5 Operation

Our goal is to derive spatial relations among multiple local data collected in different locations or at different times. To achieve this, we define operations on the local data description and check for changes in the model resulting from these operations.

Let S_0 be a set of representations for models of local data. From Theorem 1, any element of S_0 is valid.

Here, we define three operations: rotation, horizontal flip and vertical flip on S_0 (Figure 4). For $D = (L, C) \in S_0$, we define the operation o on $D = (L, C)$ as $o(D) = (o(L), o(C))$.



($(A\theta)()(ABC)(B)$, ↴) (a) original data

($(B)(A\theta)()(ABC)$, ⬏) (b) $\pi/2$ rotation

($(AB)(C)(BA\theta)()$, ↱) (c) flip horizontal

($(BA\theta)()(AB)(C)$, ⬏) (d) flip vertical

Fig. 4. Operations on S_0 [17].

5.1 $\pi/2$ rotation

Let r be the operation that rotates the model by $\pi/2$ clockwisely. This is defined as follows.

For $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$, $r(L) = (\sigma_l)(\sigma_t)(\sigma_r)(\sigma_b)$.

For C, $r($↱$) =$ ↴, $r($↴$) =$ ⬏, $r($⬏$) =$ ↳, $r($↳$) =$ ↱.

Example 4. The representation for the model in Figure 4(a) is $D =$ ($(A\theta)()(ABC)(B)$, ↴). If we draw $r(D) =$ ($(B)(A\theta)()(ABC)$, ⬏), then we can obtain the model shown in Figure 4(b), which corresponds to $\pi/2$ clockwisely rotated with respect to the original model shown in Figure 4(a).

Proposition 2 *1. The model corresponding to $r(D)$ is a figure that is $\pi/2$ clockwisely rotated relative to that corresponding to D.*

2. For each D in S_0, $r(D)$ is valid.

3. $r(r(r(r(D)))) = D$.

Proof. This can easily be proved, since the operation is only shifting segments.

5.2 Horizontal flip

Let h be the operation that flips the model horizontally.

First, we detect the layer which will occupy the top-left corner of the model after applying the operation. This is said to be *a delimiter* and is defined as follows.

For $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$,

$$delimiter(L) = \begin{cases} last(\sigma_t) \text{ (if } \sigma_t \neq \epsilon) \\ last(\sigma_l) \text{ (if } \sigma_t = \epsilon, \sigma_l \neq \epsilon) \\ last(\sigma_b) \text{ (if } \sigma_t = \sigma_l = \epsilon). \end{cases}$$

Let $I = e_1 \dots e_k$ be the layer-sequence of L and e_z be its delimiter ($1 \leq z \leq k$). Let $I' = e_{z-1}e_{z-2} \dots e_1 e_k e_{k-1} \dots e_z$.

Then we set $h(L) = (\sigma'_t)(\sigma'_r)(\sigma'_b)(\sigma'_l)$, by dividing I' into four segments by inserting the symbols '(' and ')' so that $|\sigma'_t| = |\sigma_t|$, $|\sigma'_r| = |\sigma_l|$, $|\sigma'_b| = |\sigma_b|$ and $|\sigma'_l| = |\sigma_r|$.

For C, $h($↱$) =$ ↴, $h($↴$) =$ ↱, $h($⮥$) =$ ↳, $h($↳$) =$ ⮥.

Example 5. The representation for the model in Figure 4(a) is $D = (L, C) =$ ($(A\theta)()(ABC)(B)$, ↴). $I = A\theta ABCB$, and $delimiter(L) = \theta$, since $\sigma_t \neq \epsilon$. Therefore, $I' = ABCBA\theta$. The numbers of elements on each segment are $|\sigma'_t| = |\sigma_t| = 2$, $|\sigma'_r| = |\sigma_l| = 1$, $|\sigma'_b| = |\sigma_b| = 3$ and $|\sigma'_l| = |\sigma_r| = 0$. Therefore, by cutting I' by these numbers, we get $h(L) = (AB)(C)(BA\theta)()$. In addition, $h(C) =$ ↱. As a result, we obtain $h(D) =$ ($(AB)(C)(BA\theta)()$, ↱), shown in Figure 4(c), which corresponds to the horizontally flipped original model shown in Figure 4(a).

Proposition 3 *1. The model corresponding to $h(D)$ is a figure that is horizontally flipped relative to that corresponding to D.*
2. For each D in S_0, $h(D)$ is valid.
3. $h(h(D)) = D$.

Proof. 1. Considering cyclicity, I' is equivalent to I^{-1}. The encountered order of layers on tracing the borderline of the model for $h(L)$ is the inverse of that in the original model. Moreover, the numbers of end-points on each side of the original model are the same as those on the corresponding sides of the model for $h(L)$, since $|\sigma|$ indicates the number of end-points on the side σ.

2. Assume that L is valid.
I' is equivalent to I^{-1}. In addition, for any pair X and Y of layer-symbols $chgpt(X/Y, \sigma)$ and $chgpt(Y/X, \sigma')$ are mapped to $chgpt(Y/X, \tau)$ and $chgpt(X/Y, \tau')$, respectively, by h. Then $\tau \neq \tau'$ holds since $\sigma \neq \sigma'$ holds, from the definition of h. Therefore, validity [v1] holds.
Validity [v2] holds, since I' is equivalent to I^{-1}.
Validity [v3] trivially holds.
We show that validity [v4] holds as follows. We show the case of $C \in Dn$. Since the inclinations of all the layer-borderlines are either dn or any, we consider a case in which $chgpt(X/Y, \sigma_b)$ and $chgpt(Y/X, \sigma_l)$ hold where the

inclination is $C_{XY} = dn$. In this case, the pair is mapped to $chgpt(Y/X, \sigma'_b)$ and $chgpt(X/Y, \sigma'_r)$, by h. Their inclination is up. Similarly, for the other layer-borderlines, the inclination of dn is mapped to up, and any to any. Therefore, $h(C) \in Up$ holds. It follows that validity [v4] holds in this case. We can prove the other cases similarly.

3. Let I, I' and I'' be layer-sequence of $L, h(L)$ and $h(h(L))$, respectively. And let e_z be a delimiter of L $(1 \leq z \leq k)$.
   $I = e_1 e_2 \dots e_{z-1} e_z e_{z+1} \dots e_{k-1} e_k$
   $I' = e_{z-1} \dots e_2 e_1 e_k e_{k-1} \dots e_{z+1} e_z$
   (a) When $delimiter(L) = last(\sigma_t)$, $delimiter(h(L)) = last(\sigma'_t) = e_k$, since $|\sigma_t| = |\sigma'_t|$. Therefore,
   $I'' = e_1 e_2 \dots e_{z-1} e_z e_{z+1} \dots e_{k-1} e_k$.
   Therefore, $h(h(L)) = L$.
   (b) When $delimiter(L) = last(\sigma_l)$, $last(\sigma_l) = last(I) = e_k$. Then the layer-sequences are as follows.
   $I = e_1 e_2 \dots e_{k-1} e_k$
   $I' = e_{k-1} \dots e_2 e_1 e_k$
   If $\sigma'_l \neq \epsilon$, $delimiter(h(L)) = last(\sigma'_l)$ and $last(\sigma'_l) = last(I') = e_k$, since $\sigma'_t = \epsilon$. If $\sigma'_l = \epsilon$, $delimiter(h(L)) = last(\sigma'_b)$ since $\sigma'_t = \epsilon$, and $last(\sigma'_b) = e_k$. Therefore, $delimiter(h(L)) = e_k$. Therefore,
   $I'' = e_1 e_2 \dots e_{k-1} e_k$
   Therefore, $h(h(L)) = L$.
   (c) When $delimiter(L) = last(\sigma_b), last(\sigma_b) = last(I) = e_k$. Then $delimter(L) = delimiter(h(L)) = e_k$ holds by the same discussion with that in the above case.
   Therefore, $h(h(L)) = L$.

   As for C, $h(h(C)) = C$ holds trivially from the definition of h.
   Thus, we $h(h(D)) = D$ holds.

5.3 Vertical flip

Let v be the operation that flips the model vertically. In this case, the delimiter is defined as follows.

For $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$,

$$delimiter(L) = \begin{cases} last(\sigma_b) \ (\text{if } \sigma_b \neq \epsilon) \\ last(\sigma_r) \ (\text{if } \sigma_b = \epsilon, \sigma_r \neq \epsilon) \\ last(\sigma_t) \ (\text{if } \sigma_b = \sigma_r = \epsilon). \end{cases}$$

Let $I = e_1 \dots e_k$ be the layer-sequence of L and e_z be its delimiter $(1 \leq z \leq k)$. Let $I' = e_{z-1} e_{z-2} \dots e_1 e_k e_{k-1} \dots e_z$.

Then we set $v(L) = (\sigma'_t)(\sigma'_r)(\sigma'_b)(\sigma'_l)$, by dividing I' into four segments by inserting the symbols '(' and ')' so that $|\sigma'_t| = |\sigma_b|$, $|\sigma'_r| = |\sigma_r|$, $|\sigma'_b| = |\sigma_t|$ and $|\sigma'_l| = |\sigma_l|$.

For C, $v($↱$) =$ ↳, $v($⤵$) =$ ⤴, $v($⤴$) =$ ⤵, $v($↳$) =$ ↱.

Example 6. The representation for the model in Figure 4(a) is $D = (\,(A\theta)()(ABC)(B),$ ↴$)$. If we draw $v(D) = (\,(BA\theta)()(AB)(C),$ ⮥$)$, then we can obtain the model shown in Figure 4(d) which corresponds to the vertically flipped original model shown in Figure 4(a).

Proposition 4 *1. The model corresponding to $v(D)$ is a figure that is vertically flipped relative to that corresponding to D.*
2. For each D in S_0, $v(D)$ is valid.
3. $v(v(D)) = D$.

Proof. Similar to the proof regarding h.

5.4 Combination of operations

Proposition 5 *Let S_0 be a set of representations for models of local data. For $D_1, D_2 \in S_0$ where $D_1 = (L_1, C_2)$ and $D_2 = (L_2, C_2)$, if D_2 can be obtained from D_1 by applying the operations r, h and v finite times, then the layer-sequences of L_1 and L_2 are equivalent.*

Proof. Let $L_1 = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$. Then layer-sequence of L_1 is $I = \sigma_t\sigma_r\sigma_b\sigma_l$. The layer-sequence of $r(L_1) = (\sigma_l)(\sigma_t)(\sigma_r)(\sigma_b)$ is $\sigma_l\sigma_t\sigma_r\sigma_b$, which is equivalent to I because of its cyclicity. The layer-sequences of $h(L_1)$ and $v(L_1)$ are equivalent to $I' = e_{z-1}e_{z-2}\ldots e_1e_ke_{k-1}\ldots e_z$, where e_z is the delimiter of L_1. Therefore, they are equivalent to I. Thus,the layer-sequences of L_1 and L_2 are equivalent.

The following property holds with respect to the combination of the operations, which can be proved similarly as the proof of Proposition 3.

Proposition 6 $r(r(D)) = h(v(D)) = v(h(D))$.

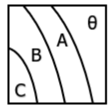
6 Reasoning Based on Connections of Models

6.1 Interconnection of models

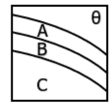
For a pair of representations for models, if the adjacency between the layers appearing in them is the same, then the configuration of the stratum is the same.

Definition 5 (same configuration) *For a pair of representations for models $D_1 = (L_1, C_1)$ and $D_2 = (L_2, C_2)$, let I_1 and I_2 be layer-sequences of L_1 and L_2, respectively. If I_1 is equivalent to I_2, then it is said that D_1 and D_2 have* the same configuration.

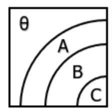
Example 7. In Figure 5, (a), (b) and (c) have the same configuration, whereas (d) does not.



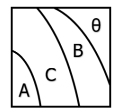
$(A\theta)()(ABC)(B)$
(a)



$()(ABC)()(BA\theta)$
(b)



$()(ABC)(BA\theta)()$
(c)



$(B\theta)(B)(CA)(C)$
(d)

Fig. 5. Same/different configuration [17].

Let S_0 be a set of representations for models of local data. When $D_1, D_2 \in S_0$ have the same configuration, we make a new model D by connecting them horizontally or vertically.

For a sequence of layer-symbols $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$, we denote $set(L)$ as a set of the layer-symbols appearing in L, that is, $set(L) = \{e|e \in \sigma_t\} \cup \{e|e \in \sigma_r\} \cup \{e|e \in \sigma_b\} \cup \{e|e \in \sigma_l\}$.

For a sequence of shape-symbols C, $cend(C)$ shows the symbol to be connected, that is either $last(C)$ or $first(C)$ depending on the locations of the end-points. If $|C| = 1$, then $cend(C) = C$.

left\right	↱	⬏	⬎	↳
↱	↱	↱⬏	↱⬎	ng
⬏	⬏↱	⬏	ng	ng
⬎	ng	ng	⬎	⬎↳
↳	ng	↳⬏	↳⬎	↳

(a) horizontal connection

top\bottom	↱	⬏	⬎	↳
↱	↱	^{T}(↱⬏)	ng	^{T}(↱↳)
⬏	^{T}(⬏↱)	⬏	ng	ng
⬎	ng	^{T}(⬎⬏)	⬎	^{T}(⬎↳)
↳	ng	ng	^{T}(↳⬎)	↳

(b) vertical connection

Table 1. Connection of shape symbols.

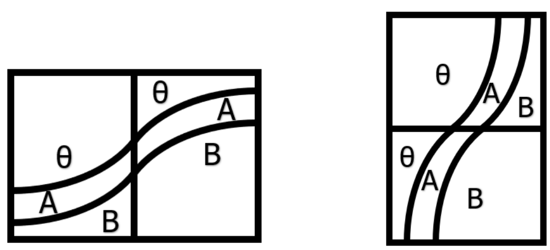
Table 1(a) shows a horizontal connection and Table 1(b) shows a vertical connection of a pair of shape-symbols. In these tables, 'ng' indicates that the connection cannot generate a smooth curve. $^T(C_1C_2)$ indicates that the curve is obtained by applying a vertical connection of C_1 and C_2. The symbol 'T' means the transpose of the sequence. If $|C| = 1$, $^TC = C$.

For example, in Table 1(a), the cell in the first row, first column shows that the result of connecting ↱ and ↱ is a single ↱; the cell in the first row, third column shows that the result of connecting ↱ on the left side and ⬎on the right side is the lined-up of these symbols, which indicates a maximal point. Similarly, the cell in the last row, second column shows that the result of connecting ↳ in the left side and ⬏ on the right side is the lined-up of these symbols, which indicates a minimal point. On the other hand, in Table 1(b), ^{T}(↱↳) and ^{T}(⤵⬏) are the extrema, since the curve is traced from top to bottom.

There are two issues to note.

First issue is the order of tracing the layer-borderline.

Compare Table 1(a) and (b). Focus on the the second row, first column of these tables, which shows a connection of ⬏ and ↱. The shapes of these connections are different, although both of the results are ⬏↱. Figure 6(a) shows the model obtained by the horizontal connection, whereas Figure 6(b) shows the one obtained by the vertical connection. We use the symbol 'T' to discriminate these shapes.



⬏↱ (a) horizontal connection

^{T}(⬏↱) (b) vertical connection

Fig. 6. Different shapes of layer-borderlines with the same sequence.

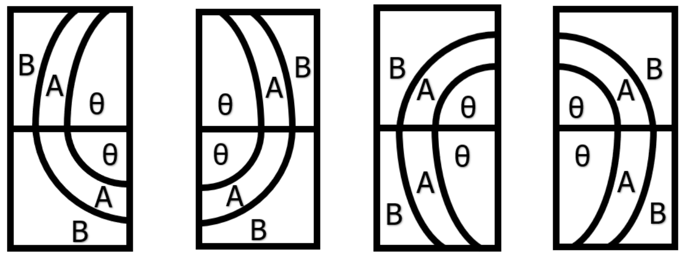
Second issue is the nondeterminacy of the shapes. We cannot determine the shape of the composed layer-borderline only from the sequence of shape-symbols.

Consider ^{T}(↱↳) in the first row, fourth column in Table 1(b). There are two models for the representation ^{T}(↱↳) shown in Figure 7(a) and (c). Similarly, there are two models for the representation ^{T}(⤵⬏) shown in Figure 7(b) and (d). As we cannot determine the shape of the composed layer-borderline only from the sequence of shape-symbols, we identify it by checking the sequence of the layer-symbols.

6.2 Horizontal connection of a pair of local data

First, we show the horizontal connection of a pair of local data.

Let S_0 be a set of representations for models of local data. For a pair of $D_1 = (L_1, C_1)$ and $D_2 = (L_2, C_2)$ in S_0 that have the same configuration, we can connect the right side of D_1 to the left side of D_2 which is denoted by $D_1 || D_2$, if the following two conditions are satisfied.



(a) D_1 (b) D_2 (c) D'_1 (d) D'_2

Fig. 7. Two kinds of shapes of curves for the same sequence of shape-symbols.

all layers occurrence Let $L_1 = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$ and $L_2 = (\tau_t)(\tau_r)(\tau_b)(\tau_l)$. Take elements c_i of the sequence-of-transitions of L_1 that satisfy $chgpt(c_i, \sigma_r)$, and put them in the order of their appearance to make the sequence $c_1 \dots c_k$. Similarly, take elements c'_j of the sequence-of-transitions of L_2 that satisfy $chgpt(c'_j, \tau_l)$, and put them in their order of appearance to make the sequence $c'_1 \dots c'_{k'}$. Then, (i) if $c_i = e_{i-1}/e_i$ then $c'_i = e_{k+1-i}/e_{k-i}$ for each i $(1 \le i \le k)$, (ii) $k = k' = |L_1|/2 = |L_2|/2$, (iii) $set(L_1) = set(L_2) = \cup_{1 \le i \le k}\{e \mid e \text{ appears in } c_i\} = \cup_{1 \le j \le k'}\{e \mid e \text{ appears in } c'_j\}$.

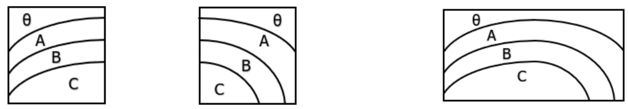
smooth curve Either of the following holds. Note that $|C_1| = |C_2| = 1$.

1. $cend(C_1), cend(C_2) \in Up$
2. $cend(C_1), cend(C_2) \in Dn$
3. $cend(C_1) =$ ↱ and $cend(C_2) =$ ↴
4. $cend(C_1) =$ ↳ and $cend(C_2) =$ ⬏

The condition of all layers occurrence means that all the end-points occurring on two connecting sides coincide, and that all the layer-symbols appear exactly once on both connecting sides. This avoids the case in which the figure corresponding to the resulting representation could contain a disconnected region. We will explain this later in Example 11. The condition of smooth curve means that the shapes of all the connected layer-borderlines are smooth.

If both conditions are satisfied, then $L = (\sigma_t\tau_t)(\tau_r)(\tau_b\sigma_b)(\sigma_l)$. $C = C_1$ if $C_1 = C_2$, and $C = C_1C_2$ otherwise. And we can connect D_1 and D_2 horizontally to generate the representation $D = (L, C)$. In horizontal connection, the shape of the connected layer-borderlines is represented by tracing it from left to right.

Example 8. The representations for the models in Figure 8(a) and (b) are $D_1 =$ ($()(ABC)()(BA\theta)$, ↱) and $D_2 =$ ($()(A)(BC)(BA\theta)$, ↴), respectively. (i) Take the specified elements from the layer-transitions $c_1 = \theta/A, c_2 = A/B, c_3 = B/C$, $c'_1 = C/B, c'_2 = B/A, c'_3 = A/\theta$, (ii) $|L_1|/2 = |L_2|/2 = 3$, (iii) $set(L_1) = set(L_2) = \cup_{1 \le i \le 3}\{e \mid e \text{ appears in } c_i\} = \cup_{1 \le j \le 3}\{e \mid e \text{ appears in } c'_j\} = \{A, B, C, \theta\}$. Therefore, the condition of all layers occurrence is satisfied. Moreover, the third condition of the smooth curve is satisfied. Thus, their horizontal connection $D_1||D_2$ is computed as ($()(A)(BC)(BA\theta)$, ↱↴), which corresponds to the representation for the model in Figure 8(c).



(a) D_1 (b) D_2 (c) $D_1||D_2$

Fig. 8. Horizontal connection of a pair of local data [17].

6.3 Vertical connection of a pair of local data

Next, we show the vertical connection of a pair of local data.

Let S_0 be a set of representations for models of local data. For a pair of $D_1 = (L_1, C_1)$ and $D_2 = (L_2, C_2)$ in S_0 that have the same configuration, we can connect the lower side of D_1 to the upper side of D_2 which is denoted by $D_1 + D_2$, if the following two conditions are satisfied.

all layers occurrence Let $L_1 = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$ and $L_2 = (\tau_t)(\tau_r)(\tau_b)(\tau_l)$. Take elements c_i of the sequence-of-transitions of L_1 that satisfy $chgpt(c_i, \sigma_b)$, and put them in the order of their appearance to make the sequence $c_1 \dots c_k$. Similarly, take elements c'_j of the sequence-of-transitions of L_2 that satisfy $chgpt(c'_j, \tau_t)$, and put them in their order of appearance to make the sequence $c'_1 \dots c'_{k'}$. Then, the same conditions with the (i)$\sim$(iii) stated in the horizontal connection should be satisfied.

smooth curve Either of the following holds. Note that $|C_1| = |C_2| = 1$.

1. $cend(C_1), cend(C_2) \in Up$
2. $cend(C_1), cend(C_2) \in Dn$
3. $cend(C_1) =$ ↱ and $cend(C_2) =$ ↳
4. $cend(C_1) =$ ↴ and $cend(C_2) =$ ⬏

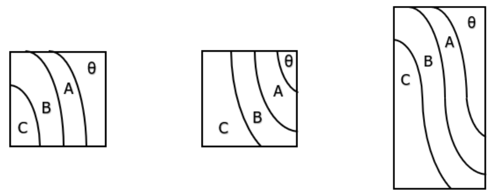
If both conditions are satisfied, then $L = (\sigma_t)(\sigma_r\tau_r)(\tau_b)(\tau_l\sigma_l)$. $C = C_1$ if $C_1 = C_2$, and $C =^T (C_1C_2)$ otherwise. And we can connect D_1 and D_2 vertically to generate the representation $D = (L, C)$. In vertical connection, the shape of the connected layer-borderlines is represented by tracing from top to bottom.

Example 9. The representations for the models in Figure 9(a) and (b) are $D_1 =$ ($(A\theta)()(ABC)(B)$, ↴) and $D_2 =$ ($(BA\theta)(AB)(C)()$, ↳), respectively, and their vertical connection can be computed as $D_1 + D_2 =$ ($(A\theta)(AB)(C)(B)$,^{T}(↴↳)) which corresponds to the representation for the model of Figure 9(c).

In the subsequent two subsections, we discuss a composed connections of multiple local data.

6.4 General horizontal connection

Consider general horizontal connection.



(a) D_1 (b) D_2 (c) $D_1 + D_2$

Fig. 9. Vertical connection of a pair of local data [17].

Let $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$. One end-points of all layer-borderlines should be located in σ_r or σ_l. When we connect (L, C) horizontally, the connecting shape-symbol *cend* can be determined only from C. However, when we connect $(L,{}^T C)$ horizontally, we have to determine *cend* by checking the location of the other end-points. We refer the value of *cend* as *fol*, which stands for 'first or last', in case of connecting $(L,{}^T C)$, in the following explanation. The value of $fol(C)$ is $first(C)$ or $last(C)$ depending on the case: if at least one of the other end-points is located in σ_t, then $cend(C) = last(C)$; if at least one of the other end-points is located in σ_b, then $cend(C) = first(C)$; otherwise, it is not always determined whether $cend(C) = last(C)$ or $cend(C) = first(C)$, and we can execute for each possibility.

Let D_1 and D_2 be the data obtained by horizontal/vertical connection of local data. We can connect the right side of D_1 to the left side of D_2.

- If $D_1 = (L_1, C_1)$ and $D_2 = (L_2, C_2)$, then $cend(C_1) = last(C_1)$, $cend(C_2) = first(C_2)$.
- If $D_1 = (L_1, C_1)$ and $D_2 = (L_2,{}^T C_2)$, then $cend(C_1) = last(C_1)$, $cend(C_2) = fol(C_2)$.
- If $D_1 = (L_1,{}^T C_1)$ and $D_2 = (L_2, C_2)$, then $cend(C_1) = fol(C_1)$, $cend(C_2) = first(C_2)$.
- If $D_1 = (L_1,{}^T C_1)$ and $D_2 = (L_2,{}^T C_2)$, then $cend(C_1) = fol(C_1)$, $cend(C_2) = fol(C_2)$.

If we apply only horizontal connections repetitively, then only the case of $cend(C_1) = last(C_1)$, $cend(C_2) = first(C_2)$ appears.

Let $L_1 = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$ and $L_2 = (\tau_t)(\tau_r)(\tau_b)(\tau_l)$. If the same conditions as in the case of horizontal connection of a pair of local data are satisfied, then we can connect D_1 and D_2 horizontally to get $D = (L, C)$ where $L = (\sigma_t \tau_t)(\tau_r)(\tau_b \sigma_b)(\sigma_l)$, and C is defined as follows.

- In case of $cend(C_1) = last(C_1)$ and $cend(C_2) = first(C_2)$
  - if $cend(C_1) = cend(C_2)$, then $C = C_1 tail(C_2)$
  - otherwise, $C = C_1 C_2$
- In case of $cend(C_1) = last(C_1)$ and $cend(C_2) = last(C_2)$
  - if $cend(C_1) = cend(C_2)$, then $C = C_1 tail(C_2^{-1})$
  - otherwise, $C = C_1 C_2^{-1}$

- In case of $cend(C_1) = first(C_1)$ and $cend(C_2) = first(C_2)$
  - if $cend(C_1) = cend(C_2)$, then $C = C_1^{-1} tail(C_2)$
  - otherwise, $C = C_1^{-1} C_2$
- In case of $cend(C_1) = first(C_1)$ and $cend(C_2) = last(C_2)$
  - if $cend(C_1) = cend(C_2)$, then $C = C_1^{-1} tail(C_2^{-1})$
  - otherwise, $C = C_1^{-1} C_2^{-1}$

Example 10. Consider the horizontal connection of $D_1 = (L_1, {}^T C_1) = (\ (A\theta)(AB)()(), {}^T(↱↳)\ )$ and $D_2 = (L_2, {}^T C_2) = (\ (AB)()()(A\theta), {}^T(⤵⤴)\ )$ shown in Figure 7(a) and (b), respectively. Let $L_1 = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$. As $cend(C_1) = fol(C_1)$, we check the location of the end-points other than $\sigma_r = AB$. Then, as $\sigma_t = A\theta \neq \epsilon$, there exists an end-point on σ_t. Therefore, $cend(C_1) = last(C_1)$. Similarly, $cend(C_2) = last(C_2)$. Thus, the conditions of all layers occurrence and smooth curve are satisfied, and we obtain $C = C_1 C_2^{-1} =$ ↱↳$($⤵⤴$)^{-1} =$ ↱↳⤴⤵. Therefore, $D_1 || D_2 = (\ (A\theta AB)()()(),$ ↱↳⤴⤵$)$.

On the other hand, consider the horizontal connection of $D_1' = (L_1, {}^T C_1) = (\ ()(A\theta)(AB)(), {}^T(↱↳)\ )$ and $D_2' = (L_2, {}^T C_2) = (\ ()()(A\theta)(AB), {}^T(⤵⤴)\ )$ shown in Figure 7(c) and (d), respectively. As $cend(C_1) = first(C_1)$ and $cend(C_2) = first(C_2)$, we obtain $C = C_1^{-1} C_2 = ($↱↳$)^{-1}$⤵⤴ $=$ ↳↱⤵⤴. $D_1' || D_2' = (\ ()()(A\theta AB)(),$ ↳↱⤵⤴$)$.

6.5 General vertical connection

Next, consider general vertical connection.

Let $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$. One end-points of all layer-borderlines should be located in σ_t or σ_b. When we connect $(L, {}^T C)$ vertically, the connecting shape-symbol $cend$ can be determined only from C. However, when we connect (L, C) vertically, we have to determine $cend$ by checking the location of the other end-points. We refer the value of $cend$ as fol, which stands for 'first or last', in case of connecting (L, C), in the following explanation. The value of $fol(C)$ is $first(C)$ or $last(C)$ depending on the case: if at least one of the other end-points is located in σ_r, then $cend(C) = first(C)$; if at least one of the other end-points is located in σ_l, then $cend(C) = last(C)$; otherwise, it is not always determined whether $cend(C) = last(C)$ or $cend(C) = first(C)$, and we can execute for each possibility.

Let D_1 and D_2 be the data obtained by horizontal/vertical connection of local data. We can connect the lower side of D_1 to the upper side of D_2.

- If $D_1 = (L_1, {}^T C_1)$ and $D_2 = (L_2, {}^T C_2)$, then $cend(C_1) = last(C_1)$, $cend(C_2) = first(C_2)$.
- If $D_1 = (L_1, {}^T C_1)$ and $D_2 = (L_2, C_2)$, then $cend(C_1) = last(C_1)$, $cend(C_2) = fol(C_2)$.
- If $D_1 = (L_1, C_1)$ and $D_2 = (L_2, {}^T C_2)$, then $cend(C_1) = fol(C_1)$, $cend(C_2) = first(C_2)$.
- If $D_1 = (L_1, C_1)$ and $D_2 = (L_2, C_2)$, then $cend(C_1) = fol(C_1)$, $cend(C_2) = fol(C_2)$.

If we apply only vertical connections repetitively, then only the case of $cend(C_1) = last(C_1)$, $cend(C_2) = first(C_2)$ appears.

Let $L_1 = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$ and $L_2 = (\tau_t)(\tau_r)(\tau_b)(\tau_l)$. If the same conditions as in the case of vertical connection of a pair of local data are satisfied, then we can connect D_1 and D_2 vertically to get $D = (L,^T C)$, where $L = (\sigma_t)(\sigma_r\tau_r)(\tau_b)(\tau_l\sigma_l)$, and C is defined as same as in the general horizontal connection.

6.6 Succeeding connections

Let S_0 be a set of representations for models of local data. We define S_1 as a union of the set of representations for the horizontally/vertically connected models and S_0.

$$\begin{aligned} S_1 = &\{ D \mid D = D_1||D_2, D_1, D_2 \in S_0 \} \\ &\cup \{ D \mid D = D_1 + D_2, D_1, D_2 \in S_0 \} \\ &\cup S_0. \end{aligned}$$

We repeat this process by generating S_n from S_{n-1} for $n > 1$.

$$\begin{aligned} S_n = &\{ D \mid D = D_1||D_2, D_1, D_2 \in S_{n-1} \} \\ &\cup \{ D \mid D = D_1 + D_2, D_1, D_2 \in S_{n-1} \} \\ &\cup S_{n-1}. \end{aligned}$$

In general, $D \in S_n$ is not always a valid representation, since a layer-borderline of D may include an extremum or an inflection.

Instead of validity, the following properties hold.

Theorem 3 (property of extended representation) $D = (L, C) \in S_n$ *satisfies the following properties.*

Let $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$, *the layer-sequence of* L, $I = e_1 \ldots e_k$ *and* $C = q_1 \ldots q_m$.

p1 *For any pair* X *and* Y *of layer-symbols, if* $chgpt(X/Y, \sigma)$ *holds, then the sequence-of-transitions of* L *includes exactly one* Y/X, *and* $chgpt(Y/X, \sigma')$ *where* $\sigma, \sigma' \in \{\sigma_t, \sigma_r, \sigma_b, \sigma_l\}$ *holds.*

p2 I *is* $X_n\theta$ *or in the form* $X_1 \ldots X_{n-1}X_nX_{n-1} \ldots X_1\theta$, *where* $X_i \neq X_j$ *(*$1 \leq i < j \leq n$*).*

p3 *For any* i *(*$1 \leq i \leq m-1$*),* $q_i \neq q_{i+1}$ *holds.*

We can apply a combination of horizontal/vertical connection to obtain the sequence $S_0, S_1, \ldots, S_n$. However, we have to choose the order of application because of the conditions of the connection.

Example 11. In Figure 10, D_1 and D_2 cannot be horizontally connected, since they do not satisfy the first condition of horizontal connection. If they were horizontally connected, disconnected regions would appear. On the other hand, $D_3 + D_1$ and $D_4 + D_2$ can be generated since the pair D_3 and D_1, and the pair D_4 and D_2 satisfy the conditions of vertical connection, respectively. In addition, $(D_3 + D_1)||(D_4 + D_2)$ can be generated since the pair $D_3 + D_1$ and $D_4 + D_2$ satisfy the conditions of horizontal connection.

As a result, a representation for the global data that may contain a maximal point is obtained.

This example indicates that the properties stated in Theorem 3 are preserved on applying the connections in an appropriate order.

6.7 Prediction of global data

Actually, there is seldom found a stratum of which the entire shape is completely exposed. When multiple data are collected at distant locations, if they have the same configuration, then we can find several possible ways to connect them, by inserting several local data between them and applying the above reasoning.

Example 12. Assume that D_1 and D_2 are collected at distant locations (Figure 11(a)), does a global stratum exist that contains both of them? One possible solution is shown in Figure 11(b), which shows that there should exist D_3 and D_4 (Figure 11(a)) to connect D_1 and D_2. Such intermediate data are often missing, and we infer the possibility that the global stratum exists, which changes in the long term as a result of crustal movements.

In this example, the right two rectangles are wider than the left ones in Figure 11(b). Actually, the sizes of collected local data are not always same or distances between their locations are different. Here, we focus on the end-points of layer-borderlines and their shapes both of which are treated qualitatively. Therefore, we can make a model for a global data by changing the size or ratio of sides of rectangles of the models for local data.

Example 13. Assume that local data P, Q and R are collected at distant locations shown in Figure 12. Does a global stratum exist that connects p_b and r_b, q_l and r_l, respectively?

In this case, we have a solution $(D_1||D_2) + (D_3||D_4) + (\ (D_5 + D_6)||D_7\ )$ shown in Figure 13(a), which is obtained by the operations $||$ and $+$ repetitively. In this figure, the local data D_5, D_2 and D_4 correspond to P, Q and R, respectively. The representation for this model is ($()(ABC)()(BA\theta),^T$(⤵↱↳⤵⬏↳⤵)). There is another solution $(D_1||D_2) + (\ (D_5 + D_6)\ ||\ ((D_3||D_4) + D_7)\ )$ shown in Figure 13(b) for which the representation is the same. This example shows that global structure can be generated by the reasoning procedure, and that multiple solutions can be found.

7 Related Works

To the best of the authors' knowledge, there has been almost no research on strata that uses symbolic representation and logical reasoning.

To combine AI techniques and structural geology, application of machine learning using big data is one possibility. However, the currently available strata-data archives are quite small and the stored data are not sufficiently categorized. Moreover, in most data archives, figures and landscapes are stored using numerical data.

On the other hand, in the QSR research field, several methods for symbolic treatment of shapes have been proposed. Almost all of them treat spatial data on a two-dimensional plane.

Leyton proposed a grammar that represents changes in the shape of a closed curve, starting from a simple smooth curve. He explained changes in shape based on a force acting from inside or outside the curve [12]. He showed that any shape of a smooth closed curve can be represented using language based on the proposed grammar. Tosue et al. extended the grammar so that it can represent phenomena such as the creation of a tangent point and division of the curve [18]. They applied the method to a process of organogenesis. Galton et al. proposed another grammar that can apply not only to a smooth curve but also to a straight line or a curve with cusps [8]. They showed that objects of various shapes can be symbolically represented by connecting a finite number of primitive segments. They also referred to transformation between representations differing in granularity. Cabedo et al. proposed a representation for the borderline of an object with further information such as relative lengths and relative angles, and also showed the juxtaposition of objects [2, 1]. Falomir et al. extended this representation to develop a new language for a qualitative shape representation of a two-dimensional object. They also considered the connection of the objects and formalized the composition of qualitative lengths, angles and convexities. Falomir et al. also defined similarity between qualitative shapes described in their extended model [6]. They focused only on the shape of the boundary of an object, whereas we treat the inner configuration of an object, which consists of multiple regions in a single object as well as the shape of the regions. They used straight lines to represent a shape, whereas we use curves, which is another difference.

All of these expressions adopted methods that represent the shape of an object by connecting primitive segments when tracing its borderline. On the other hand, Cohn took a different approach to represent a concave object [5]. He regarded differences in the closure and the object itself as regions and represented the spatial relations of these regions. Kumokawa et al. also proposed a different representation for a concave shape using closure [11].

A study by Kulik et al. applied QSR to landscape silhouettes [10]. They proposed a description language for the shape of an open line. They defined several primitives comprising two consecutive vectors depending on relative lengths and angles; regarded the borderline of a silhouette of a landscape as a pattern of connections between these primitives; and deduced landscape features, including mountain, valley, and plateau. They also proposed a transformation from the refined level to the abstract level. The differences between Kulik's method and ours are: first, he used straight lines as primitives, whereas we use curves; second, his target silhouette was always in the vertical direction, whereas our method can be applied to rotated forms; third, he neither formalized the method nor discussed the validity of the representation, whereas we both define the validity of the representation and prove one-to-one relation with the model.

In addition, while all existing studies treated the essentially one-dimensional data of a borderline, we treat the two-dimensional data of a stratum consisting of multiple regions.

8 Conclusions

We have discussed qualitative representation and reasoning for strata.

We provided a model for local data from a typical form of fold strata, and proposed its representation in the form of a pair of sequences of symbols that stand for the configuration of a layer and the shapes of the borderlines between layers. This representation is suitable to show the main features of strata: one layer extends in one direction if there is no fault, and the relations of interconnections between layers are unchanged even if the width of a layer, shape, or axis of a fold changes.

We defined the validity of the representation, and then proved that the representation for a model is valid and that conversely, there exists the model of a valid representation. Moreover, we defined several operations on the representation, and showed that they preserve the validity. We also showed reasoning about the connection of local data. We formalized the connection and showed that global data connecting several local data can be constructed.

This enables derivation of relations among multiple local data collected in different locations or at different times. Our main contribution is to show symbolic treatment of strata and provide a basis for logically explaining the process of landscape generation.

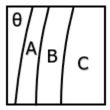
In future studies, we intend to identify sets of representations obtained from repetitive application of connections of local data. We are considering a relaxation of the condition of all layers occurrence for the connection so that more flexible reasoning is available. We are also considering the formalization needed to explain the strata-generation process, as well as a qualitative simulation for possible future morphological changes.

Acknowledgements This research is supported by JSPS Kakenhi JP21K12020. The authors would like to thank Motohiro Tsuboi for giving useful advice from the field of geology.

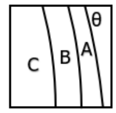
References

1. Cabedo, L.M., Abril, L.G., Morente, F.V., Falomir, Z.: A pragmatic qualitative approach for juxtaposing shapes. Journal of Universal Computer Science **16**(11), 1410–1424 (2010)
2. Cabedo, L.M., Escrig, M.T.: A qualitative theory for shape representation and matching for design. Proceedings of the Sixteenth European Conference on Artificial Intelligence, ECAI'2004 pp. 858–862 (2004)
3. Chen, J., Cohn, A.G., Liu, D., Wang, S., Ouyang, J., Yu, Q.: A survey of qualitative spatial representations. The Knowledge Engineering Review **30**, 106–136 (2013)

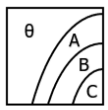
4. Cohn, A., Renz, J.: Qualitative spatial representation and reasoning. In: Handbook of Knowledge Representation, chap. 13. Elsevier (2008)
5. Cohn, A.G.: Hierarchical representation of qualitative shape based on connection and convexity. Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science (COSIT95) pp. 311–326 (1995)
6. Falomir, Z., Abril, L.G., Cabedo, L.M., Ortega, J.A.: Measures of similarity between objects based on qualitative shape descriptions. Spatial Cognition & Computation **13**(3), 181–218 (2013)
7. Falomir, Z., Pich, A., Costa, V.: Spatial reasoning about qualitative shape compositions: Composing qualitative lengths and angles. Annals of Mathematics and Artificial Intelligence **88**, 589-–621 (2020)
8. Galton, A., Meathrel, R.: Qualitative outline theory. Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence pp. 1061–1066 (1999)
9. Kano, K., Murata, A.: Structural Geology. Asakura Publishing Co., Ltd. (in Japanese) (1998)
10. Kulik, L., Egenhofer, M.J.: Linearized terrain: Languages for silhouette representations. Spatial Information Theory. Foundations of Geographic Information Science, International Conference, COSIT 2003 pp. 118–135 (2003)
11. Kumokawa, S., Takahashi, K.: Qualitative spatial representation based on connection patterns and convexity. AAAI08 Workshop on Spatial and Temporal Reasoning pp. 40–47 (2008)
12. Leyton, M.: A process-grammar for shape. Artificial Intelligence **34**, 213–247 (1988)
13. Ligozat, G.: Qualitative Spatial and Temporal Reasoning. Wiley (2011)
14. Nishimura, Y., et al.: Earth Science (high school textbook, in Japanese). Daiichi Gakushusha Co. (2021)
15. Pich, A., Falomir, Z.: Logical composition of qualitative shapes applied to solve spatial reasoning tests. Cognitive Systems Research **52**, 82–102 (2018)
16. Sioutis, M., Wolter, D.: Qualitative spatial and temporal reasoning: current status and future challenges. Proceedings of 30th International Joint Conference on Artificial Intelligence pp. 4594–4601 (2021)
17. Taniuchi, Y., Takahashi, K.: Qualitative spatial representation and reasoning about fold strata. Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART2023), pp. 211–220 (2023)
18. Tosue, M., Takahashi, K.: Towards a qualitative reasoning on shape change and object division. In: Spatial Information Theory. 11th International Conference, COSIT 2019. pp. 7:1–7:15 (2019)



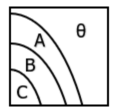
($(ABC)()(BA\theta)()$,↱)
D_1



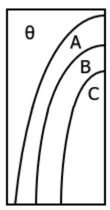
($(BA\theta)()(ABC)()$, ↴)
D_2



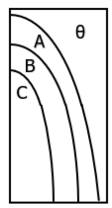
($()(ABC)(BA\theta)()$,↱)
D_3



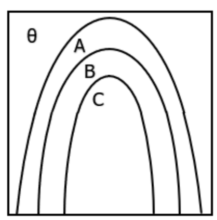
($()()(ABC)(BA\theta)$, ↴)
D_4



($()(ABC)(BA\theta)()$,↱)
$D_3 + D_1$



($()()(ABC)(BA\theta)$, ↴)
$D_4 + D_2$



($()()(ABCBA\theta)()$,↱↴)
$(D_3 + D_1)||(D_4 + D_2)$

Fig. 10. Process of connections [17].

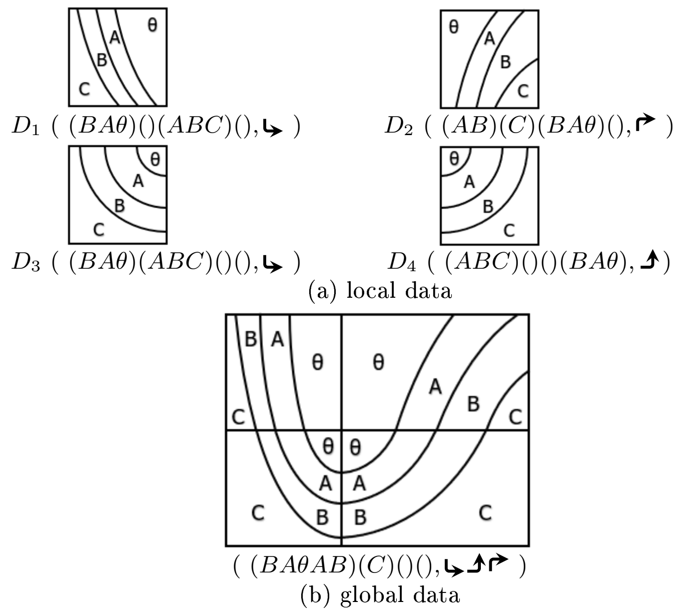


Fig. 11. Prediction of global data [17].

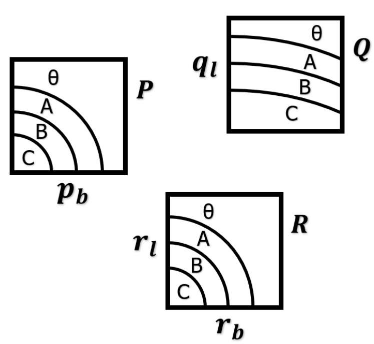
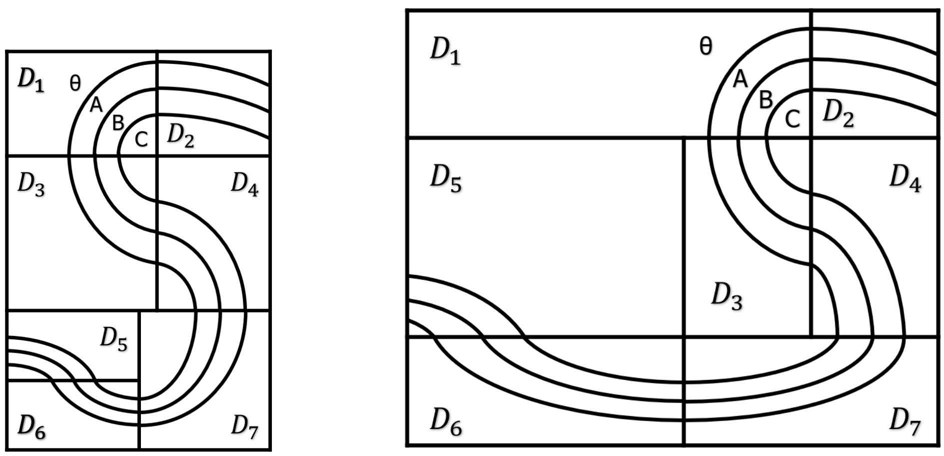


Fig. 12. Collected local data.



(a) solution 1 (b) solution 2

Fig. 13. A global data obtained by repetitive horizontal/vertical connections.