A Spatial Similarity Measure based on Games: Theory and Practice

MARCO AIELLO, ILLC and ISIS, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam. The Netherlands. Email: aiellom@ieee.org

Abstract

Model comparison games can be used not only to decide whether two specific models are equivalent or not, but also to establish a measurement of difference among a whole class of models. We show how this is possible in the case of the spatial modal logic $S_4$ of Bennett. The approach results in a spatial similarity measure based on topological model comparison games. After establishing the theoretical framework, we move towards practice by giving an algorithm to effectively compute the similarity measure for a class of topological models widely used in computer science applications: polygons of the real plane. In the appendix, we briefly overview an implemented system based on the theoretical framework.

Keywords: model comparison games, similarity measures, modal logics of space, image retrieval based on spatial relationships

1 Introduction

There are various ways to take space qualitatively. Topology, orientation or distance have been investigated in a non-quantitative manner. The literature especially is abundant in mereotopological theories, i.e., theories of parthood $P$ and connection $C$. Even though the two primitives can be axiomatized independently, the definition of part in terms of connection suffices for AI applications. Usually, some fragment of topology is axiomatized and set inclusion is used to interpret parthood.\footnote{L. J. of the IGPL, Vol. 10 No. 1, pp. 1–22 2002 © Oxford University Press}

Most of the efforts in mereotopology have gone into the axiomatization of specific theories, disregarding important model theoretic questions. Issues such as model equivalence are seldom (if ever) addressed. Seeing an old friend from high-school yields an immediate comparison with the image one had from the school days. Most often, one immediately notices how many aesthetic features have changed. Recognizing a place as one already visited involves comparing the present sensory input against memories of the past sensory inputs. “Are these trees the same as I saw six hours ago, or are they arranged differently?” An image retrieval system seldom yields an exact match, more often it yields a series of ‘close’ matches. In computer vision, object occlusion cannot be disregarded. One ‘sees’ a number of features of an object and compares them with other sets of features to perform object recognition. Vision is not a matter of precise matching, it is more closely related to similarity. The core of the problem lies in the precise definition of ‘close’ match, thus the question shall be: How similar are two spatial patterns?

The fundamental issues in order to answer this question involve finding an agreement on spatial representation and finding an agreement on a language to describe
spatial patterns. Our choice here falls on modal logics topologically interpreted. The language, called \( S4_u \), is a multi-modal \( S4^*S5 \) logic interpreted on topological spaces equipped with valuation functions. \( S4_u \) is an extension of the simple modal logic \( S4 \) with universal and existential modal operators. Thank to the extension one can get rid of \( S4 \)'s intrinsic ‘locality’, a known technique used in modal logic, \( [21] \). Bennett introduced \( S4_u \) in the spatial setting \( [10] \) to encode decidable fragments of the region connection calculus RCC (a fundamental qualitative spatial reasoning calculus in AI \( [24] \) extending Allen’s ideas \( [5] \) from temporal reasoning to spatial reasoning). The encoding also proved essential to identify maximal tractable fragments of RCC \( [26] \). \( S4_u \) has recently been used in \( [28] \) as a logic complete with respect to connected topological spaces. Finally, in the recent and important trend of combining spatial and temporal formalism, \( S4_u \) plays an important role \( [30] \). Even though the logical technique we deploy is similar to that of \( [10, 26] \), we would like to remark a shift in perspective. First, we consider \( S4_u \) not as a decidable access to RCC but as a general language of (mereo-)topology: a way to ‘logically’ access topological and mereological notions. Second, we stress issues of model equivalence and model comparison, not only spatial representation.

Spatial representation is not only interesting in itself, but also when considering its applications. It is essential in vision, in spatial reasoning for robotics, in geographical information systems, just to name a few. Of paramount importance in applications is the comparison of spatial patterns which must be represented in the same way, in short, similarity measures are of great importance. We consider similarity measures and look at their application to image retrieval. Image retrieval is concerned with the indexing and retrieval of images from a database, according to some desired set of image features. These features can be as diverse as textual annotations, color, texture, object shape, and spatial relationships among objects. The way the features from different images are compared, in order to have a measure of similarity among images, is what really distinguishes an image retrieval architecture from another one. We refer to \( [17] \) for an overview of image retrieval and more specifically to \( [27] \) for image similarity measures. Here we concentrate on image retrieval based on spatial relationships at the qualitative level of mereotopology, that is, part-whole relations, topological relations and topological properties of individual regions (see for instance \( [6] \)). Other image retrieval systems are based on spatial relationships as the main retrieval feature. The work in \( [24] \) is founded on transformation of Voronoi diagrams and that in \( [23] \) on graph matching. An older and known approach to image retrieval by spatial relationship is in \( [14] \). This seminal work considers the projections of regions onto two axes imposed on the picture and simple interval relations over the projections. This approach suffers from not being orientation invariant and from the inability to deal with overlapping objects. On the positive side is the compactness of the topological representation of spatial relationships (called 2D strings).

The organization of the paper reflects the transition from theory to practice we are interested in. We begin by giving the formal (theoretical) details of the modal logic \( S4_u \) in Section 2. To get a feeling for its expressive power, we place it in the taxonomy of mereotopological theories of Cohn-Varzi. In Section 3, we introduce the notion of topological bisimulation for \( S4_u \) and show its adequacy. The main theoretical result of the paper is provided in Section 4 where Ehrenfeucht-Fraïssé style\(^1\) model comparison

\(^1\)For an introduction to Ehrenfeucht-Fraïssé games see, for instance, \( [18] \).
2. A GENERAL FRAMEWORK FOR MEREOTOPOLOGY

games adequate for S4u are presented and it is shown how the games can be turned into a distance measure on the space of all topological models of S4u. In Section 5, we make an ontological commitment and show the distance measure to be decidable by providing an algorithm to compute it. The proofs of the theorems throughout the paper are collected at the end in Appendix A. The techniques described in the paper have been used to implement an image retrieval prototype named IRIS (Image Retrieval based on Spatial relationships) which is rapidly overviewed in Appendix B.

The paper is based and extends [1, 2].

2 A general framework for Mereotopology

The proposed framework takes the beaten road of mereotopology by extending topology with a mereological theory based on the interpretation of set inclusion as part-hood. Hence, a brief recall of the basic topological definitions is in order.

A topological space is a couple $\langle X, O \rangle$, where $X$ is a set and $O \subseteq \mathcal{P}(X)$ such that: $\emptyset \in O$, $X \in O$, $O$ is closed under arbitrary union, $O$ is closed under finite intersection. An element of $O$ is called an open. A subset $A$ of $X$ is called closed if $X - A$ is open. The interior of a set $A \subseteq X$ is the union of all open sets contained in $A$. The closure of a set $A \subseteq X$ is the intersection of all closed sets containing $A$.

To capture a considerable fragment of topological notions a multi-modal language $S4u$ interpreted on topological spaces (à la Tarski [30]) is used. A topological model $M = \langle X, O, \nu \rangle$ is a topological space $\langle X, O \rangle$ equipped with a valuation function $\nu : \mathcal{P} \rightarrow \mathcal{P}(X)$, where $\mathcal{P}$ is the set of proposition letters of the language.

The definition and interpretation of $S4u$ follows that given in [3], which in turn is a rewriting of the one in [10]. In [3] though, emphasis is given to the topological expressivity of the language rather than the mereotopological implications. Every formula of $S4u$ represents a region. Two modalities are available. $\Box \varphi$ to be interpreted as “interior of the region $\varphi$”, and $U \varphi$ to be interpreted as “it is the case everywhere that $\varphi$.” The truth definition can now be given. Consider a topological model $M = \langle X, O, \nu \rangle$ and a point $x \in X$:

- $M, x |= p$ iff $x \in \nu(p)$ (with $p \in \mathcal{P}$)
- $M, x |= \neg \varphi$ iff not $M, x |= \varphi$
- $M, x |= \varphi \rightarrow \psi$ iff not $M, x |= \varphi$ or $M, x |= \psi$
- $M, x |= \Box \varphi$ iff $\exists o \in O : x \in o \land \forall y \in o : M, y |= \varphi$
- $M, x |= \Diamond \varphi$ iff $\forall o \in O : x \notin o \land \exists y \in o : M, y |= \varphi$
- $M, x |= U \varphi$ iff $\forall y \in X : M, y |= \varphi$
- $M, x |= E \varphi$ iff $\exists y \in X, M, y |= \varphi$

Since $\Box$ is interpreted as interior and $\Diamond$ (defined dually as $\Diamond \varphi \leftrightarrow \neg \Box \neg \varphi$, for all $\varphi$) as
closure, it is not a surprise that these modalities obey the following axioms:

\[ \Box A \rightarrow A \quad (T) \]
\[ \Box A \rightarrow \Box \Box A \quad (4) \]
\[ \Box \top \quad (N) \]
\[ \Box A \land \Box B \leftrightarrow \Box (A \land B) \quad (R) \]

(4) is idempotence, while (N) and (R) are immediately identifiable in the definition of topological space. For the universal—existential modalities \( U \) and \( E \) (defined dually: \( E \varphi \leftrightarrow \neg U \neg \varphi \)) the axioms are those of S5:

\[ U(\varphi \rightarrow \psi) \rightarrow (U \varphi \rightarrow U \psi) \quad (K) \]
\[ U \varphi \rightarrow \varphi \quad (T) \]
\[ U \varphi \rightarrow U U \varphi \quad (4) \]
\[ \varphi \rightarrow UE \varphi \quad (B) \]

In addition, the following ‘connecting’ principle is part of the axioms:

\[ \Diamond \varphi \rightarrow E \varphi \quad (Con) \]

The axiomatization of \( \Box \) as interior is due to [30] and is generally known as S4 in modal logics. Though, in the context of Kripke semantics one gives an equivalent set of axioms to the one here provided. The axiomatization of the full \( S4_u \) was first introduced in [21], then by Bennett [10] with the topological interpretation for spatial reasoning.

Before defining the similarity measure based on model comparison games for \( S4_u \), we take a look at the mereotopological expressive power of the language. This to get acquainted with the language and to get an intuition for what \( S4_u \) can and what it cannot express.

### 2.1 Expressivity

The language \( S4_u \) is perfectly suited to express mereotopological concepts. The relation of parthood \( P(A, B) \) of a region \( A \) being inside the region \( B \) holds whenever it is the case everywhere that \( A \) implies \( B \):

\[ P(A, B) := U(A \rightarrow B) \]

This captures exactly the set-inclusion relation of the models. As for connection \( C \), two regions \( A \) and \( B \) are connected if there exists a point where both \( A \) and \( B \) are true:

\[ C(A, B) := E(A \land B) \]

From here it is immediate to define all the usual mereotopological predicates such as proper part, tangential part, overlap, external connection, and so on. Notice that the choice made in defining \( P \) and \( C \) is arbitrary. So, why not take a more restrictive definition of parthood? Say, \( A \) is part of \( B \) whenever the closure of \( A \) is contained in the interior of \( B \)?

\[ P(A, B) := U(\Diamond A \rightarrow \Box B) \]
3. When are two spatial patterns the same?

As this formula shows, $\mathbf{S4}_u$ is expressive enough to capture also this definition of parthood. In [15], the logical space of mereotopological theories is systematized. Based on the intended interpretation of the connection predicate $C$, and the consequent interpretation of $P$ (and fusion operation), a type is assigned to mereotopological theories. More precisely, a type is a triple $\tau = (i, j, k)$, where the first $i$ refers to the adopted definition of $C_i$, $j$ to that of $P_j$ and $k$ to the sort of fusion. The index $i$, referring to the connection predicate $C$, accounts for the different definition of connection at the topological level. Using $\mathbf{S4}_u$ one can repeat here the three types of connection:

$$C_1(A, B) := E(A \land B)$$

$$C_2(A, B) := E(A \land \Diamond B) \lor E(\Diamond A \land B)$$

$$C_3(A, B) := E(\Diamond A \land \Diamond B)$$

Looking at previous mereotopological literature, one remarks that RCC uses a $C_3$ definition, while the system proposed in [6] uses a $C_1$. Similarly to connectedness, one can distinguish the various types of parthood, again in terms of $\mathbf{S4}_u$:

$$P_1(A, B) := U(A \to B)$$

$$P_2(A, B) := U(A \to \Diamond B)$$

$$P_3(A, B) := U(\Diamond A \to \Diamond B)$$

In [15], the definitions of the $C_i$ are given directly in terms of topology, and the definitions of $P_j$ in terms of a first order language with the addition of a predicate $C_i$. Finally, a general fusion $\phi_k$ is defined in terms of a first order language with a $C_1$ predicate. Fusion operations are like algebraic operations on regions, such as adding two regions (product), or subtracting two regions. One cannot repeat the general definition given in [15] at the $\mathbf{S4}_u$ level. Though, one can show that various instances of fusion operations are expressible in $\mathbf{S4}_u$. For example, the product $A \times_k B$:

$$A \times_1 B := A \land B$$

$$A \times_2 B := (\Diamond A \land B) \lor (A \land \Diamond B)$$

$$A \times_3 B := (\Diamond A \land \Diamond B)$$

The above discussion has shown that $\mathbf{S4}_u$ is a general language for mereotopology. All the different types $\tau = (i, j, k)$ of mereotopological theories are expressible within $\mathbf{S4}_u$.

3 When are two spatial patterns the same?

One is now ready to address questions such as: When are two spatial patterns the same? or When is a pattern a sub-pattern of another one? More formally, one wants to define a notion of equivalence adequate for $\mathbf{S4}_u$ and the topological models. In first-order logic the notion of ‘partial isomorphism’ is the building block of model equivalence. Since $\mathbf{S4}_u$ is multi-modal language, one resorts to bisimulation, which is the modal analogue of partial isomorphism. Bisimulations compare models in a structured sense, ‘just enough’ to ensure the truth of the same modal formulas [22].
Definition 3.1 (Topological bisimulation) Given two topological models \( \langle X, O, \nu \rangle \) and \( \langle X', O', \nu' \rangle \), a total topological bisimulation is a non-empty relation \( \equiv \subseteq X \times X' \) defined for all \( x \in X \) and for all \( x' \in X' \) such that if \( x \equiv x' \):

\[
\text{(base): \quad } x \in \nu(p) \text{ iff } x' \in \nu'(p) \text{ (for any proposition } p) \\
\text{(forth condition): \quad } \text{if } x \in o \in O \text{ then } \exists o' \in O' : x' \in o' \text{ and } \forall y' \in o' : \exists y \in o : y \equiv y' \\
\text{(back condition): \quad } \text{if } x' \in o' \in O' \text{ then } \exists o \in O : x \in o \text{ and } \forall y \in o : \exists y' \in o' : y \equiv y' 
\]

If only conditions (i) and (ii) hold, the second model simulates the first one.

The notion of bisimulation is used to answer questions of ‘sameness’ of models, while simulation will serve the purpose of identifying sub-patterns. Though, one must show that the above definition is adequate with respect to the mereotopological framework provided in this paper.

Theorem 3.2 Let \( M = \langle X, O, \nu \rangle \), \( M' = \langle X', O', \nu' \rangle \) be two models, \( x \in X \), and \( x' \in X' \) bisimilar points. Then, for any modal formula \( \varphi \) in \( S4_u \), \( M, x \models \varphi \) iff \( M', x' \models \varphi \).

Theorem 3.3 Let \( M = \langle X, O, \nu \rangle \), \( M' = \langle X', O', \nu' \rangle \) be two models with finite \( O, O' \), \( x \in X \), and \( x' \in X' \) such that for every \( \varphi \) in \( S4_u \), \( M, x \models \varphi \) iff \( M', x' \models \varphi \). Then there exists a total bisimulation between \( M \) and \( M' \) connecting \( x \) and \( x' \).

In words, extended modal formulas are invariant under total bisimulations, while finite modally equivalent models are totally bisimilar. One may notice, that in Theorem 3.3 a finiteness restriction is posed on the open sets. This will not surprise the modal logician, since the same kind of restriction holds for Kripke semantics and does not affect the proposed use for bisimulations in the mereotopological framework.

4 How different are two spatial patterns?

If topological bisimulation is satisfactory from the formal point of view, one needs more to address qualitative spatial reasoning problems and computer vision issues. If two models are not bisimilar, or one does not simulate the other, one must be able to quantify the difference between the two models. Furthermore, this difference should behave in a coherent manner across the class of all models. Informally, one needs to answer questions like: How different are two spatial patterns?

To this end, we recall the game theoretic definition of topo-games, and then prove the main theoretical result of this paper, namely the fact that topo-games induce a distance on the space of all topological models for \( S4_u \). First, we give the definition and the theorem that ties together the topo-games, \( S4_u \) and topological models.

Definition 4.1 (Topo-game) Consider two topological models \( \langle X, O, \nu \rangle \), \( \langle X', O', \nu' \rangle \) and a natural number \( n \). A topo-game of length \( n \), notation \( TG(X, X', n) \), consists of \( n \) rounds between two players, Spoiler and Duplicator, who move alternatively.
Spoiler is granted the first move and always chooses which type of round to engage. The two sorts of rounds are as follows:

\[
\begin{align*}
\text{global} & \quad \begin{align*}
(i) & \text{ Spoiler chooses a model } X_s \text{ and picks a point } \bar{x}_s \text{ anywhere in } X_s \\
(ii) & \text{ Duplicator chooses a point } \bar{x}_d \text{ anywhere in the other model } X_d
\end{align*} \\
\text{local} & \quad \begin{align*}
(i) & \text{ Spoiler chooses a model } X_s \text{ and an open } o_s \\
& \text{ containing the current point } x_s \text{ of that model} \\
(ii) & \text{ Duplicator chooses an open } o_d \text{ in the other model } X_d \text{ containing its current point } x_d \\
(iii) & \text{ Spoiler picks a point } \bar{x}_d \text{ in Duplicator’s open } o_d \text{ in the } X_d \text{ model} \\
(iv) & \text{ Duplicator replies by picking a point } \bar{x}_s \text{ in Spoiler’s open } o_s \text{ in } X_s
\end{align*}
\end{align*}
\]

The points \(\bar{x}_s\) and \(\bar{x}_d\) become the new current points. A game always starts by a global round. By this succession of actions, two sequences are built: \(\{x_1, x_2, \ldots, x_n\}\) and \(\{x'_1, x'_2, \ldots, x'_n\}\). After \(n\) rounds, if \(x_i\) and \(x'_i\) (with \(i \in [1, n]\)) satisfy the same propositional atoms, Duplicator wins, otherwise, Spoiler wins. A winning strategy (w.s.) for Duplicator is a function from any sequence of moves by Spoiler to appropriate responses which always ends in a win for him. Spoiler’s winning strategy is defined dually.

The multi-modal rank of a S4u formula is the maximum number of nested modal operators appearing in it (i.e. \(\Box, \Diamond, U\) and \(E\) modalities). The following adequacy of the games with respect to the mereotopological language holds.

**Theorem 4.2 (Adequacy)** Duplicator has a winning strategy for \(n\) rounds in \(TG(X, X', n)\) iff \(X\) and \(X'\) satisfy the same formulas of multi-modal rank at most \(n\).

Various examples of plays and a discussion of winning strategies can be found in \(\text{[3]}\).

The interesting result is that of having a game theoretic tool to compare topological models. Given any two models, they can be played upon. If Spoiler has a winning strategy in a certain number of rounds, then the two models are different up to a certain degree. The degree is exactly the minimal number of rounds needed by Spoiler to win. On the other hand, one knows (see \(\text{[3]}\)) that if Spoiler has no w.s. in any number of rounds, and therefore Duplicator has in all games, including the infinite round game, then the two models are bisimilar.

A way of comparing any two given models is not of great use by itself. It is essential instead to have some kind of measure. It turns out that topo-games can be used to define a distance measure.

\[\text{Note: For example, one may find interesting that a normal form is available for the language (one for which every formula has one universal modal operator ranging over boolean combinations of local modal operators). The normal form is tied to the winning strategies of either player.}\]
A Spatial Similarity Measure based on Games: Theory and Practice

Definition 4.3 (isosceles topo-distance) Consider the space of all topological models $T$. Spoiler’s shortest possible win is the function $spw : T \times T \to \mathbb{N} \cup \{\infty\}$, defined as:

$$spw(X_1, X_2) = \begin{cases} n & \text{if Spoiler has a winning strategy in } TG(X_1, X_2, n), \\ & \text{but not in } TG(X_1, X_2, n - 1) \\ \infty & \text{if Spoiler does not have a winning strategy in } \\ & TG(X_1, X_2, \infty) \end{cases}$$

The isosceles topo-model distance (topo-distance, for short) between $X_1$ and $X_2$ is the function $tmd : T \times T \to [0, 1]$ defined as:

$$tmd(X_1, X_2) = \frac{1}{spw(X_1, X_2)}$$

The distance was named ‘isosceles’ since it satisfies the triangular property in a peculiar manner. Given three models, two of the distances among them (two sides of the triangle) are always the same and the remaining distance (the other side of the triangle) is smaller or equal. On the left of Figure 1, three models are displayed: a spoon, a fork and a plate. Think these cutlery objects as subsets of a dense space, such as the real plane, which evaluate to $\phi$, while the background of the items evaluates to $\neg \phi$. The isosceles topo-distance is displayed on the left next to the arrow connecting two models. For instance, the distance between the fork and the spoon is $\frac{1}{2}$ since the minimum number of rounds that Spoiler needs to win the game is 2.

**Fig. 1:** On the left, three models and their relative distance. On the right, the distinguishing formulas.
To see this, consider the formula $E \Box \phi$, which is true on the spoon (there exists an interior point of the region $\phi$ associated with the spoon) but not on the fork (which has no interior points). On the right of the figure, the formulas used by spoiler to win the three games between the fork, the spoon and the plate are shown. Next the proof that $tmd$ is really a distance, in particular the triangular property, exemplified in Figure $\text{F}$, is always satisfied by any three topological models.

**Theorem 4.4 (isosceles topo-model distance)** $tmd$ is a distance measure on the space of all topological models.

The nature of the isosceles topo-distance triggers a question. Why, given three spatial models, the distance between two couples of them is always the same?

First an example, consider a spoon, a chop-stick and a sculpture by Henry Moore. It is immediate to distinguish the Moore’s sculpture from the spoon and from the chop-stick. The distance between them is high and the same. On the other hand, the spoon and the chop-stick look much more similar, thus, their distance is much smaller. Mereotopologically, it may even be impossible to distinguish them, i.e., the distance may be null.

In fact one is dealing with models of a qualitative spatial reasoning language of mereotopology. Given three models, via the isosceles topo-distance, one can easily distinguish the very different patterns. In some sense they are far apart as if they were belonging to different equivalence classes. Then, to distinguish the remaining two can only be harder, or equivalently, the distance can only be smaller.

## 5 Computing similarities

The fundamental step to move from theory to practice has been taken when shifting from model comparison games to a distance. To complete the journey towards practice one needs to identify ways of effectively compute the distance in cases actually occurring in real life domains. We do not have an answer to the general question of whether the topo-distance is computable for any two topological models or not. Though, by restricting to a specific class of topological models widely used in real life applications, we can show the topo-distance to be computable when one makes an ontological commitment. The commitment consists of considering topological spaces made of polygons. This is common practice in various application domains such as geographical information systems (GIS), in many branches of image retrieval and of computer vision, just to mention the most common.

Consider the real plane $\mathbb{R}^2$, any line in $\mathbb{R}^2$ cuts it into two half-planes. We call a half-plane **closed** if it includes the cutting line, **open** otherwise.

**Definition 5.1 (region)** A **polygon** is the intersection of finitely many open or closed half-planes. An **atomic region** of $\mathbb{R}^2$ is the union of finitely many polygons.

An atomic region is denoted by one propositional letter. More in general, any set of atomic regions, simply called **region**, is denoted by a S4$_u$ formula. The polygons of the plane equipped with a valuation function, denoted by $M_{\mathbb{R}^2}$, are in full rights a topological model as defined in Section 2, a basic topological fact. A similar definition of region can be found in $\mathbb{R}^2$. In that article Pratt and Lemon also provide a collection of fundamental results regarding the plane, polygonal ontology just defined (actually one in which the regions are open regular).
From the model theoretic point of view, the advantage of working with $M_{G_2}$ is that we can prove a logical finiteness result and thus give a terminating algorithm to compute the topo-distance. The preliminary step is thus that of proving a finiteness lemma for $S4_u$ over $M_{G_2}$ models.

**Lemma 5.2 (finiteness)** There are only finitely many modally definable subsets of a finite set of regions $\{r_i | r_i \text{ is an atomic region}\}$.

Here is a proof sketch. We work by enumerating cases, i.e., considering boolean combinations of planes, adding to an ‘empty’ space one half-plane at the time, first to build one region $r$, and then to build a finite set of regions. The goal is to show that only finitely many possibilities exist. We begin by placing a half plane denoted by $r$ on an empty bidimensional space, Figure 2.a. Let us follow what happens to points in the space from left to right. On the left, points satisfy the formula $E(r \wedge \square r)$ and its subformulas $Er$ and $E\square r$. This is true until we reach the frontier point of the half-plane. Either $E(\neg r \wedge \square r)$ or $E(r \wedge \square \neg r)$ are true depending on whether the half-plane is open or closed, respectively. Once the frontier has been passed to the right, the points satisfy $E(\neg r \wedge \square r)$ and its subformulas $E \neg r$ and $E \square \neg r$, better seen in Figure 2.a. In fact, if we consider negation in the formulas the role of $r$ and $\neg r$ switch. Consider now a second plane in the picture:

- **Intersection:** the intersection may be empty (no new formula), may be a polygon with two sides and vertices (no new formula, the same situation as with one polygon), or it may be a line, the case of two closed polygons that share the side (in this last case depicted in Figure 2.b—spike—we have a new formula, namely, $E(r \wedge \square \neg \neg r)$).

---

3. Of course, in general this is not true. There are infinitely many non equivalent $S4_u$ formulas and one can identify appropriate Kripke models to show this. See, e.g., [11].

4. The finiteness lemma is the extension to two dimensions of the theorem for serial sets of [10]. In two-dimensions one has 8 non-equivalent formulas rather than 6, as in the one dimensional case proved in [10].
• **Union:** the union may be a polygon with either one or two sides (no new formula), two separated polygons (no new formula), or two open polygons sharing the open side (this last case depicted in Figure 5c). In such a case, one inverts the roles of \( r \) and \( \neg r \) in the formula: \( E(\neg r \land \Box \neg r) \).

Finally, consider combining cases (a) and (b). By union, we get Figure 5d. The only situation bringing new formulas is the latter. In particular, the point where the line intersects the plane satisfies the formula: \( E(\Diamond \Box r \land \Diamond (r \land \Box \neg r)) \). By intersection, we get no segment, or the empty space, thus, no new formula.

The four basic configurations just identified yield no new configuration from the S4n point of view. To see this, consider the boolean combinations of the above configurations. We begin by negation (complement):

\[
\begin{array}{c|c|c|c|c}
\neg & a & b & c & d \\
\hline
\neg a & \neg b & \neg a, \neg b & \neg a, \neg b, \neg d \\
\end{array}
\]

Union straightforwardly follows (where a stands for both a and \( \neg a \), as both configurations always appear together):

\[
\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c}
\cup & a & b & c & d \\
\hline
a & a, \neg b, \neg d & a, c, d & a, \neg b, c, d, \neg d & a, \neg b, d, \neg d \\
b & a, c, d & b & c, d & d \\
c & a, \neg b, c, d, \neg d & c, d & a, \neg b, c, d, \neg d & a, \neg b, c, d, \neg d \\
d & a, \neg b, d, \neg d & d & a, \neg b, c, d, \neg d & a, \neg b, d, \neg d \\
\end{array}
\]

The table for intersection follows, with the proviso that the combination of the two regions can always be empty (not reported in the table) and again a and \( \neg a \) are represented simply by a:

\[
\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c}
\cap & a & b & c & d \\
\hline
a & a, b, c, d & b & a, b, c & a, b, d \\
b & b & b & b & b \\
c & a, b, c & b & a, b, c, d & a, b, c, d \\
d & a, b, d & b & a, b, c, d & a, b, c, d \\
\end{array}
\]

We call topo-vector associated with the region \( r \), notation \( \vec{r} \), an ordered sequence of ten boolean values. The values represent whether the region \( r \) satisfies or not the ten formulas

\[
\{ E_r, E_{\neg r}, E_{\Box r}, E_{\Box \neg r}, E(\neg r \land \Box \neg r), E(r \land \Box \neg r), E(\neg r \land \Box \neg r), E(\Diamond \Box r \land \Diamond (r \land \Box \neg r)), E(\Diamond \Box r \land \Diamond (r \land \Box \neg r)) \}.
\]
The ten formulas are those identified in Figure 2, which we have shown to be the only one definable by boolean combinations of planes denoting the same one region \( r \). For example, the topo-vector associated with a plate—a closed square \( r \) in the plane—is \{true, true, true, true, false, true, false, false, false, false\}. Adding half-planes with different denotations \( r_2, r_3, \ldots \) increases the number of defined formulas. The definition of topo-vector is extended to an entire \( \mathbb{M} \mathbb{R}^2 \) model:

\[
\{ E \bigwedge_i [\neg] r_i, \ E \bigwedge_i \Box [\neg] r_i, \ E (\bigwedge_i [\neg]^+ r_i \land \bigwedge_i \Diamond [\neg]^* r_i), \ E (\bigwedge_i [\neg]^+ r_i \land \bigwedge_i \Box [\neg]^* r_i), \\
E (\bigwedge_i \Diamond [\neg]^+ r_i \land \Diamond (\bigwedge_i [\neg]^+ r_i \land \bigwedge_i \Box [\neg]^* r_i)) \},
\]

where \( [\ ] \) denotes an option and if the option \( [\ ]^+ \) is used then the option \( [\ ]^* \) is not and vice-versa. The topo-vector is built such that the modal rank of the formulas is not decreasing going from the positions with lower index to those with higher. The size of such a vector is \( 5 \cdot 2^i \) where \( i \) is the number of denoted regions of the model. The fact that the size of the topo-vector grows exponentially with the number of regions might seem a serious drawback. Though, as we shall show in a moment, the topo-vector stores all the information relevant for \( \text{S}_4 \mathbb{u} \) about the model. Furthermore, the size of a topo-vector is most often considerably smaller than that of a topological model. In fact, a topo-vector is of exponential size in the number of regions, while a topological model is of exponential size in the number of points of the space because of the set of opens. As a final argument, one should add that in practical situations the number of regions is always much smaller than the number of points of the space.

We are now in a position to devise an algorithm to compute the topo-distance between two topological models. The algorithm works by first computing the associated topo-vectors and then comparing them. By the comparison it is possible to establish which formulas differentiate the two models and therefore the distance between the two models. Here is the general algorithm (in pseudo-code) to compute the topo-distance between two topological models \( M_1 \) and \( M_2 \):

\[
\text{topo-distance}(M_1, M_2) \\
\quad v_1^* = \text{topo-vector}(M_1) \\
\quad v_2^* = \text{topo-vector}(M_2) \\
\quad \text{align } v_1^* \text{ and } v_2^* \\
\quad \text{loop on } v_1^*, v_2^* \text{ with index } i \\
\quad \quad \text{if } v_1^*(i) \neq v_2^*(i) \\
\quad \quad \quad \text{return } \text{modal rank}(v_1^*(i)) \\
\quad \quad \text{return } 0
\]

The idea is of retrieving the topo-vectors associated with the two input models and then looping over their elements. The inequality check can also be thought of as a xor, since the elements of the array are booleans. If the condition is never satisfied, the two topo-vectors are identical, the two-models are topo-bisimilar and thus the topo-distance is null. The \text{align} command makes the topo-vectors of the same length and aligns the formulas of the two, i.e., such that to the same index in the vector
5. Computing similarities

Fig. 3. Computing the topo-vector on a simple model.

corresponds the same formula. If a topo-vector contains a formula that the other one does not, the entry is added to the vector missing it with a false value. To complete the description of the algorithm, we provide the function to compute the topo-vector associated with an $M_{i\mathcal{E}}$ model:

```
topo-vector(M)
  \vec{v} = initialized to all false values
  loop on regions $r$ of $M$ with index $i$
    loop on atomic regions $a$ of $r(i)$ with index $j$
      loop on vertices $v$ of $a(j)$ with index $k$
        update $\vec{v}$ with the point $v(k)$
        if $v(k)$ is not free
          loop on intersections $x$ of $a(j)$ with all regions of $M$ with index $l$
            update $\vec{v}$ with the point $x(l)$
  return $\vec{v}$$
```

If a point $v(k)$ of an atomic region $a(j)$ is contained in any polygon different from $a(j)$ and it is not contained in any other region, then the condition $v(k)$ is not free is satisfied. Standard computational geometry algorithms exist for this task, [16]. When the “update $\vec{v}$ with the point $p$” function is called, one checks in which case $p$ is (as shown after Lemma 5.2), then one considers the position of the corresponding topo-vector and puts in a true value. An obvious optimization to the algorithm is to avoid checking points for which all the associated formulas are already true. Consider the simple model of Figure 3 composed of two closed regions $r$ and $q$. Since there are two regions, the topo-vector will be of size $5 \cdot 2^2 = 20$ elements: $\{E(r \land q), E(r \land \neg q), E(\neg r \land \neg q \land \Box r \land \Box q)\}$. After initialization, the region $r$ is considered and one starts looping on the vertices of its polygons, first the point 1. The point is free, it is the vertex of a full polygon (not a segment) and therefore the topo-vector is updated directly in the positions corresponding to $Er \land \neg q$, $E\Box r \land \Box \neg q$, $Er \land \neg q \land \Box r \land \Box \neg q$, $Er \land \neg q \land \Diamond \Box r \land \Diamond \Box q$. The points 2 and 3 would update the values for the same formula and are not considered. The point 4 falls inside the first
polygon of \( r \), the topo-vector does not need update. Intersections are then computed and the point 5 is found. The point needs to update the vector for the formula \( E \Diamond \Box r \land \Diamond \Box \neg q \land \Diamond (r \land \neg q \land \Box \Diamond \neg r \land \Box \Diamond \neg q) \). Finally, the point 6 is considered and the point needs to update the vector for the formula \( E(r \land \neg q \land \Diamond \Box \neg r \land \Diamond \Box \neg q) \). The algorithm proceeds by considering the second region, \( q \) and its vertices 7, 8, and 9. The three vertices all fall inside the region \( r \) and provide for the satisfaction of the formulas \( Er \land q, E \Box r \land \Box q, \ldots \).

**Lemma 5.3 (termination)** The topo-distance algorithm terminates.

The property is easily shown by noticing that a segment (a side of a polygon) can have at most one intersection with any other segment, that the number of polygons forming a region of \( M_{\mathbb{R}^2} \) is finite, and that the number of regions of \( M_{\mathbb{R}^2} \) is finite. Putting this result together with Lemma 5.2 one gets the hoped decidablility result for polygonal topological models.

**Theorem 5.4 (decidability of the topo-distance)** In the case of polygonal topological models \( M_{\mathbb{R}^2} \) over the real plane, the problem of computing the topo-distance among any two models is decidable.

Given the definition of topo-distance, the fact that two models have a null topo-distance implies that in the topo-game Duplicator has a winning strategy in the infinite round game. In the case of \( M_{\mathbb{R}^2} \), Theorem 5.4 implies that the two models are topo-bisimilar. Note that, in general, this is not the case: Duplicator may have a winning strategy in an infinite model comparison game adequate for some modal language and the models need not be bisimilar [9].

**Corollary 5.5 (decidability of topo-bisimulations)** In the case of polygonal topological models over the real plane, the problem of identifying whether two models are topo-bisimilar or not is decidable.

6 Conclusions

We have followed the line from theory to practice in a context of spatial reasoning. First, we have considered a general mereotopological framework, placing the language \( S_4_u \) where it belongs: \( S_4_u \) is a general mereotopological language not committed to any specific definition of connection, but rather with high topological discriminating power. We addressed issues of model equivalence and especially of model comparison, thus, looking at mereotopology from a new angle. Defining a distance that encodes the mereotopological difference between spatial models has important theoretical and application implications, as we have shown. Our journey has ended by illustrating the actual decidability of the devised similarity measure for a practically interesting class of models.

The theoretical framework proposed is much more general than what we have shown here. We were interested in a mereotopological framework and have therefore used the language \( S_4_u \) interpreted on topological models, but an isosceles distance can be used for any modal language equipped with negation and for which one has adequate notions of model comparison games and bisimulation. Even the restriction to modal logic is not necessary, one can think of first-order logic, of the usual Ehrenfeucht-Fraïssé games, of elementary equivalence in place of bisimulation, and an isosceles
distance is then definable. The decidability result for the distance is the only thing that does not necessarily extend, rather one has to consider the class of models and the logic case by case. Of particular importance is then how the adequate topological games are defined.

We would like to stress the fact that the use of model comparison games presented in this paper is novel. Model comparison games have been used only to compare two given models, but the issue of setting a distance among a whole class of models has not been previously addressed. The technique employed in Theorem 4.4 for the language $S_4 u$ is, as we have just mentioned, much more general. A question interesting *per se*, but out of the scope of the present paper, is: which is the class of games (over which languages) for which a notion of isosceles distance holds? We believe the class of such languages and model comparison games to be quite vast.

Having implemented a system based on the above framework is also an important step in the presented research. Experimentation is essential to assess applicability, but some preliminary considerations are possible. We have noticed that the prototype is very sensible to the labeling of segmented areas of images, i.e., to the assignment of propositional letters to regions. We have also noticed that the mereotopological expressive power appears to enhance the quality of retrieval and indexing over pure textual searches, but the expressive power of $S_4 u$ is still too limited. Notions of qualitative orientation, shape or geometry appear to be important, especially when the user expresses his desires in the form of an image query or of a sketch.

The generality of the framework described in the paper allows for optimism about future developments. Once one has identified an appropriate language of, say, qualitative shape with adequate model comparison games, a newer version of IRIS can be built. We will be researching in two directions. On the one hand, qualitative notions of shape are expressible via mathematical morphology, which in turn is closely related to modal logics $[2]$. On the other hand, axiomatizations of the notions of betweenness (also originated by Tarski in $[31]$) may provide for qualitative notions of geometry. Again, one can stay on the ground of modal logics and one can look at languages for incidence geometries. In the approach, one distinguishes the sorts of elements that populate space and considers the incidence relation between elements of the different sorts (see $[8, 7, 33]$).

All in all, there is much more to model-comparison games than simply laying down two peculiar models and start playing on them. We have looked at spatial reasoning and at image similarity, but many more roads are viable.

Aknowledgements

The author is thankful to Johan van Benthem for fruitful discussions on the topic, to Kees Doets for feedback on a previous version of the proof of Theorem 4.4, and to Michael Zakharyaschev for helpful discussion. This work was supported in part by CNR grants 203.15.10 and 203.7.27.

References

A Spatial Similarity Measure based on Games: Theory and Practice


A. **Proofs of Theorems**

Proof of Theorem 3.2 on page 16.

Induction on \( \varphi \). The case of a proposition letter \( p \) is the first condition on \( \models \). As for conjunction, \( M, x \models \varphi \land \psi \) is equivalent by the truth definition to \( M, x \models \varphi \) and \( M, x \models \psi \), which by the induction hypothesis is equivalent to \( M', x' \models \varphi \) and \( M', x' \models \psi \), which by the truth definition amounts to \( M', x' \models \varphi \land \psi \). The other boolean cases are similar. For the modal case, we do one direction. First the ‘local’ modal operators \( \Box \) and \( \Diamond \). If \( M, x \models \Box \varphi \), then by the truth definition we have that \( \exists o \in O : x \in o \land \forall y \in o : M, y \models \varphi \). By the forth condition, corresponding to \( o \), there must exist an \( o' \in O' \) such that \( \forall y' \in o' \exists y \in o : y \equiv y' \). By the induction hypothesis applied to \( y \) and \( y' \) with respect to \( \varphi \), then \( \forall y' \in o' : M', y' \models \varphi \). By the truth definition of the modal operator we have \( M', x' \models \Box \varphi \). Using the back condition one proves the other direction likewise. Now the ‘global’ modal operators \( U \) and \( E \): If \( M, x \models E \varphi \), then by the truth definition we have that \( \exists y : M, y \models \varphi \). The point \( x \) must be in the open set \( X \) (the whole space). By the forth condition, we know that there must be a corresponding open set \( o' \) in \( X' \). Though, we need a little bit more, therefore we show next that \( X' \) itself is one of such \( o' \).

**Fact A.1** If the topological models \( \langle X, O, \nu \rangle \) and \( \langle X', O', \nu' \rangle \) are totally topologically bisimilar, a special instance of the forth condition holds (similarly for the back condition):

\[
\text{if } x \in X \in O \text{ then } \exists X' \in O' : x' \in X' \text{ and } \forall y' \in X' : \exists y \in X : y \equiv y'
\]

In general, if we instantiate the forth condition on the whole space \( X \), we do not know which open set \( o' \) will correspond on \( \langle X', O', \nu' \rangle \):

\[
\text{if } x \in X \in O \text{ then } \exists o' \in O' : x' \in o' \text{ and } \forall y' \in o' : \exists y \in X : y \equiv y'
\]

Obviously, \( o' \subseteq X' \). Ab absurdum, \( o' \neq X' \) for all \( o' \) as just defined by the forth condition. Thus, there must exist an element \( y' \) of \( X' \) that belongs to none of the sets \( o' \). Two cases are now possible. Either there exists a \( p \in X \) such that \( p \equiv y' \) or it does not exist. In both cases one reaches a contradiction. If it exists, then \( X' \) is an \( o' \) open set. If it does not, then the two models are not totally topologically bisimilar.

We can now proceed in the original proof. By the forth condition and the above fact, corresponding to \( X \) there is the open whole space \( X' \) such that \( \forall z \in X' \exists z' \in X \equiv z' \). By the induction hypothesis applied to \( z \) and \( z' \) with respect to \( \varphi \), \( \forall z' \in X' : M', z' \models \varphi \). One of such \( z \) is the \( y \) of the truth definition for \( E \varphi \), therefore \( M', y \models \varphi \). By the truth definition, we have \( M', x' \models E \varphi \). Using the back condition one proves the other direction likewise.


Proof of Theorem 3.3 on page 15

To get a bisimulation between the two finite models, we stipulate that \( u \equiv u' \) if and only if \( u \) and \( u' \) satisfy the same modal formulas. The atomic preservation condition for a bisimulation holds since the modal \( \varphi \) include all proposition letters. We now prove the forth condition. Suppose that \( u \equiv u' \) where \( u \in o \). We must find an open \( o' \) such that \( u' \in o' \) and \( \forall y \in o' \exists y \in o : y \equiv y' \). Now, suppose there is no such \( o' \). Then for every \( o' \) containing \( x' \exists y' \in o' : \forall y \in o : \exists \varphi_y : y \neq \varphi_y \) and \( y' \models \varphi_y \). In words, every open \( o' \) contains a point \( y' \) with no modally equivalent point in \( o \). Taking the finite conjunction of all formulas \( \varphi_y \), we get a formula \( \Phi_{\varphi} \) such that \( y' \models \Phi_{\varphi} \) and \( \neg \Phi_{\varphi} \) is true everywhere in \( o \). Slightly abusing notation, we write \( o \models \neg \Phi_{\varphi} \). This line of reasoning holds for any open \( o' \) containing \( x' \) as chosen. Therefore, there exists a collection of formulas \( \neg \Phi_{\varphi} \) for which \( o \models \neg \bigwedge_{o'} \Phi_{\varphi} \).

Since \( x \in o \), by the truth definition we have \( x \models \Box_{o'} \neg \Phi_{\varphi} \). By the fact that \( x \) and \( x' \) satisfy the same modal formulas, it follows that \( x' \models \Box_{o'} \neg \Phi_{\varphi} \). But then, there exists an open \( o' \) (with \( x' \in o' \)) such that \( o' \models \bigwedge_{o'} \neg \Phi_{\varphi} \). Since \( o' \) is an open containing \( x' \), it is one of the \( o' \), i.e. \( o' \models \neg \bigwedge_{o'} \Phi_{\varphi} \). But we had supposed that for all opens \( o' \) there was a point \( y' \models \Phi_{\varphi} \), so in particular the \( y' \) of \( o' \) satisfies \( \Phi_{\varphi} \). We have thus reached a contradiction: which shows that some appropriate open \( o' \) must exist. The back clause is proved analogously.

Proof of Theorem 4.2 on page 15

The left to right direction is proven by induction on the length \( n \) of the game \( TG(X, X', n) \). If \( n = 0 \) and Duplicator has a winning strategy, this means that \( X \) and \( X' \) satisfy the same propositional letters, hence the same boolean combinations of propositional letters, i.e., the same modal formulas of modal rank 0. Now for the inductive step. Suppose that Duplicator has a winning strategy \( \sigma \) in \( TG(X, X', n) \). We want to show that \( X, x \models \varphi \) iff \( X', x' \models \varphi \) when the modal rank of \( \varphi \) is \( k \).

By simple syntactic inspection, \( \varphi \) must be a boolean combination of formulas of the form \( \Box_{o'} \psi \) or \( U_{o'} \psi \) where \( \psi \) has modal rank less or equal to \( n - 1 \). Thus, it suffices to prove that \( X \models \Box_{o'} \psi \) iff \( X' \models \Box_{o'} \psi \) and that \( X \models U_{o'} \psi \) iff \( X' \models U_{o'} \psi \). Without loss of generality, let us consider the first model. Suppose that \( X \models \Box_{o'} \psi \). By the truth definition there exists an open \( o \) (with \( x \in o \)) such that \( \forall \psi \in o : X, x \models \psi \). Now, assume that the \( n \)-round game starts with Spoiler choosing \( o \) in \( X \).

Using the strategy \( \sigma \), Duplicator can pick an open \( o' \) such that \( x' \in o' \) and \( \forall \psi \in o' : X, x' \models \psi \). Now Spoiler can pick any point \( u \) in \( o' \). Duplicator can use the information in \( \sigma \) to respond with a point \( u \in oBox \), concluding the first round, so that the remaining strategy \( \sigma' \) is still winning for Duplicator in \( TG(X, X', n - 1) \). By the inductive hypothesis, the fact that \( X, u \models \psi \) (where \( \psi \) has modal rank \( n - 1 \)) implies that \( X' \models \psi \). Thus we have shown that all \( u' \in o' \) satisfy \( \psi \), and hence \( X' \models \psi \). Therefore, the direction is analogous. Suppose now that \( X \models U_{o'} \psi \). By the truth definition, for all \( x \in X \) such that \( X, x \models \psi \). Ab absurdum, \( X - \models U_{o'} \psi \), hence \( X \models E \neg \psi \). By the truth definition, \( \exists x' \models X' \) such that \( X', x' \models \neg \psi \). Spoiler can choose the \( x' \) point as his first move. Duplicator’s choice on \( X \) is necessarily a point \( x \) such that \( X, x \models \psi \), hence Duplicator cannot win the game \( TG(X, X', n - 1) \), contradicting the induction hypothesis. The other direction is analogous.

The right to left direction is again proven by induction on \( n \). If \( n = 0 \), then \( X \) and \( X' \) satisfy the same non-modal formulas. In particular, they satisfy the same atoms, which is winning for Duplicator, by the definition of topological game. For the inductive step, without loss of generality, let us assume that in the first (global) round of \( TG(X, X', n) \) Spoiler chooses the point \( x \). Consider a generic open \( o \) containing \( x \). Now, take the set \( \{ DES_{n-1}(z) : z \in o \} \), where \( DES_{n-1}(z) \) denotes all the formulas up to modal rank \( n - 1 \) satisfied at \( z \). This set is not finite per se, but we can simply prove the following.

Fact A.2 (Logical Finiteness) There are only finitely many formulas of modal depth \( k \) up to logical equivalence.

Therefore, we can write one boolean formula to describe this open set \( o \), namely \( \bigvee \bigwedge DES_{n-1}(z) \). Since this is true for all \( z \in o \), by the truth definition that \( X, x \models \Box \bigvee \bigwedge DES_{n-1}(z) \) (a formula of modal rank \( n \)). By hypothesis, \( x \) and \( x' \) satisfy the same modal formulas of modal rank \( n \), so \( X', x' \models \Box \bigvee \bigwedge DES_{n-1}(z) \). This last fact, together with the truth definition implies that there
exists an open \( o' \) such that \( \forall z' \in o' : X', z' \models \bigvee \bigwedge \text{DES}_{n-1}(z) \). This is the open that Duplicator must choose to reply to Spoiler’s move. Now Spoiler can pick any point \( u' \) in \( o' \). Such a point satisfies at least one disjunct \( \bigwedge \text{DES}_{n-1}(z) \), and we let Duplicator respond with \( z \in o \). As a result of this first round, \( z, u' \) satisfy the same modal formulas. Thus, one correctly gets of this first round, \( z, u' \) satisfy the same modal formulas. Thus, one correctly gets

Proof of Theorem 4.4 on page 19

\( tmd \) satisfies the three properties of distances; i.e., for all \( X_1, X_2 \in T \):

(i) \( tmd(X_1, X_2) \geq 0 \) and \( tmd(X_1, X_2) = 0 \iff X_1 = X_2 \)

(ii) \( tmd(X_1, X_2) = tmd(X_2, X_1) \)

(iii) \( tmd(X_1, X_2) + tmd(X_2, X_3) \geq tmd(X_1, X_3) \)

As for (i), from the definition of topo-games it follows that the amount of rounds that can be played is a positive quantity. Furthermore, the interpretation of \( X_1 = X_2 \) is that the spaces \( X_1, X_2 \) satisfy the same modal formulas. If Spoiler does not have a w.s. in \( \lim_{n \to \infty} TG(X_1, X_2, n) \) then \( X_1, X_2 \) satisfy the same modal formulas. Thus, one correctly gets

\[
tmd(X_1, X_2) = \lim_{n \to \infty} \frac{1}{n} = 0.
\]

Equation (ii) is immediate by noting that, for all \( X_1, X_2, TG(X_1, X_2, n) = TG(X_2, X_1, n) \).

As for (iii), the triangular property, consider any three models \( X_1, X_2, X_3 \) and the three games playable on them,

\[
TG(X_1, X_2, n), TG(X_2, X_3, n), TG(X_1, X_3, n)
\] (A.1)

Two cases are possible. Either Spoiler does not have a winning strategy in all three games (A.1) for any amount of rounds, or he has a winning strategy in at least one of them.

If Spoiler does not have a winning strategy in all the games (A.1) for any number of rounds \( n \), then Duplicator has a winning strategy in all games (A.1). Therefore, the three models satisfy the same modal formulas, \( spw \to \infty \), and \( tmd \to 0 \). Trivially, the triangular property (iii) is satisfied.

Suppose Spoiler has a winning strategy in one of the games (A.1). Via Theorem 4.2 (adequacy), one can shift the reasoning from games to formulas: there exists a modal formula \( \gamma \) of multi-modal rank \( m \) such that \( X_i \models \gamma \) and \( X_j \models \neg \gamma \). Without loss of generality, one can think of \( \gamma \) as being in normal form:

\[
\gamma = \bigvee \bigwedge \neg U(\varphi_{S4})
\] (A.2)

This last step is granted by the fact that every formula \( \varphi \) of \( S4_k \) has an equivalent one in normal form whose modal rank is equivalent or smaller to that of \( \varphi \). Let \( \gamma^* \) be the formula with minimal multi-modal depth \( m^* \) with the property: \( X_i \models \gamma^* \) and \( X_j \models \neg \gamma^* \). Now, the other model \( X_k \) either satisfies \( \gamma^* \) or its negation. Without loss of generality, \( X_k \models \gamma^* \) and therefore \( X_j \) and \( X_k \) are distinguished by a formula of depth \( m^* \). Suppose \( X_j \) and \( X_k \) to be distinguished by a formula \( \beta \) of multi-modal depth \( h < m^* \). \( X_j \models \beta \) and \( X_k \models \neg \beta \). By the minimality of \( m^* \), one has that \( X_i \models \beta \), and hence, \( X_j \) and \( X_k \) can be distinguished at depth \( h \). As this argument is symmetric, it shows that either

- one model is at distance \( \frac{1}{m^*} \) from the other two models, which are at distance \( \frac{1}{h} \) (\( \leq \frac{1}{m^*} \)), or
- one model is at distance \( \frac{1}{h} \) from the other two models, which are at distance \( \frac{1}{m^*} \) (\( \leq \frac{1}{h} \)) one from the other.

It is a simple matter of algebraic manipulation to check that \( m^*, l \) and \( h, m^* \) (as in the two cases above), always satisfy the triangular inequality.

\[\text{In the proof, the availability of the normal form is not strictly necessary, but it gives a better impression of the behavior of the language.}\]
The IRIS prototype

The ultimate step toward practice of the spatial framework presented in the paper is the actual implementation of the similarity measure in a prototype. The topo-distance is a building block of an image retrieval system, named IRIS Image Retrieval based on Spatial relationships, coded in Java and enjoying a Swing interface (Figure 5).

The actual similarity measure is built in IRIS to both index and retrieve images on the basis of:
(i) The spatial intricacy of each region,
(ii) The binary spatial relationships between regions, and
(iii) The textual description accompanying the image.

Referring to Figure 4, one can get a glimpse of the conceptual organization of IRIS. A spatial model, as defined in Section 4, and a textual description (central portion of the figure) are associated with each image of the collection (on the left). Each topological model is represented by its topo-distance vector, as built by the algorithm in Section 5, and by a matrix of binary relationships holding between regions. Similarly, each textual description is indexed holding a representative textual vector of the text (right portion of the figure). In Figure 5, a screen-shot from IRIS after querying a database of about 50 images of men and cars is shown. On the top-right is the window for sketching queries. The top-center window serves to write textual queries and to attach information to the sketched regions. The bottom window shows the results of the query with the thumbnails of the retrieved images (left to right are the most similar). Finally, the window on the top-left controls the session.

We remark again the importance of moving from games to a distance measure and of identifying the topo-vectors for actually being able to implement the spatial framework. In particular, in IRIS once an image is place in the data-base the topo-vector for its related topological model is computed, thus off-line, and it is the only data structure actually used in the retrieval process. The representation is quite compact both if compared with the topological model and with the image itself. In addition, the availability of topo-vectors as indexing structures enables us to use a number of information retrieval optimizations, [20].

In IRIS, the similarity measure is built of three components:

\[
similarity(I_q, I_j) = \frac{1}{k_n} (k_n^{\text{topo}} \cdot d_{\text{topo}}(I_q, I_j) + k_n^{b} \cdot d_b(I_q, I_j) + k_n^{\text{text}} \cdot d_{\text{text}}(I_q, I_j))
\]
where $I_q$ is the query image (equipped with its topological model and textual description), $I_j$ is the $j$-th image in the visual database, $k_{\text{topo}}$, $k_b$, and $k_{\text{text}}$ are user-defined factors to specify the relative importance of topological intricacy, binary relationships and text in the querying process, $k_n$ is a normalizing factor, $d_{\text{topo}}(I_q, I_j)$ is the topo-distance between $I_q$ and $I_j$, $d_b(I_q, I_j)$ and $d_{\text{text}}(I_q, I_j)$ are the distances for the binary spatial relationships and for the textual descriptions, respectively.

The entire Section 5 is concerned with the computation of $d_{\text{topo}}(I_q, I_j)$. The topo-distance component is simply:

$$d_{\text{topo}}(I_q, I_j) = \text{topo-distance}(\text{t-vec}(I_q), \text{t-vec}(I_j))$$

The second component $d_b(I_q, I_j)$ of the similarity measure accounts for the binary spatial relationships between objects. When an image is indexed, a matrix is built. This is a square matrix whose indices range over the regions present in the model. The generic entry $e_{i,j}$ of the matrix represents the spatial relationship between region $i$ and region $j$ and can be one of the following: disconnected, externally connected, overlap, equal, tangential part, non-tangential part, and the inverses of the last two (RCC8). Following [19], we define a topological distance using RCC8 in the following way. Any two relations are at distance $n$ if there is a path of length $n$ in the graph in Figure 6 connecting the two nodes representing the relations. Our distance is slightly different from that in [19] since we use a modification of its original graph, though the underlying idea is the same. In the similarity measure, one compares matrices $b(M_1, M_2)$:

$$d_b(I_q, I_j) = b(b_{\text{matrix}}(I_q), b_{\text{matrix}}(I_j))$$

where $b_{\text{matrix}}(I_j)$ is the matrix of binary RCC8 relations associated with the regions identified in the $j$-th image.

The third and last component $d_{\text{text}}(I_q, I_j)$ of the similarity measure deals with textual annotation. The motivation comes from captions accompanying images in paper documents or present ‘near’ images in hyper-media documents. We employ quite standard textual information retrieval techniques, see for instance [20], and therefore omit further explanation of this part of the similarity

**Fig. 5.** The result of querying a database of men and cars.
A Spatial Similarity Measure based on Games: Theory and Practice

Fig. 6. The binary relationships graphs.

measure behalf for the standard definition of ‘textual distance’ between two image descriptions:

\[ d_{\text{text}}(I_q, I_j) = (1 - \frac{\text{weighted}\text{occurrences}(\text{text}\text{vector}(I_q), \text{text}\text{vector}(I_j))}{\text{length}(\text{text}\text{vector}(I_q))}) \]

where text\text{vector}(I_j) is the list of meaningful words found in the description of the \(j\)-th image, weighted\text{occurrences} counts the number of instances of a word appearing in two textual vectors weighted by a factor indicating the indexing power of the word. A word is more powerful if it discriminates more, which in turns means that it occurs in less descriptions in the whole collection of image captions. The \(d_{\text{ext}}(I_q, I_j)\) follows a common way of defining a cosine distance among word vectors, see for instance [34].

Received 2001