Deeply Semantic Inductive Spatio-Temporal Learning

Jakob Suchan¹, Mehul Bhatt¹, and Carl Schultz²

Human-Centred Cognitive Assistance¹ hcc.uni-bremen.de and The DesignSpace Group.,¹² www.design-space.org
University of Bremen¹, and University of Münster², GERMANY

Abstract. We present an inductive spatio-temporal learning framework rooted in inductive logic programming. With an emphasis on visuo-spatial language, logic, and cognition, the framework supports learning with relational spatio-temporal features identifiable in a range of domains involving the processing and interpretation of dynamic visuo-spatial imagery. We present a prototypical system, and an example application in the domain of computing for visual arts and computational cognitive science.

Keywords: Spatio-Temporal Learning; Dynamic Visuo-Spatial Imagery; Declarative Spatial Reasoning; Inductive Logic Programming; AI and Art

1 INTRODUCTION

Cognitive assistive technologies and computer-human interaction systems involving an interplay of space, dynamics, and cognition necessitate capabilities for explainable reasoning, learning, and control about space, actions, change, and interaction [1]. Prime application scenarios, for instance, include (A1–A5): (A1). activity grounding from video and point-clouds; (A2). modelling and analysis of environmental processes at the geospatial scale; (A3). medical computing scenarios replete with visuo-spatial imagery; (A4). visuo-locomotive human behavioural data concerning aspects such as mobility or navigation, eye-tracking based visual perception research; (A5). embodied human-machine interaction and control for commonsense cognitive robotics. A crucial requirement in relevant application contexts (such as A1–A5) pertains to the semantic interpretation of multi-modal human behavioural or socio-environmental data, with objectives ranging from knowledge acquisition (e.g., medical computing, computer-aided learning) and data analyses (e.g., activity interpretation) to hypothesis formation in experimental settings (e.g., empirical visual perception studies). The focus of our research is the processing and interpretation of dynamic visuo-spatial imagery with a particular emphasis on the ability to learn commonsense knowledge that is semantically founded in spatial, temporal, and spatio-temporal relations and patterns.

DEEP VISUO-SPATIAL SEMANTICS The high-level semantic interpretation and qualitative analysis of dynamic visuo-spatial imagery requires the representational and inferential mediation of commonsense abstractions of space, time, action, change, interaction and their mutual interplay thereof. In this backdrop, deep visuo-spatial semantics denotes the existence of declaratively grounded models — e.g., pertaining to
space, time, space-time, motion, actions & events, spatio-linguistic conceptual knowledge— and systematic formalisation supporting capabilities such as: (a) mixed quantitative qualitative spatial inference and question answering (e.g., about consistency, qualification and quantification of relational knowledge); (b) non-monotonic spatial reasoning (e.g., for abductive explanation); (c) relational learning of spatio-temporally grounded concepts; (d) integrated inductive-abductive spatio-temporal inference; (e) probabilistic spatio-temporal inference; (f) embodied grounding and simulation from the viewpoint of cognitive linguistics (e.g., for knowledge acquisition and inference based on natural language).

Recent perspectives on deep visuo-spatial semantics encompass methods for declarative (spatial) representation and reasoning —e.g., about space and motion— within frameworks such as constraint logic programming (rule-based spatio-temporal inference [4, 24]), answer-set programming (for non-monotonic spatial reasoning [27]), description logics (for spatio-terminological reasoning [3]), inductive logic programming (for inductive-abductive spatio-temporal learning [5, 6]) and other specialised forms of commonsense reasoning based on expressive action description languages for modelling space, events, action, and change [1, 2]. In general, deep visuo-spatial semantics driven by declarative spatial representation and reasoning pertaining to dynamic visuo-spatial imagery is relevant and applicable in a variety of cognitive interaction systems and assistive technologies at the interface of (spatial) language, (spatial) logic, and (visuo-spatial) cognition.

INDUCTIVE SPATIO-TEMPORAL LEARNING (WITH DEEP SEMANTICS)

This research is motivated by the need to have a systematic inductive logic programming [15] founded spatio-temporal learning framework and corresponding system that:

- provides an expressive spatio-linguistically motivated ontology to predicate primitive and complex (domain-independent) relational spatio-temporal features identifiable in a broad range of application domains (e.g., A1–A5) involving the processing and interpretation of dynamic visuo-spatial imagery.
- supports spatio-temporal relations natively such that the semantics of these relations is directly built into the underlying ILP-based learning framework.
- supports seamless mixing of, and transition between, quantitative and qualitative spatial data.

We particularly emphasise and ensure compatibility with the general setup of (constraint) logic programming framework such that diverse knowledge sources and reasoning mechanisms outside of inductive learning may be directly interfaced, and reasoning / learning capabilities be combined within large-scale integrated systems for cognitive computing.

2 LEARNING FROM RELATIONAL SPATIO-TEMPORAL STRUCTURE: A GENERAL FRAMEWORK AND SYSTEM

We present a general framework and working prototype for an inductive spatio-temporal learning system with an elaborate ontology supporting a range of space-time features;
we demonstrate the functional capabilities from the viewpoint of AI-based computing for the arts & social sciences, and computational cognitive science.

2.1 THE SPATIO-TEMPORAL DOMAIN $O_{SP}$, AND QS

The spatio-temporal ontology $O_{sp} \equiv_{def} <E, R>$ is characterised by the basic spatial entities ($E$) that can be used as abstract representations of domain-objects and the relational spatio-temporal structure ($R$) that characterises the qualitative spatio-temporal relationships amongst the supported entities in ($E$). The following primitive spatial entities are sufficient to characterise the learning mechanism and its sample application for this paper:

- **a point** is a pair of reals $x, y$;  
- **a vector** is a pair of reals $v_x, v_y$;  
- **an oriented point** consists of a point $p$ and a vector $v$;  
- **a line segment** is a pair of end points $p_1, p_2$ ($p_1 \neq p_2$);  
- **a rectangle** is a point representing the bottom left corner, a direction vector $v$ defining the orientation of the base of the rectangle, and a real width and height $w, h$ ($0 < w, 0 < h$);  
- **an axis-aligned rectangle** is a rectangle with fixed direction vector $v = (1, 0)$;  
- **a circle** is a centre point $p$ and a real radius $r$ ($0 < r$);  
- **a simple polygon** is defined by a list of $n$ vertices (points) $p_1, \ldots, p_n$ (spatially ordered counter-clockwise) such that the boundary is non-self-intersecting, i.e., there does not exist a polygon boundary edge between vertices $p_i, p_{i+1}$ that intersects some other edge $p_j, p_{j+1}$ for all $1 \leq i < j < n$ and $i+1 < j$.

Spatio-temporal relationships ($R$) between the basic entities in $E$ may be characterised with respect to arbitrary spatial and spatio-temporal domains such as mereotopology, orientation, distance, size, motion; Table 1 lists the relevant supported relations from the viewpoint of established spatial abstraction calculi such as the Region Connection Calculus [16], Rectangle Algebra and Block Algebra [7], LR Calculus [20], Oriented-Point Relation Algebra (OPRA) [14], and Space-Time Histories [8, 9].

QS – ANALYTIC SEMANTICS FOR $O_{SP}$  
We adopt an analytic approach to spatial reasoning, where the semantics of spatial relations are encoded as polynomial constraints within a (constraint) logic programming setup. The analytic method supports the integration of qualitative and quantitative spatial information, and provides a means for sound, complete and approximate spatial reasoning [4]. For example, let axis-aligned rectangles $a, b$ each be defined by a bottom-left vertex $(x_i, y_i)$ and a width and height $w_i, h_i$, for $i \in \{a, b\}$ such that $x_i, y_i, w_i, h_i$ are reals. The relation that $a$ is a non-tangential proper part of $b$ corresponds to the polynomial constraint:

$$(x_b < x_a) \land (x_a + w_a < x_b + w_b) \land (y_b < y_a) \land (y_a + h_a < y_b + h_b)$$

Continuing with the example, this is generalised to arbitrarily oriented rectangles. Determining whether a point is inside an arbitrary rectangle is based on vector projection. Point $p$ is projected onto vector $v$ by taking the dot product:

$$(x_p, y_p) \cdot (x_v, y_v) = x_p x_v + y_p y_v.$$  

With this approach, the task of determining whether a set of spatial relations is consistent then becomes the task of determining whether a system of polynomial constraints is satisfiable. We emphasise that our approach and framework are not limited to the above
<table>
<thead>
<tr>
<th><strong>Spatial Domain (QoS)</strong></th>
<th><strong>Formalisms</strong></th>
<th><strong>Spatial Relations (R)</strong></th>
<th><strong>Entities (E)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mereotopology</td>
<td>RCC-5, RCC-8 [16]</td>
<td>disconnected (dc), external contact (ec), partial overlap (po), tangential proper part (tpp), non-tangential proper part (npp), proper part (pp), part of (p), discrete (d), overlap (o), contact (c)</td>
<td>arbitrary rectangles, circles, polygons, cuboids, spheres</td>
</tr>
<tr>
<td>Rectangle &amp; Block algebra [7]</td>
<td>proceeds, meets, overlaps, starts, during, finishes, equals</td>
<td>axis-aligned rectangles and cuboids</td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td>LR [20]</td>
<td>left, right, collinear, front, back, on</td>
<td>2D point, circle, polygon with 2D line</td>
</tr>
<tr>
<td></td>
<td>OPRX [14]</td>
<td>facing towards, facing away, same direction, opposite direction</td>
<td>oriented points, 2D/3D vectors</td>
</tr>
<tr>
<td>Distance, Size</td>
<td>QDC [10]</td>
<td>adjacent, near, far, smaller, equi-sized, larger</td>
<td>rectangles, circles, polygons, cuboids, spheres</td>
</tr>
<tr>
<td>Dynamics, Motion</td>
<td>Space-Time Histories [8, 9]</td>
<td>moving: towards, away, parallel; growing / shrinking: vertically, horizontally; passing: in front, behind; splitting / merging</td>
<td>rectangles, circles, polygons, cuboids, spheres</td>
</tr>
</tbody>
</table>

Table 1. The Spatio-Temporal Domain \( \mathcal{O}_{sp} \) supported within the Learning Framework

entities; a wider class of 2D and 3D spatial entities are supported and may be defined as per domain-specific and computational needs [4, 18, 27, 19].

**INDUCTIVE LEARNING WITH THE SPATIAL SYSTEM \( \langle O_{sp}, QS \rangle \)** Learning is founded on the Aleph ILP system [21]. Learning spatio-temporal structures, is based on integrating the spatial ontology \( O_{sp} \) described above, into the basic learning setup of ILP.

**Given:** (1) A set of examples \( E \), consisting of positive and negative examples for the desired spatio-temporal structure, i.e., \( E = E^+ \cup E^- \), where each example is given by a set of spatio-temporal observations in the domain; (2) the (spatio-temporal) background knowledge \( B \).

The spatio-temporal learning domain is defined by basic spatial entities \( (E) \) constituting the domain objects, the relational spatial structure \( (R) \) describing the spatio-temporal configuration of spatial entities in the domain, and rules defining spatio-temporal phenomena and characteristics of the domain. In this context, spatio-temporal facts characterising the learning examples \( E \) can be given as, (a) numerical representation of domain objects, (b) qualitative relations between spatial entities, or (c) a mixed qualitative-quantitative representations, where the facts are partially grounded in numerical observations.

**Learning:** The learning task is defined as finding hypothesis \( H \) consisting of spatio-temporal relations \( (R) \) holding between basic spatial entities \( (E) \), such that \( H \cup B \models E^+ \), and \( H \cup B \not\models E^- \).

As such, the spatial ontology \( O_{sp} \) constitutes an integrated part of the learning setup and spatio-temporal semantics are available throughout the learning process.

### 3 LEARNING CINEMATOGRAPHIC PATTERNS AND THEIR VISUAL RECEPTION: THE CASE OF SYMMETRY

Aimed at cognitive film studies and visual perception research, we present a use-case pertaining to the (visual) learning of cinematographic patterns of symmetry and its vi-
Fig. 1. Positive examples for symmetric scene structures at the object level

To demonstrate the temporal aspect of the learning framework, we demonstrate the capability to learn “axioms of visual perception” from dynamic eye-tracking data; both the chosen films and their corresponding eye-tracking data are obtained from a large-scale experiment in visual perception of films [23, 22]. The presented example translates to a variety of cases involving visual perception and human behaviour studies.

**Learning Spatial Structures: Object-Level Symmetry**

As an example for learning spatial structures, we consider symmetry in the relative object placement in a movie scene (see Fig. 1). In particular, learning is based on the spatial configuration of people, faces, and their facing direction, directly obtained from computer vision algorithms as described in [23]. In this context, positive and negative examples, are given as numerical spatial facts about domain objects in the image.

```prolog
... detection(id(0), image(3), class(person), rectangle(point(319, 194), 319, 456)).
> detection(id(1), image(3), class(person), rectangle(point(678, 215), 367, 452)).
> detection(id(0), image(3), class(face), rectangle(point(438, 246), 86, 86)).
> detection(id(1), image(3), class(face), rectangle(point(745, 284), 87, 87)).
> 2d_facing_dir(id(0), image(3), vector(0.550864, 0.834595), magnitude(6.26042)).
> 2d_facing_dir(id(1), image(3), vector(-0.500519, 0.865726), magnitude(4.82556)).
...```

We define representations of domain objects linking the numerical description of objects in the image to basic spatial entities describing different aspects of these objects, e.g. the bounding box (rectangles), or the center-point (points).

```prolog
entity(center(person(P)), point(X, Y), image(Img)) :-
  detection(_T, image(Img), class(person), rectangle(point(Xr, Yr), W, H)),
  X is Xr + W/2, Y is Yr + H/2.
```

In addition to the detected domain objects, we define abstract geometric objects needed to describe symmetry, e.g. the symmetry axis in the center of the image.

```prolog
entity(symmetry_obj(center_axis), line(X, 0, X, Y), image(Img)) :-
  img(image(Img)), media_size(size(MediaWidth, MediaHeight), image(Img)),
  X is MediaWidth/2, Y is MediaHeight.
```

**Learning:** We learn the relational spatial structure consisting of qualitative spatial relationships characterising symmetry in the configuration of the spatial entities in the image, i.e. we consider relations of topology, orientation, distance, and size.

```prolog
:- modeh(1, symmetric(+img)).
:- modeb(*, entity(#ent,-obj,+img)).
:- modeb(*, topology(rcc8(#rel),+obj,+obj)).
:- modeb(*, distance(#rel,+obj,+obj)).
:- modeb(*, orientation(#rel,+obj,+obj)).
:- modeb(*, size(#rel,+obj,+obj)).
```

1 Our case-study is motivated by a broader multi-level interpretation of symmetry from the viewpoint of film cinematography [25]; however, the specific example of this paper focuses on one aspect of this multi-level symmetry characterisation involving relative object placement in a movie scene.
Exemplary symmetrical spatial structures, learned by the system include the following.

\[
\text{symmetric}(A) \leftarrow \text{entity(center(person(0)), B, A)}, \text{entity(center(person(1)), C, A)}, \text{entity(symmetry_object(center_axis), D, A)}, \text{distance(equidistant, D, C, B)}.
\]

\[
\text{symmetric}(A) \leftarrow \text{entity(person(0), B, A)}, \text{entity(person(1), C, A)}, \text{size(same, C, B)}.
\]

**Learning Spatio-Temporal Dynamics: Axioms of Perception**

We illustrate learning of spatio-temporal dynamics in the context of visual perception, by learning perceptual patterns from eye-tracking data and people tracks in a movie scene. As an example we focus on attention of a person switching from one individual to another.

\[
\text{detection(id(0), frame(426), class(person), rectangle(point(385,66),244,271))}.
\]

\[
\text{detection(id(1), frame(426), class(person), rectangle(point(111,68),332,276))}.
\]

\[
\text{gazepoint(frame(426), point(859,212))}.
\]

Learning:

We adapt the general learning setup of the example above, for learning spatio-temporal dynamics by introducing the predicate `holds_in/2` to denote that a spatial relation holds between two entities at a time point.

\[
\text{:- modeb(*, holds_in(topology(#rel, +ent, +ent), +time)).}
\]

\[
\text{:- modeb(*, time(#rel, +time, +time)).}
\]

\[
\text{...}
\]

Spatio-temporal dynamics constituting attention switches include the following.

\[
\text{att_switch(B) :- holds_in(topology(inside, gaze, person(1)), A)}, \text{holds_in(topology(inside, gaze, person(2)), B), time(consecutive, A, B)}.\]

**4 DISCUSSION AND OUTLOOK**

Directly comparable to this research is the line of work on integrated inductive-abductive reasoning for learning spatio-temporal relational models from video in [5, 6]; here, spatio-temporal learning in the context of ILP has only been addressed for the case of topological relations. Furthermore, the ILP learning framework does not have built-in semantics for the topological relations. Aside from this, learning relational spatial structures was investigated in the context of learning spatial relations from language[12], and within the geospatial domain [13, 26]. Probabilistic Logic Programming frameworks such as PRISM [17] and ProbLog [11] have been used for learning parameters, and the structure, of probabilistic logic programs, although (qualitative) spatial reasoning has not been directly addressed. The main point-of-departure of this paper with respect to the state of the art in (qualitative) spatial learning is that the semantics of spatial, temporal, and spatio-temporal relations are directly built within the inductive learning framework of ILP. Pragmatically, what this implies is that it is possible to seamlessly describe a learning problem using a generic relational spatio-temporal ontology directly as part of a logic programming based learning environment. To the best of our knowledge, such a general spatio-temporal learning framework with built in semantics for mixed qualitative-quantitative spatio-temporal reasoning capabilities has not been available before. Furthermore, the ontology of space-time features supported in our framework goes much beyond topological relations addressing orientation, distance, and size. Future research will focus on enhancing the expressivity of the spatio-temporal relations to cover a wider range of domain-independent features characterising spatio-temporal dynamics.
Bibliography


