Physical Bongard Problems

Erik Weitnauer and Helge Ritter

CoR-Lab, CITEC, Bielefeld University,
Universitätsstr. 21-23, 33615 Bielefeld, Germany
`{eweitnau,helge}@techfak.uni-bielefeld.de`

Abstract. In this paper, we introduce Physical Bongard Problems (PBPs) as a novel and potentially rich approach to study the impact the constraints of a physical world have on mechanisms of concept learning and scene categorization. Each PBP consists of a set of 2D physical scenes which are positive or negative examples of a concept that must be identified. We discuss the properties that make PBPs challenging, analyze computational and representational requirements for a computational solver, and describe a first implementation of such a system. It can solve a subset of non-trivial PBPs using a version space approach for achieving its scene categorizations. The key element is a physics engine that is used both for the construction of information-rich physical features and for the prediction of how a given situation might evolve.

Keywords: concept learning, scene categorization, physical understanding, physics simulation, analogy making

1   Introduction

Despite the complex and dynamic nature of the world we live in, we are able to make sense of what happens around us. Already in early childhood, we build a sophisticated conceptual knowledge of our physical reality and are able to predict and visualize the outcome of many dynamic situations. Attempts to understand these striking abilities and their underlying processes need to take into account the important role of our physical embodiment [1]. Being embodied in a physical world requires the ability to rapidly capture the 'essence' of situations with respect to their physical interaction properties. This includes recognizing configurations that can provide physical support for an intended action, judging the feasibility of moving pieces, or 'perceiving' the imminent instability in a particular arrangement.

Having studied aspects of this challenge using advanced robot platforms in the context of grasping and manipulation [2], we here wish to introduce a complementary approach whose aim is to provide a maximally parsimonious, yet very rich framework to study mechanisms of *physics-based categorization*. To this end we introduce a novel class of problems inspired by and extending earlier work on pattern categorization [3]. Their essential characteristic is the embedding of an analogy detection task in the domain of physical situations. We call

these problems *Physical Bongard Problems* (PBPs). Each PBP consists of a set of 2D physical scenes, containing four positive and four negative examples of the concept to be learned.

We argue that this problem class is well suited for research on our ability to conceptualize physical situations and make appropriate decisions in dynamic and interactive settings. Insights can be gained both by analyzing how humans solve PBPs, e.g., using questionnaires or eye tracking techniques, and by combining this empirical work with the development and analysis of computational solvers. In this paper, we introduce the domain of PBPs and give an overview of their properties and what makes them intricate to solve. We discuss the role of physical knowledge for PBPs, how it can be modeled using a standard physics engine and how a particular version space based solver implementation performs.
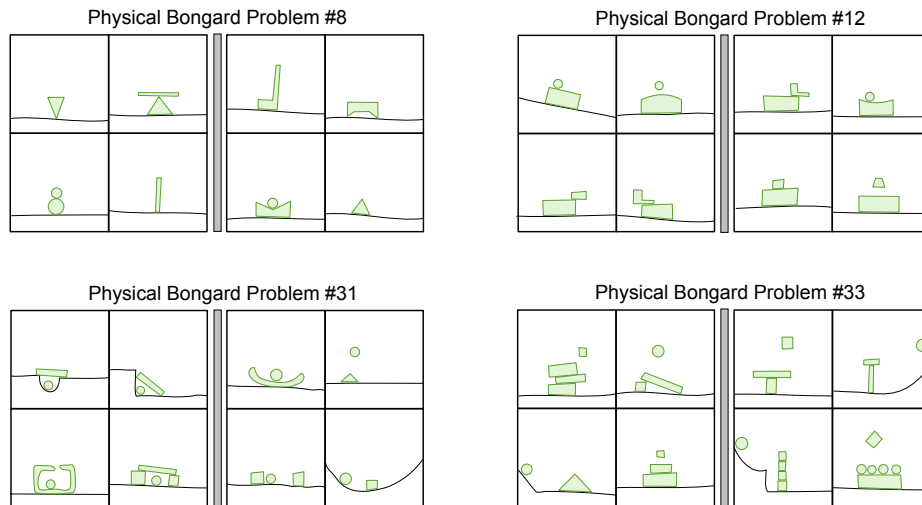
2   Physical Bongard Problems

In the design of Physical Bongard Problems we took inspiration from the class of Bongard problems (BPs), which are a set of 100 visual pattern recognition and categorization tasks, originally created by M. Bongard and extended by Douglas Hofstadter and Harry Foundalis [3–5]. Each BP consists of twelve images, six of them on the left and six on the right, all with an arbitrary pattern in black and white. The task is to identify the conceptual distinction between both sides. While many of the BPs are solved by humans rather intuitively, their computational solution is still an outstanding challenge.

In Physical Bongard Problems, while the task is the same, the images are taken from a physical domain, shifting the focus away from low-level visual processing towards dynamics and interaction. Instead of arbitrary static patterns, the images contain snapshots of 2D physical scenes depicted from a side perspective. PBPs can be considered as BPs which are more constrained by being 'embedded in physics', on the one hand, but can, on the other hand, as a consequence represent concepts not within the reach of the non-physical BPs. The scenes in PBPs may contain arbitrary-shaped non-overlapping rigid objects which do not move at the time $t = t_0$ of the snapshot. The solution of PBPs can be based on descriptions of the whole scene or parts of the scene at any point in time or on the reaction of objects to simple kinds of interaction, e.g., pushing. We have so far designed 34 PBPs, which can be viewed online [6]. Figure 1 depicts four of them.

2.1   Challenging Aspects

There are several challenging aspects of PBPs that make them both intricate to solve and an interesting object of research. In the following list, the first three aspects are unique to PBPs, while the further ones are shared by PBPs and classical BPs.

Fig. 1. Four Physical Bongard Problems. Solutions: (try yourself, first!) #8: configuration unstable vs. stable. #12: small object falls down vs. stays on large one. #31: circle is blocked vs. can be lifted. #33: construction stays intact vs. gets destroyed. See [6].

Physics. The need to invoke implicit physical knowledge of how the depicted object configuration will evolve (or respond to imagined physical interventions) for solving a problem is the main distinguishing characteristic of a PBP. This involves 'natural' assumptions, such as the association of some mass with each object and the presence of a downward directed gravity force. Using these assumptions, we can make physical judgments, e.g., about the stability of a configuration or predict likely states of motion (e.g., a 'ball' accelerating on a ramp).

Interaction. Physical understanding includes judgments about how objects might respond to imagined interventions. This is important in many situations in life, e.g., to judge whether some location can support my body, or how objects can be moved in a scene without causing unwanted inference to others.

Time. To see a scene as physical allows us to see it as a snapshot of a dynamical process. This connection generates a rich set of additional features arising from forward and backward predictions of the expected changes and can augment the scene with events that themselves are not depicted, like the collision of objects.

Grouping. Based on common features, relations or roles, several objects of one scene might have to be interpreted as a group to find a solution. There can be relations between groups or even groups of groups.

Focusing. In scenes with many objects, it is inefficient to consider the relations between all object pairs. Instead, a few important objects might have to be picked out while the others can be considered as 'background'.

Correspondence. When scenes contain groups of objects or relations between objects play a role, the mapping of two scenes requires to identify correspondence between two structured representations, which is highly non-trivial. This task is often referred to as analogy-making, an exciting research topic in itself.

Context. A suitable representation of a physical scene cannot be given a-priori, but depends on the context that is set by the other scenes. A single scene could be used in several PBPs and have a different interpretation in each of them, e.g., a different choice of what is the main object and what is the 'background'.

3   Computational Solvers for PBPs

3.1   Modeling Physical Knowledge

It is essential for solving PBPs to be able to predict and visualize the outcome of dynamic situations and interactions. We model this ability by giving the solver access to a physics engine[1] (PE). It is used in two ways: First, for the prediction of the unfolding of actions in the scenes. By constructing and simulating the scenes in the PE, the solver can inspect them at any time between the initial snapshot t_0 and the time all motion has stopped. Second, the engine is used to estimate physical object features. This includes features like object speed, acceleration and collision events, as well as concepts depending on interactions with objects in the scenes like pushing and pulling. We construct a basic notion of object stability by pushing the object briefly and observing its reaction. Its stability is judged by the distance it moves, where less movement correlates with more stability. A notion of the 'motion potential' or 'movability' of objects can be constructed by measuring the distance the objects can be pulled using a small force. A last feature derived from interaction is the role of an object as supporter of other objects in the scene. By imagining the scene without the object, i.e., by removing the object, it can be observed how the stability of the other objects depends on the removed one.

3.2   Implementation of a Basic Solver

An important and non-trivial decision for implementing any solver is the choice of a suitable input representation. Since PBPs are embedded in a physical world and only contain closed objects above some ground, the outlines and positions of these objects can be used as input representation without restricting the problem domain or making the problems significantly easier to solve. Using this input representation, we implemented a basic solver based on the version space algorithm for concept learning [7]. The hypothesis space contains all vectors `<side, numbers, distances, sizes, shapes, stabilities>`, where `side` is the side of the PBP ('left' or 'right'), `numbers` is a range of object count and the

---

[1] A physics engine is a piece of software that can perform physical simulations. We used the free Box2D physics engine in our experiments. See `http://box2d.org/`.

remaining elements are disjunctions of feature values ('small', 'medium', 'large' or 'rectangle', 'circle', 'triangle', 'other' or 'stable', 'unstable', 'moving' or 'near', 'far', respectively). All elements except `side` can also take the value '?', in which case they match any scene. For example, the meaning of the hypothesis `<left, 1-3, ?, small or large, ?, stable>` is "all left scenes (and none of the right scenes) contain one to three objects that are small or large-sized and stable". The algorithm starts with a set of all possible hypotheses and then removes the incompatible ones for each scene. Finally, among the remaining hypotheses, the one with the shortest length is chosen as solution. If no solution could be found at $t = t_0$, the algorithms is applied to the scenes at $t = t_{\mathrm{end}}$.

Results The presented algorithm can solve the PBPs 1 to 5, 8, 11 and 18. It demonstrates the successful application of a physics engine in concept learning of dynamic physical scenes and constitutes a baseline for PBP solvers. Yet, due to its simplicity, the subset of PBPs that it solves still is small. It could be extended by adding more object and scene features. However, there are some principal limitations that cannot be overcome this way. Of the challenging aspects listed in Section 2.1, the present algorithm addresses *physics*, *interaction* and *time*, where the handling of time is only rudimentary and not sensitive to changes, durations or events. The other complexity sources of *grouping*, *focusing*, *correspondence* and *context sensitivity* cannot be adequately addressed by the present algorithm because 'flat' feature vectors are used to describe the scenes. Therefore, e.g., the selection of key objects in the scene and relations between objects cannot be captured (See PBP 12, 31 and 33 in Figure 1). The step to a more powerful algorithm will involve the use of structured scene representations. Building and mapping these representations is a task of analogy-making, and we will report in a subsequent paper on extensions along this line.

4   Related Work

The interpretation of and reasoning about physical scenes has a long tradition in artificial intelligence in the field of qualitative physics, where physical knowledge is represented as high-level logical rule systems [8, 9]. We chose to provide physical knowledge in another, less rigid and more analog form: with a physics engine. This way, we are not committed to a certain level of abstraction and can start building representations at a low level.

Traditionally, much research on concept learning has been done in the context of unstructured domains where a concept can be represented as a set of attribute values [10]. Recently, more attention was paid to learning structured concepts in the domain of description logics [11]. The learning of concepts from dynamic examples as presented in this work, has not been in the focus of concept learning research, so far.

There have only been few attempts to develop computational solvers for classical Bongard problems. The only solver that is able to come up with solutions of some BPs without using hand-crafted input representations is the 'Phaeaco'

system by H. Foundalis [5]. It builds and maps representations in a process of analogy-making performed by a complex adaptive system. See [12] for a summary of computational approaches to analogy-making.

5   Conclusion

In this paper, we made two main contributions. First, we introduced Physical Bongard Problems as a novel research tool for concept learning and scene categorization by agents situated in a physical world. We discussed the aspects of Physical Bongard Problems that make them a challenge for computational solvers, which are *physics, interaction, time, grouping, focusing, correspondence* and *context sensitivity*. As a second contribution, we demonstrated how a physics engine can be effectively used to equip an algorithm with the physical understanding necessary to solve PBPs. The engine is used for both scene prediction and construction of information-rich physical features through simulated object interactions. We showed the feasibility of this approach with a basic PBP solver implementation and discussed its limitations, which are mainly a result of using unstructured collections of features as scene representations.

The step to a more powerful solver will require the use of structured representations and the extension of the basic solver with dynamic scene-encoding and structure-mapping capabilities. These two abilities are central topics in analogy-making and we are currently exploring how existing analogy-making algorithms can be adapted for the use in a PBP solver.

References

1. Pfeifer, R., Bongard, J., Grand, S.: How the body shapes the way we think: a new view of intelligence. The MIT Press (2007)
2. Ritter, H., Haschke, R., Rthling, F., Steil, J.: Manual intelligence as a rosetta stone for robot cognition. Robotics Research (2011) 135146
3. Bongard, M.M.: Pattern Recognition. Rochelle Park, N.J.: Hayden Book Co., Spartan Books (1970)
4. Hofstadter, D.R.: Gödel, Escher, Bach: an eternal golden braid. Harvester Press (1979)
5. Foundalis, H.E.: Phaeaco: A cognitive architecture inspired by Bongard's problems. PhD thesis, Indiana University (2006)
6. Weitnauer, E.: Physical bongard problems. http://naive-physics.com/pbp/ (2012)
7. Mitchell, T.M.: Generalization as search. Artificial intelligence **18**(2) (1982) 203226
8. Forbus, K.D.: Qualitative process theory: Twelve years after. Artificial Intelligence in Perspective **59**(1) (1994) 115
9. Kurtoglu, T., Stahovich, T.F.: Interpreting schematic sketches using physical reasoning. In: AAAI Spring Symposium on Sketch Understanding. (2002) 7885
10. Goodman, N.D., Tenenbaum, J.B., Feldman, J., Griffiths, T.L.: A rational analysis of Rule-Based concept learning. Cognitive Science **32**(1) (2008) 108154
11. Lehmann, J.: DL-Learner: learning concepts in description logics. The Journal of Machine Learning Research **10** (2009) 26392642
12. Gentner, D., Forbus, K.D.: Computational models of analogy. Wiley Interdisciplinary Reviews: Cognitive Science **2**(3) (2011) 266276