Looking for Logic in Vision

Paulo Santos, Derek Magee and Anthony Cohn
{psantos,drm,agc}@comp.leeds.ac.uk
School of Computing, University of Leeds, UK

Abstract

This paper presents how the axioms of equivalence can be induced from visual observation without assuming background knowledge. The present research was conducted upon a system that combines computer vision with inductive logic programming that was first designed to learn protocol behaviour from observation.

1 Introduction

Two important requirements for an autonomous agent to understand and act in any real environment are the ability to hypothesise about perceptual information and the ability to act according to the acceptable behaviour (*protocol behaviour*) in that environment. In this paper we use a system that combines computer vision with an inductive logic programming (ILP) language, which is here used to construct general mathematical structures from visual data; PROGOL, Muggleton (1995, 1996), is utilised as the ILP language. Our system has been initially designed to learn protocol behaviours from sensor data that could further be incorporated in a virtual agent. In this work, however, we use this system to show how some general mathematical structures (such as the equivalence axioms) can be induced from the visual data.

The framework proposed in this this paper has been evaluated on the visual observation of simple game scenarios. In brief, the system observes two players engaged in the game described as follows: two objects (that are implicitly ordered) are initially put on a board, the game consists of keeping on the table the greater object while the other is replayed. Both objects are withdrawn from the table when their faces show the same figure. The challenge here is to generate the transitivity, reflexivity and symmetry axioms from the observation of this game without assuming any preconceived notion of number or any pseudo definition of ordering. This represents a key difference of this work and previous research on high-level image interpretation such as Nagel (2000); Fern et al. (2002); Santos and Shanahan (2002), which were characterised by the interpretation of sensor data given domain-specific background knowledge.

By setting our system the task of learning basic axioms of mathematics we do not intend to develop a system to assist mathematicians (such as the systems described by Colton et al. (2000) and Lenat (1983)) but to investigate how common sense knowledge can be automatically induced from computer vision data. The long term purpose of this research is provide an autonomous system with the necessary machinery that will allow it to formulate its own logical explanations about its environment.

The experimental setup used in this work is composed of a video camera observing a table top where two players are engaged in playing a game. The output of the vision system is sent to a PROLOG meta program whose tasks are three fold. First, it re-writes the vision data into multiple PROGOL experiments. The meta program also runs these experiments and, finally, evaluates their outputs. Although the automatic evaluation of the output is an important issue in this work, it is outside the scope of the present paper.

In effect the meta program selects subsequent pairs of state descriptions from the vision data generating, for each pair, a predicate $trans(s_i, s_{i+1})$, read as *the transition from state s_i to state s_{i+1}*. Or aim, thus, is to search for recursive rules from the data defined with the predicate $trans/2$.

2 Looking for equivalence

Let t_i ($i \in [1, \ldots, n]$) be a time point and ci,cj,ck,cr,cs, cw and cn represent seven of the fifteen classes assumed in the segmentation algorithm. A typical sequence of the assumed game (and the relative output of the vision system) is shown in Figure 1

In order to allow the system to learn the axioms of equivalence, an over classification of the paper-scissors-stone figures was assumed. The three objects of the game were classified into 15 distinct classes. From the symbolic-learning standpoint, this assumption is equivalent to showing, at each round of the game, one of 15 different shapes of papers, scissors and stones. The axioms of equivalence are obtained from analysing the states in which a draw occurs, i.e., states containing pairs of objects that are followed by an empty state (e.g., Figures 1(i) and 1(iii)).

Eight data sets (containing an average of 30 examples each) were utilised in this experiment. The following formulae were obtained from our meta program:

$$trans([A, B], []) :- \quad trans([A, C], []),$$

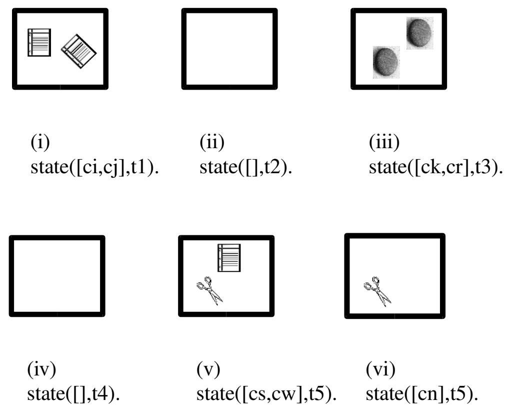

Figure 1: Six states of the paper-scissor-stone game.

$$\begin{aligned}
& trans([C,B],[]). && (1)\\
trans([A,A],[]). & && (2)\\
trans([A,B],[C]) :- \; & trans([A,D],[C]), && \\
& trans([B,A],[C]). && (3)\\
trans([A,B],[]) :- \; & trans([B,A],[]). && (4)\\
trans([A,B],[]) :- \; & trans([C,A],[]), && \\
& trans([C,B],[]). && (5)\\
trans([A,B],[]) :- \; & trans([C,A],[]), && \\
& trans([B,C],[]). && (6)\\
trans([A,A],[B]) :- \; & trans([A,C],[B]). && (7)\\
trans([A,B],[C]) :- \; & trans([A,D],[C]), && \\
& trans([B,D],[C]). && (8)
\end{aligned}$$

The answer set above includes a transitivity rule (Formula 1) followed by the reflexivity axiom (Formula 2). The symmetry axiom (Formula 4) came in fourth, preceded by a spurious rule (Formula 3). Formula 5 and 6 and the final rule (Formula 8) are three rules expressing the transitivity of $trans/2$ that are equivalent to Formula 1. They were not subsumed by Formula 1 because the symmetry of the predicate $trans/2$ was not a priori assumed in the learning process, but obtained as a result (Formula 4). Finally, Formulae 7 seems to be the result of the generalisation of a noisy portion of the data. These formulae were ranked according to the voting criteria proposed in Santos et al. (2004), the discussion of the ranking method, however, is outside the scope of this paper.

Last but not least, in addition to the axioms of equivalence, the process of constructing the axioms of equivalence produced a minimal set of examples of ground atomic formulae representing the equivalence between symbols in the given data set. With this minimal set, and the transitivity and symmetry axioms, a complete set of equivalences may be determined.

3 Conclusion

In this paper, transitivity, reflexivity and symmetry axioms were induced from computer vision data obtained from the observation of a simple game playing scenario. It is worth pointing out that the simplicity of the scenarios does not compromise the importance of the findings. The application of the axioms obtained is not constrained to the scenarios where they were inferred, but they are general rules that are present in a variety of reasoning processes. The further use of these rules on diverse domains is an issue to be taken into account by future investigations.

Also subject to future investigations is the inclusion of an actuator module in the system. The theories constructed from observation could, thus, be executed in the real world and, as a result, be further refined according to the interaction of the agent with its environment.

References

Colton, S., Bundy, A., and Walsh, T. (2000). Automatic identification of mathematical concepts. In *Proc. 17th International Conf. on Machine Learning*, pages 183–190. Morgan Kaufmann, San Francisco, CA.

Fern, A., Siskind, J. M., and Givan, R. (2002). Learning temporal, relational, force-dynamic event definitions from video. In *Proc. of AAAI/IAAI*, pages 159–166.

Lenat, D. (1983). Eurisko: A program which learns new heuristics and domain concepts. *Artificial Intelligence*, 21:61–98.

Muggleton, S. (1995). Inverse entailment and Progol. *New Generation Computing, Special issue on Inductive Logic Programming*, 13(3-4):245–286.

Muggleton, S. (1996). Learning from positive data. In Muggleton, S., editor, *ILP96*, volume 1314 of *LNAI*, pages 358–376. SV.

Nagel, H.-H. (2000). Image sequence evaluation: 30 years and still going strong. In *Proc. of ICPR*, pages 1149–1158, Barcelona, Spain.

Santos, P., Magee, D., and Cohn, A. (2004). Combining multiple answers for learning mathematical structures from visual observation. manuscript.

Santos, P. and Shanahan, M. (2002). Hypothesising object relations from image transitions. In van Harmelen, F., editor, *Proc. of ECAI*, pages 292–296, Lyon, France.