Interpreting imprecise diagrams

Conference or Workshop Item

Interpreting Imprecise Diagrams

Neil Smith, Pete Thomas, and Kevin Waugh

Open University, Milton Keynes, MK7 6AA, UK,
{N.Smith, P.G.Thomas, K.G.Waugh}@open.ac.uk

Abstract. Imprecise diagrams occur in many situations. We propose a five-stage architecture for interpreting such diagrams, and have implemented this approach in an automatic grading tool for answers to examination questions. The approach is based on identifying (possibly malformed) minimal meaningful units and interpreting them to yield a meaninful result. Early indications are that the tool's performance is similar to human markers.

1 Introduction

An imprecise diagram is one where required features are either malformed or missing, or extraneous features are included. Imprecise diagrams frequently occur in student answers to examination questions. This problem has arisen from our investigations of the automatic grading of free-form text answers [5].

The automatic marking of answers in a textual form has received much attention over the years [2]. Our approach is similar in that we currently do not attempt to address any higher-order semantic structures above the "association"; this is similar to looking for key words or phrases in a sentential answer.

Much of the activity in diagrammatic reasoning has been directed towards precise diagrams, such as the use of diagrams in mathematical proofs [4] and visual query interfaces to GISs [1, 3].

DATsys [6] is a diagrammatic front end to the CourseMaster marking system. DATsys provides a method for creating bespoke diagram editors, but does little to address how those diagrams are marked. In contrast, our work concentrates on the marking of possibly ill-formed or inaccurate diagrams.

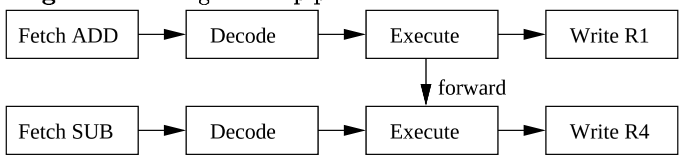
2 Our Approach

Our general approach interpreting imprecise diagrams consists of five stages: segmentation, assimilation, identification, aggregation, and interpretation. *Segmentation* and *assimilation* together translate a raster-based input into a set of diagrammatic primitives, e.g. boxes and text. In the *identification* stage, domain knowledge is used to identify low-level, domain-specific features. These features are *aggregated* into higher-level, abstact features. Finally, the diagram is *interpreted* to produce meaningful results, such as a grade.

Fig. 1. An example examination question

Draw a diagram showing how the data hazard inherent in the execution of the pair of instructions `ADD R2,R3,R1` ; `SUB R1,R5,R4` by a 4-stage pipeline, is overcome.

Fig. 2. A drawing of two pipelines with a link between them



We have followed this approach to implement a simple automatic marking tool. Since our primary interest is in the interpretation of imprecise diagrams, we avoided the segmentation and assimilation phases through the use of a drawing tool. Following the lead of statistical NLP, we surmised that the aggregation phase would not be needed for simple interpretation tasks, such as marking.

Malformed features are handled by an inference mechanism in the identification stage which results in a well-formed diagram. Missing and extraneous features are dealt with in the interpretation phase; the former by the appropriate allocation of marks, the latter are simply ignored.

3 Identification and Interpretation

The question shown in figure 1 was set in an online examination; its model solution is shown in figure 2. The significant features in this solution are the pair of four-stage piplines and the "forward" link between them.

In this domain, we have taken the smallest meaningful unit to be an *association*, which is a pair of boxes connected by a link. Given this definition, we treat a diagram as a (possibly overlapping) set of associations.

Students often represented the ordering of the stages in a pipeline by the relative positioning of boxes on the canvas as shown in Figure 3. These implicit associations were recognized in the identification stage by assuming that that two adjacent boxes were associated. This limited form of interpretation was sufficient for the examples we have seen to date.

Interpretation was made on the basis of matching each association in the given answer with each association in the solution and deriving a similarity

Fig. 3. Order indicated by relative position

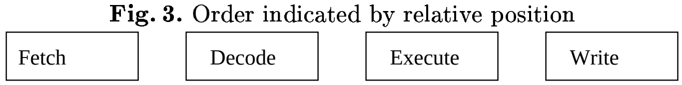


Table 1. Human *vs.* automatic marking

	Human	Automatic
Mean	2.7	3.1
St. Dev.	0.837	0.894

measure (a value in the range [0, 1]) for each pair, based on the texts in the association. Each value was multiplied by the weight of the solution association, and the matches that maximised the total mark were used to determine the total mark.

4 Results: Testing the automatic marker

In an on-line mock exam, 13 students used a simple drawing tool to create their answers to the question posed above. The answers were marked by the marking tool and four independent human markers. The results are given in Table 1; the maximum mark for the question was 4.0. A Student's t-test of these data indicates that we cannot reject the null hypothesis that the automatic marker is significantly different from the human markers.

Despite the small sample, the results are sufficiently encouraging to suggest that this method of diagram marking is a feasible approach to the problem. Therefore, we intend to apply this method to more complex domains (such as E-R diagrams). We also need to test whether the aggregation stage is of real value, and devise alternative techniques in the identification and intrepreation stages.

References

1. Anderson, M., McCartney, R. (2003) "Diagram processing: Computing with diagrams", *Artificial Intelligence* **145** (1-2): 181–226.
2. Burstein, J., Leacock, C., Swartz, R. (2001). "Automated Evaluation of Essays and Short Answers". *Fifth International Computer Assisted Assessment Conference*, Loughborough University, UK, pp. 41–54.
3. Donlon, J.J., Forbus, K.D. (1999). "Using a geographic information system for qualitative spatial reasoning about trafficability", *Proceedings of the Qualitative Reasoning Workshop*, Loch Awe, Scotland.
4. Jamnik, M (1998) *Automating Diagrammatic Proofs of Arithmetic Arguments*, PhD thesis, University of Edinburgh.
5. Thomas, P.G., Price, B., Paine, C. and Richards, M. (2002) "Remote Electronic Examinations: an architecture for their production, presentation and grading". *British Journal of Educational Technology*, **33** (5) 539–552.
6. Tsintsifas A., (2002), "A Framework for the Computer Based Assessment of Diagram-Based Coursework", Ph.D. Thesis, Computer Science Department, University of Nottingham, UK