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 (those with malformed, missing, or extraneous features) occur in many situations. We propose a five-stage architecture for interpreting such diagrams and have implemented a tool, within this architecture, for automatically grading answers to examination questions. The approach is based on identifying (possibly malformed) minimal meaningful units and interpreting them to yield a meaningful result. Early indications are that the tool's performance is similar to that of 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. The need to handle such diagrams has arisen from our investigations of the automatic grading of free-form text answers [5].

The automatic grading 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, which is equivalent 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]. One diagrammatic grading system is DATsys [6], 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 graded. In contrast, our work concentrates on the grading of possibly ill-formed or inaccurate diagrams.

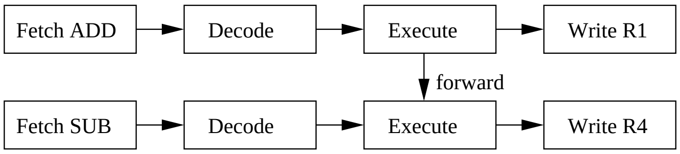
2 Our Approach

Our general approach to 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, the *minimal*

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



meaningful units. These features are *aggregated* into higher-level, abstract features. Finally, the diagram is *interpreted* to produce meaningful results, such as a grade.

We have implemented the identification and interpretation stages in a simple automatic grading tool. Since our initial interest is in these stages, we avoided the need for the segmentation and assimilation phases by using a drawing tool for input. Given that our previous work in automatic grading [5] has been successful without the recognition of abstract features, we decided to explore the problem, in the first instance, without aggregation.

Malformed features are handled by an inference mechanism in the identification stage which identifies and repairs such features. Missing and extraneous features are dealt with in the interpretation phase: the former through the appropriate allocation of marks, while 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 pipelines and the “forward” link between them.

In this domain, we have taken the minimal 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.

Under exam conditions, students often represented the ordering of the stages in a pipeline by the relative positioning of unconnected boxes (e.g. left-to-right, top-to-bottom). These implicit associations were recognized in the identification stage by assuming that two adjacent boxes were associated. This limited form of identification was sufficient for the examples we have seen to date.

Interpretation of the diagram was guided by a “marking scheme”, where each association in the solution has a mark allocated. Each association in the answer was compared to each association in the solution, to give a set of similarity measures. The mark for each comparison was calculated by scaling the allocated

Table 1. Human *vs.* automatic grading

	Human	Automatic
Mean	2.78	2.73
σ	1.05	1.09

mark by the similarity measure. The final mark was found by summing the highest scaled marks.

4 Preliminary Evaluation

In an on-line mock exam, 13 students used a simple drawing tool to create their answers to the question posed in figure 1. The answers were graded by the tool and by four independent human markers. The results are given in table 1; the maximum mark for the question was 4.0. The Pearson correlation coefficient for these data is 0.75 which is significant for $n = 13$ at the $p < 0.01$ level (2-tailed), indicating that the automatic grader performs very similarly to the human markers.

Despite the small sample, the results are sufficiently encouraging to suggest that this method of diagram grading 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 aggregation will be of real value, and to devise alternative techniques in the identification and interpretation 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