Generating Mathematical Word Problems

Sandra Williams

Computing Department
The Open University
Milton Keynes, MK7 6AA, United Kingdom
s.h.williams@open.ac.uk

Abstract

This paper describes a prototype system that generates mathematical word problems from ontologies in unrestricted domains. It builds on an existing ontology verbaliser that renders logical statements written in Web Ontology Language (OWL) as English sentences. This kind of question is more complex than those normally attempted by question generation systems, since mathematical word problems consist of a number of sentences that communicate a short narrative (in addition to providing the relevant numerical information required to solve the underlying mathematical problem). Thus, they embody many research issues that do not crop up with single-sentence questions. As well as describing the prototype system, I discuss five ways in which the difficulty of the generated questions may be controlled automatically during generation.

Introduction

Numeracy amongst the general population is shockingly low, even in developed countries such as the U.K. where more than one in five adults do not have the skills to perform everyday tasks such as calculating the correct change when paying for a small item with cash (Williams et al. 2003). Automatic aids such as intelligent tutoring systems might help alleviate the problem; however, a key bottleneck in their development is collecting a sufficiently large number of questions to present to students. Ideally, such systems require access to many thousands of questions. Students need practice questions that present similar kinds of problems so that they learn to generalise problem-solving techniques and cope with a wide variety of contexts. Natural Language Generation (NLG) seems an obvious solution. It has already been applied to the problem of generating single-sentence multiple choice questions (Mitkov and Ha 2003; Papasalouros, Kanaris, and Kotis 2008). With NLG, there is potential for great flexibility in the content and presentation of questions through variations in choice of input data, discourse structures, sentence structures, and words.

I consider the problem of generating more complex types of questions that pose a greater challenge for both NLG systems and students than multiple choice: mathematical word

problems (MWPs) (Verschaffel, Greer, and Corte 2000). These questions are expressed as multi-sentence short narratives in which some numerical information is given; they end with a request for the student to calculate an unknown quantity. Normally, solving such problems requires a range of skills: literacy skills for reading the question, skills to recognise the problem type and skills to apply simple arithmetical operators. One type of MWP is known as Multiplicative Compare (Cetintas et al. 2009):

> *A formica japonica worker ant weighs 0.004 gram. It can walk while holding in its mouth an object weighing 5 times as much as its own body. How many grams can a worker ant carry?*

Here, two numerical expressions '0.004 gram' and '5 times' are woven into a story in which the main protagonist is an ant. We are told that the ant is of type 'formica japonica worker'; that the ant walks; that the ant holds an object in its mouth; and so on; pieces of information that are extraneous to the main problem. At the end of a story, there is a request for an unknown value that the student must calculate. MWP questions may also contain additional 'distractor' numerical values that make identifying the ones required for the arithmetical operation more difficult. In this paper I will focus on arithmetical problems.

Another bottleneck in the development of intelligent tutoring systems is finding and adapting suitable input data from which to build questions. Our research investigates whether the explosion of ontologies and linked data on the semantic web will provide a solution. I propose a method for locating patterns in statements in ontologies that will enable us to populate mathematical problem templates with suitable knowledge. Exploiting the logical structures in ontologies should ensure that the knowledge is semantically valid.

I adapted an existing NLG toolset for generating syntactically correct multi-sentence MWP questions and demonstrate five ways in which the difficulty of the generated problems may be varied automatically.

A worked example from a prototype system

I demonstrate a method by which MWP questions may be generated from any ontology that contains OWL DataPropertyAssertion statements (axioms) with literal values that are

OWL statements	SWAT English sentences
ClassAssertion(Class(#island) NamedIndividual(#benbecula))	Benbecula is an island
ObjectPropertyAssertion(ObjectProperty(#memberOf) NamedIndividual(#benbecula) NamedIndividual(#uistAndBarraArchipelago))	Benbecula is member of the Uist and Barra Archipelago
DataPropertyAssertion(dataProperty(#hasPopulation) NamedIndividual(#benbecula) Literal(datatypeIRI=#integer 1219))	Benbecula has a population of 1219
ClassAssertion(class(#island) NamedIndividual(#southUist))	South Uist is an island
ObjectPropertyAssertion(ObjectProperty(#memberOf) NamedIndividual(#southUist) NamedIndividual(#uistAndBarraArchipelago))	South Uist is member of the Uist and Barra Archipelago
DataPropertyAssertion(DataProperty(#hasPopulation) NamedIndividual(#southUist) Literal(datatypeIRI=#integer 1818))	South Uist has a population of 1818

Table 1: OWL statements describing two individual that are comparable for MC problems.

defined as integers. For example, the travel ontology[1] contains the following statement:

```
DataPropertyAssertion(
  DataProperty(#hasPopulation)
  NamedIndividual(#benbecula)
  Literal(DatatypeIRI=integer 1219))
```

This means that an individual, #benbecula has the data property #hasPopulation with an integer value of 1219.

The first task is to locate two individuals in the ontology that are similar enough to be compared in an MC problem. We can define them as members of the same class, possessing the same data property with different values and optionally sharing some additional property (not essential since it merely provides 'padding' for the question). Table 1 (left column) shows OWL statements for two such individuals (Benbecula and South Uist) that are both members of the island class with different values for data property #hasPopulation and an additional shared property #memberOf. The right-hand column shows English sentences produced by SWAT Tools (Stevens et al. 2010)[2] which, amongst other things, derive lexical entries for each identifier (e.g., #memberOf becomes the singular phrase 'is member of' and the plural phrase 'are members of') and English sentences for each OWL statement.

The next tasks are to refactor and aggregate (Williams and Power 2010) the retrieved OWL statements and arrange them to form a MWP question, e.g.:

```
ClassAssertion(ObjectIntersectionOf(
    Class(#Island)
    ObjectHasValue(
        ObjectProperty(#memberOf)
        NamedIndividual(#uists_
          and_barra_archipelago)))
      [NamedIndividual(#benbecula)
      NamedIndividual(#south_uist)])
DataPropertyAssertion(
    DataProperty(#hasPopulation)
```

---

[1] The travel ontology is available from the TONES repository at owl.cs.manchester.ac.uk/repository/

[2] http://swat.open.ac.uk/tools/

```
    NamedIndividual(#benbecula)
    Literal(DatatypeIRI=#integer 1219))
DataPropertyAssertion(
    DataProperty(#hasPopulation)
    NamedIndividual(#south_uist)
    Literal(DatatypeIRI=#integer 1818))
What is the ratio of the two
    <DataProperty(#hasPopulation)>?
```

With some extensions to the SWAT Tools lexicon module (improvement of lexical entries derived from data properties) and to the grammar (addition of a rule to produce a relative clause for the first statement), the first three OWL statements can be realised directly by the grammar. The final sentence is produced by slotting a noun phrase derived from the lexical entry for the data property #hasPopulation into a template. The output is then:

Benbecula and South Uist are islands that are members of the Uists and Barra Archipelago. Benbecula has a population of 1219. South Uist has a population of 1818. What is the ratio of the two populations?

Key factors affecting the difficulty of MWPs

There are a number of factors that affect the difficulty of MWPs, some that can be controlled automatically within our prototype question generator and others that have not yet been tackled.

Factors that can be controlled automatically

In the prototype MWP question generation system, I have identified some key factors that can be varied automatically:

1. Readability of the output text

2. Inclusion of distractor numerical values

3. Introduction of extraneous information into the narrative

4. Order of presentation of numerical values

5. Conceptual difficulty of the mathematical problem

Variations in any of these can affect the difficulty of a MWP.

The first, readability, can be *increased* by, for instance, generating *simple* sentences and, conversely, *decreased* by generating more *complex* ones. Sentence complexity is related to sentence length, a factor known to affect readability (Coleman 1962); indeed, it is a component of readability formulae (Flesch 1949). Regarding the relationship between readability and difficulty of MWPs, Wheeler and McNutt (1983) found that low-achieving students performed significantly better on MWPs with simple sentences than on ones with complex sentences. Sentence complexity can be controlled within our current generator by switching on or off the aggregation and refactoring processes so that a version of the MWP described above could be generated with shorter sentences:

Benbecula is an island. South Uist is an island. Benbecula is a member of the Uists and Barra Archipelago. South Uist is a member of the Uists and Barra Archipelago. Benbecula has a population of 1219. South Uist has a population of 1818. What is the ratio of the two populations?

The second factor is distractor numerical values which are often added to increase the difficulty of MWPs. Provided that an ontology contains additional data properties relating to the two named individuals, extra sentences with distractor values can be generated, e.g.:

Benbecula has an area of 124.

However, although this sentence reflects what is present in the ontology, a unit of measurement for the land area is missing. Units were not present in the ontology and the present system does not handle them (see Limitations).

The third factor, extraneous information, discussed in the introduction, may be removed so that a question is simplified by stripping it down to essentials, e.g.:

Benbecula has a population of 1219. South Uist has a population of 1818. What is the ratio of the two populations?

The fourth factor is order of presentation of numerical values. Cohen and Stover (1981) found that when asked to modify MWPs for peers, students rearranged the order of numerical values to facilitate mathematical operations. It seems, then, that this ordering is at least perceived as playing a role in question difficulty. In the MWP above, the smaller population comes first which is the natural ordering for calculating a ratio, e.g., 2:3, but the reverse order would be more natural for performing a subtraction. Of course, it is easy to generate sentences in different orders to modify the difficulty of a MWP.

The fifth factor, conceptual difficulty of the mathematical problem, I address by consulting a mathematics curriculum for schools (Qualification and Curriculum Authority 1999). This builds on previous work (Williams and Power 2009) in which we assumed that in basic skills mathematics courses, concepts are taught in order of increasing difficulty. Thus we view addition of single-digit numbers (taught at an early stage) as simpler than calculating the ratio of two numbers and giving the answer as a percentage (taught at later stage).

Thus, the system could generate a question that would be easier from this point of view by requesting an addition calculation rather than a ratio:

Benbecula is an island. Benbecula has a population of 1219. South Uist is an island. South Uist has a population of 1818. What is the total population of the two islands?

Other important factors for future research

Of course, there are other factors that control the difficulty of MWPs that are outside the scope of the present work. For example, a factor that has been discussed widely in the literature is authenticity (Palm 2009). MWPs are normally regarded as a means to teach students that abstract mathematical principles learnt in the classroom can be applied in real world situations. If that is the aim, then the closer a MWP models reality the better (i.e., the model should be 'authentic'). Certain MWPs can then be criticised for anticipating 'correct' answers that are unrealistic, thus discouraging students from applying real problem solving techniques and instead coercing them to apply arithmetical operations in an incorrect and formulaic manner; an example is the MWP:

John's best time to run 100 meters is 17 seconds. How long will it take him to run 1 kilometer?

where the so-called 'correct' answer, 170 seconds, clearly does not reflect reality (Verschaffel, Greer, and Corte 2000). Ideally, authenticity should be addressed, or at the very least, non-authentic MWPs should be filtered out.

Limitations

The system is yet to be tested within a suitable intelligent tutoring system and with users.

The current prototype is a proof-of-concept demonstrator that has not been extensively tested on a large set of ontologies. Currently it generates only a few types of MWP (additions, subtractions, ratios). Of course, many more types could be added and the natural language generator could be more sophisticated. Currently, there is no handling of units of measurement, for instance. Although some ontologies neglect to model these correctly (see the land area example above), where they do, the system should be extended to generate numerical quantities with units in English.

Conclusions

I have presented a prototype system for generating mathematical word problems (MWPs) from ontologies, a novel application in the question generation community. I have also identified and discussed five key factors that affect the relative difficulty of these problems by drawing on research from education and psychology that was carried out some years ago. This has led to insights as to how these factors might be addressed by simple means within the current prototype but there is much scope for further development, extensions to other kinds of MWP and, in particular, future evaluations with mathematics teachers and students.

Acknowledgements

References

Cetintas, S.; Si, L.; Xin, Y. P.; Zhang, D.; and Park, J. Y. 2009. Automatic text categorization of mathematical word problems. In *Proceedings of the 22nd International FLAIRS Conference*, 27–32.

Cohen, S. A., and Stover, G. 1981. Effects of teaching sixth-grade students to modify format variables of math word. *Reading Research Quarterly* 16(2):175–200.

Coleman, E. 1962. Improving comprehensibility by shortening sentences. *Journal of Applied Psychology* 46:131–134.

Flesch, R. 1949. *The Art of Readable Writing*. Harper, USA.

Mitkov, R., and Ha, L. A. 2003. Computer-aided generation of multiple-choice tests. In *Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing - Volume 2 (HLT-NAACL-EDUC '03)*, 17–22.

Palm, T. 2009. Theory of authentic task situations. In Verschaffel, L.; Greer, B.; Dooren, W. V.; and Mukhopadhyay, S., eds., *Words and Worlds: Modelling Verbal Descriptions of Situations*, 3–19. Sense Publishers.

Papasalouros, A.; Kanaris, K.; and Kotis, K. 2008. Automatic Generation Of Multiple Choice Questions From Domain Ontologies. In Nunes, M. B., and McPherson, M., eds., *e-Learning*, 427–434. IADIS.

Qualification and Curriculum Authority. 1999. Mathematics: the National Curriculum for England. Department for Education and Employment, London.

Stevens, R.; Malone, J.; Williams, S.; and Power, R. 2010. Automating class definitions from OWL to English. In *Proceedings of Bio-Ontologies 2010: Semantic Applications in Life Sciences SIG at the 18th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB 2010)*.

Verschaffel, L.; Greer, B.; and Corte, D. E. 2000. *Making sense of word problems*. Taylor and Francis.

Wheeler, L. J., and McNutt, G. 1983. The effect of syntax on low-achieving students' abilities to solve mathematical word problems. 17(3):309–315.

Williams, S., and Power, R. 2009. Precision and mathematical form in first and subsequent mentions of numerical facts and their relation to document structure. In *Proceedings of the 12th European Workshop on Natural Language Generation*, 118–121.

Williams, S., and Power, R. 2010. Grouping axioms for more coherent ontology descriptions. In *6th International Natural Language Generation Conference (INLG 2010)*, 197–202.

Williams, J.; Clemens, S.; Oleinikova, K.; and Tarvin, K. 2003. The Skills for Life survey: A national needs and impact survey of literacy, numeracy and ICT skills. Technical Report Research Report 490, Department for Education and Skills.