Datasets for Math Word Problem Solving

(Document version: 0.10)

This document introduces the datasets we collected in the SigmaDolphin project [1] for evaluating an automatic math problem solving system. One dataset named *number_word_std* is made publicly available so far. We are preparing more and will release them once they are ready. This document will be updated accordingly when new datasets are released. Latest version of the document is available from the project website [1].

Dataset name	Description
number_word_std	File name: number_word_std.zip (280KB) Containing 1,878 number word problems[1], used as evaluation data in our EMNLP'15 paper [2]. Some subsets are included: Linear, LinearT2, and LinearT6. Each subset is further divided into a dev set and a test set.
…	…

Table 1. Datasets

1. File format

A dataset contains one or multiple subsets. All problems in each subset are stored in a JSON [3] file. Each problem is a JSON object as shown in Figure 1.

```
{
  "id": "algebra.com.117395",
  "index": 1043,
  "text": "one number is 11 more than another number. Find the two numbers if three
times the larger exceeds four times the smaller number by 4.",
  "sources": "algebra.com.117395",
  "equations": [
    "unkn: x,y",
    "equ: x=y+11",
    "equ: 3*x=4*y+4"
  ],
  "ans": "{40;29}",
  "ans_simple": [
    40,
    29
  ]
}
```

Figure 1. JSON format of an example problem

Fields of a problem include:

id: A string used as problem identifier
index: An integer used as problem identifier
text: Problem text

[1] Number word problems are math word problems on numbers.

sources: List of sources from which the problem is obtained. When there is one source, the field value is a JSON string (see Figure 1). If there are multiple sources, the field value is an array of JSON strings (Figure 2).

equations: Manually annotated equations for the problem. The value is an array of JSON strings. Each string is either a list of unknown values (with prefix "unkn:"), or an equation (with prefix "equ:").

ans: Problem gold answers provided by human annotators. Gold answers are used by an evaluation algorithm to automatically determine whether the output of a math problem solving system is correct or not.

ans_simple: Please ignore this field.

```
"sources": [
  "algebra.com.107783",
  "https:\/\/answers.yahoo.com\/question\/index?qid=20071110150449AA3zeMl"
],
```

Figure 2. The "sources" field of problem "algebra.com.107783"

2. Gold answer format and evaluation

Gold answers are used by an evaluation algorithm to automatically determine whether the output of a math problem solving system is correct or not. When the answer to a problem is an integer, the gold answer can have a very simple format. For example the "ans" field of the following problem is "9":

ID: yahoo.answers.20080724143616aaciqtd

Text: Fifteen more than four times a number is 6 more than five times the number. What's the value of the number?

Problem	**Gold Answer**
algebra.com.289589 One positive integer is 3 less than a second positive integer. The sum of the squares of the two integers is 65. Find both positive integers.	4; 7
yahoo.answers.20091207064212aaklfpn The sum of two numbers is twenty-three, and the larger number is five more than the smaller number. Find these numbers.	{9; 14}
yahoo.answers.20070731081228aae3oxl When the reciprocal of 4 times a number is subtracted from 2, the result is twice the reciprocal of the number. Find the number.	9/8 | 1.125
yahoo.answers.20071214233524aamomnu The numerator of a fraction is 5 less than the denominator. If 1 is added to both the numerator and the denominator the fraction would become 2/3. Find the fraction.	9/14
algebra.com.212803 The sum of a number and its square is 6. Find the number	2 or -3
algebra.com.348547 Find two positive numbers so that twice their sum equals their product and one number is 9 times the other number.	{20; 20/9} | {20; 2.222}
algebra.com.141735 The product of the smaller two of three consecutive integers is equal to 23 plus the largest. Find the integers.	5;6;7 or -5;-4;-3

Table 2. Problem gold answer examples

However, in many cases, gold answers should have specific structures to facilitate evaluation. We have the following guidelines for annotators in building the gold answer field:

- ✓ Use ";" to separate the values of different variables
- ✓ Use "or" to separate different answers (in the case that one problem has multiple answers)
- ✓ Use "|" to separate different answer formats
- ✓ Place results in "{}" if any order of the variables is allowed

Some examples are shown in Table 2. Please pay attention that

- ✓ For every decimal value in the gold answer, at most three significant digits after the decimal point are kept. In other words, if there are more than three significant digits after the decimal point of a decimal value, only three significant digits will be kept.
- ✓ When there are no valid answers to a problem (e.g., algebra.com.367618), its answer string is set to be "*ans_no_result*".

Table 3 shows the evaluation results of some example system outputs w.r.t. gold answers. Please use them as test cases to test an evaluation algorithm.

Problems and gold answers	**System output**	**Evaluation**
ID: yahoo.answers.20091207064212aaklfpn Text: "The sum of two numbers is twenty-three, and the larger number is five more than the smaller number. Find these numbers." Gold ans.: {9; 14}	9; 14	Correct
	9; 14 or 14; 9	Correct
	9 or 14	Wrong
	9 and 14	Wrong
	9	Wrong
ID: yahoo.answers.20080928211132aayel0h Text: "Find two numbers so that twice their sum equals their product and one number is 9 times the other number. Enter the smaller number first." Gold ans.: 20/9; 20 or 0; 0 | 2.222; 20 or 0; 0	20/9; 20 or 0; 0	Correct
	2.222; 20 or 0; 0	Correct
	0; 0 or 2.222; 20	Correct
	20; 20/9 or 0; 0	Wrong
	20/9; 20	Wrong
	0; 0	Wrong
ID: algebra.com.348547 Text: "Find two positive numbers so that twice their sum equals their product and one number is 9 times the other number." Gold ans.: {20; 20/9} | {20; 2.222}	20; 20/9	Correct
	20/9; 20	Correct
	2.222; 20	Correct
	20	Wrong
	2.222	Wrong
	20/9	Wrong
ID: yahoo.answers.20071214233524aamomnu Text: "The numerator of a fraction is 5 less than the denominator. If 1 is added to both the numerator and the denominator the fraction would become 2/3. Find the fraction." Gold ans.: 9/14	9/14	Correct
	18/28	Wrong
	0.643	Wrong
ID: algebra.com.367618 Text: "find two consecutive multiples of 7 whose sum is 84" Gold ans.: ans_no_result	ans_no_result	Correct
	38.5; 45.5	Wrong
	<empty-string>	Wrong
	no solution	Wrong

Table 3. Problem gold answer examples

3. More details of the datasets

The dataset number_word_std contains 8 subsets, whose statistic information is shown in Table 4. Here "dev" means the subset is for algorithm development and debugging; "test" means the subset is for evaluation. "Linear" indicates that all problems in the subset correspond to linear equations (i.e., the

"equations" field of each problem contains only linear equations). "T2" means each equation template in the subset corresponds to at least 2 problems. Similarly, "T6" means each equation template in the subset corresponds to at least 6 problems. Please refer to [2] for more details.

Subset		File name	#problems	#sentences (average)	#words (average)
All	dev	number_word_std.dev.json	374	1.79	20.3
	test	number_word_std.test.json	1,504	1.75	22.5
Linear	dev	number_word_std.linear.dev.json	247	1.78	19.6
	test	number_word_std.linear.test.json	986	1.72	19.0
LinearT2	dev	number_word_std.linear_t2.dev.json	172	1.85	18.8
	test	number_word_std.linear_t2.test.json	669	1.71	17.4
LinearT6	dev	number_word_std.linear_t6.dev.json	71	1.96	16.8
	test	number_word_std.linear_t6.test.json	348	1.80	16.1

Table 4. Subsets of the number_word_std dataset

References

[1] Project SigmaDolphin: http://research.microsoft.com/en-us/projects/dolphin/

[2] Shuming Shi, Yuehui Wang, Chin-Yew Lin, Xiaojiang Liu and Yong Rui. 2015. Automatically Solving Number Word Problems by Semantic Parsing and Reasoning. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Lisbon, Portugal.

[3] JSON: http://json.org/