Program Induction by Rationale Generation: Learning to Solve and Explain Algebraic Word Problems

Wang Ling♠   **Dani Yogatama**♠   **Chris Dyer**♠   **Phil Blunsom**♠◇

♠DeepMind   ◇University of Oxford

{lingwang,dyogatama,cdyer,pblunsom}@google.com

Abstract

Solving algebraic word problems requires executing a series of arithmetic operations—a program—to obtain a final answer. However, since programs can be arbitrarily complicated, inducing them directly from question-answer pairs is a formidable challenge. To make this task more feasible, we solve these problems by generating *answer rationales*, sequences of natural language and human-readable mathematical expressions that derive the final answer through a series of small steps. Although rationales do not explicitly specify programs, they provide a scaffolding for their structure via intermediate milestones. To evaluate our approach, we have created a new 100,000-sample dataset of questions, answers and rationales. Experimental results show that indirect supervision of program learning via answer rationales is a promising strategy for inducing arithmetic programs.

1 Introduction

Behaving intelligently often requires mathematical reasoning. Shopkeepers calculate change, tax, and sale prices; agriculturists calculate the proper amounts of fertilizers, pesticides, and water for their crops; and managers analyze productivity. Even determining whether you have enough money to pay for a list of items requires applying addition, multiplication, and comparison. Solving these tasks is challenging as it involves recognizing how goals, entities, and quantities in the real-world map onto a mathematical formalization, computing the solution, and mapping the solution back onto the world. As a proxy for the richness of the real world, a series of papers have used natural language specifications of algebraic word problems, and solved these by either learning to fill in templates that can be solved with equation solvers (Hosseini et al., 2014; Kushman et al., 2014) or inferring and modeling operation sequences (programs) that lead to the final answer (Roy and Roth, 2015).

In this paper, we learn to solve algebraic word problems by inducing and modeling programs that generate not only the answer, but an **answer rationale**, a natural language explanation interspersed with algebraic expressions justifying the overall solution. Such rationales are what examiners require from students in order to demonstrate understanding of the problem solution; they play the very same role in our task. Not only do natural language rationales enhance model interpretability, but they provide a coarse guide to the structure of the arithmetic programs that must be executed. In fact the learner we propose (which relies on a heuristic search; §4) fails to solve this task without modeling the rationales—the search space is too unconstrained.

This work is thus related to models that can explain or rationalize their decisions (Hendricks et al., 2016; Harrison et al., 2017). However, the use of rationales in this work is quite different from the role they play in most prior work, where interpretation models are trained to generate plausible sounding (but not necessarily accurate) post-hoc descriptions of the decision making process they used. In this work, the rationale is generated as a latent variable that gives rise to the answer—it is thus a more faithful representation of the steps used in computing the answer.

This paper makes three contributions. First, we have created a new dataset with more than 100,000 algebraic word problems that includes both answers and natural language answer rationales (§2). Figure 1 illustrates three representative instances

Problem 1:
Question: Two trains running in opposite directions cross a man standing on the platform in 27 seconds and 17 seconds respectively and they cross each other in 23 seconds. The ratio of their speeds is:
Options: A) 3/7   B) 3/2   C) 3/88   D) 3/8   E) 2/2
Rationale: Let the speeds of the two trains be x m/sec and y m/sec respectively. Then, length of the first train = 27x meters, and length of the second train = 17 y meters. (27x + 17y) / (x + y) = 23 → 27x + 17y = 23x + 23y → 4x = 6y → x/y = 3/2.
Correct Option: B

Problem 2:
Question: From a pack of 52 cards, two cards are drawn together at random. What is the probability of both the cards being kings?
Options: A) 2/1223   B) 1/122   C) 1/221   D) 3/1253   E) 2/153
Rationale: Let s be the sample space.
Then n(s) = 52C2 = 1326
E = event of getting 2 kings out of 4
n(E) = 4C2 = 6
P(E) = 6/1326 = 1/221
Answer is C
Correct Option: C

Problem 3:
Question: For which of the following does $p(a) - p(b) = p(ab)$ for all values of a and b?
Options: A) $p(x) = x^2$, B) $p(x) = x/2$, C) $p(x) = x + 5$, D) $p(x) = 2x1$, E) $p(x) = |x|$
Rationale: To solve this easiest way is just put the value and see that if it equals or not.
with option A. $p(a) = a^2$ and $p(b) = b^2$
so L.H.S = $a^2 - b^2$
and R.H.S = $(a - b)^2 \rightarrow a^2 + b^2 - 2ab$.
so L.H.S not equal to R.H.S
with option B. $p(a) = a/2$ and $p(b) = b/2$
L.H.S = $a/2 - b/2 \rightarrow 1/2(a - b)$
R.H.S = $(a - b)/2$
so L.H.S = R.H.S which is the correct answer.
answer:B
Correct Option: B

Figure 1: Examples of solved math problems.

from the dataset. Second, we propose a sequence to sequence model that generates a sequence of instructions that, when executed, generates the rationale; only after this is the answer chosen (§3). Since the target program is not given in the training data (most obviously, its specific form will depend on the operations that are supported by the program interpreter); the third contribution is thus a technique for inferring programs that generate a rationale and, ultimately, the answer. Even constrained by a text rationale, the search space of possible programs is quite large, and we employ a heuristic search to find plausible next steps to guide the search for programs (§4). Empirically, we are able to show that state-of-the-art sequence to sequence models are unable to perform above chance on this task, but that our model doubles the accuracy of the baseline (§6).

2 Dataset

We built a dataset with 100,000 problems with the annotations shown in Figure 1. Each question is decomposed in four parts, two inputs and two outputs: the description of the problem, which we will denote as the **question**, and the possible (multiple choice) answer options, denoted as **options**. Our goal is to generate the description of the rationale used to reach the correct answer, denoted as **rationale** and the **correct option** label. Problem 1 illustrates an example of an algebra problem, which must be translated into an expression (i.e., $(27x + 17y)/(x + y) = 23$) and then the desired quantity (x/y) solved for. Problem 2 is an example that could be solved by multi-step arithmetic operations proposed in (Roy and Roth, 2015). Finally, Problem 3 describes a problem that is solved by testing each of the options, which has not been addressed in the past.

2.1 Construction

We first collect a set of 34,202 seed problems that consist of multiple option math questions covering a broad range of topics and difficulty levels. Examples of exams with such problems include the GMAT (Graduate Management Admission Test) and GRE (General Test). Many websites contain example math questions in such exams, where the answer is supported by a rationale.

Next, we turned to crowdsourcing to generate new questions. We create a task where users are presented with a set of 5 questions from our seed dataset. Then, we ask the Turker to choose one of the questions and write a similar question. We also force the answers and rationale to differ from the original question in order to avoid paraphrases of the original question. Once again, we manually check a subset of the jobs for each Turker for quality control. The type of questions generated using this method vary. Some turkers propose small changes in the values of the questions (e.g., changing the equality $p(a)p(b) = p(ab)$ in Problem 3 to a different equality is a valid question, as long as the rationale and options are rewritten to reflect the change). We designate these as replica problems as the natural language used in the question and rationales tend to be only minimally unaltered. Others propose new problems in the same topic where the generated questions tend to differ more radically from existing ones. Some Turkers also copy math problems available on the web, and we

		Question	Rationale
Training Examples		100,949	
Dev Examples		250	
Test Examples		250	
Numeric	Average Length	9.6	16.6
	Vocab Size	21,009	14,745
Non-Numeric	Average Length	67.8	89.1
	Vocab Size	17,849	25,034
All	Average Length	77.4	105.7
	Vocab Size	38,858	39,779

Table 1: Descriptive statistics of our dataset.

Figure 2: Distribution of examples per length.

define in the instructions that this is not allowed, as it will generate multiple copies of the same problem in the dataset if two or more Turkers copy from the same resource. These Turkers can be detected by checking the nearest neighbours within the collected datasets as problems obtained from online resources are frequently submitted by more than one Turker. Using this method, we obtained 70,318 additional questions.

2.2 Statistics

Descriptive statistics of the dataset is shown in Figure 1. In total, we collected 104,519 problems (34,202 seed problems and 70,318 crowdsourced problems). We removed 500 problems as heldout set (250 for development and 250 for testing). As replicas of the heldout problems may be present in the training set, these were removed manually by listing for each heldout instance the closest problems in the training set in terms of character-based Levenstein distance. After filtering, 100,949 problems remained in the training set.

We also show the average number of tokens (total number of tokens in the question, options and rationale) and the vocabulary size of the questions and rationales. Finally, we provide the same statistics exclusively for tokens that are numeric values and tokens that are not.

Figure 2 shows the distribution of examples based on the total number of tokens. We can see that most examples consist of 30 to 500 tokens, but there are also extremely long examples with more than 1000 tokens in our dataset.

3 Model

Generating rationales for math problems is challenging as it requires models that learn to perform math operations at a finer granularity as each step within the solution must be explained. For instance, in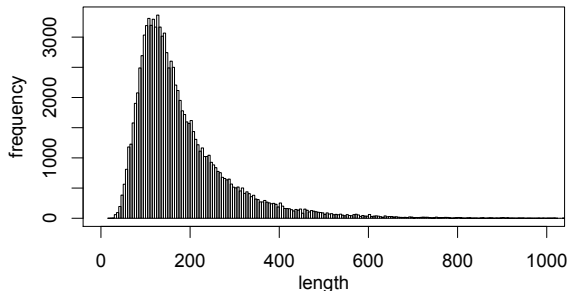 Problem 1, the equation $(27x +$ $17y)/(x + y) = 23$ must be solved to obtain the answer. In previous work (Kushman et al., 2014), this could be done by feeding the equation into an expression solver to obtain $x/y = 3/2$. However, this would skip the intermediate steps $27x + 17y = 23x + 23y$ and $4x = 6y$, which must also be generated in our problem. We propose a model that jointly learns to generate the text in the rationale, and to perform the math operations required to solve the problem. This is done by generating a program, containing both instructions that generate output and instructions that simply generate intermediate values used by following instructions.

3.1 Problem Definition

In traditional sequence to sequence models (Sutskever et al., 2014; Bahdanau et al., 2014), the goal is to predict the output sequence $\boldsymbol{y} = y_1, \ldots, y_{|\boldsymbol{y}|}$ from the input sequence $\boldsymbol{x} = x_1, \ldots, x_{|\boldsymbol{x}|}$, with lengths $|\boldsymbol{y}|$ and $|\boldsymbol{x}|$.

In our particular problem, we are given the problem and the set of options, and wish to predict the rationale and the correct option. We set $\boldsymbol{x}$ as the sequence of words in the problem, concatenated with words in each of the options separated by a special tag. Note that knowledge about the possible options is required as some problems are solved by the process of elimination or by testing each of the options (e.g. Problem 3). We wish to generate $\boldsymbol{y}$, which is the sequence of words in the rationale. We also append the correct option as the last word in $\boldsymbol{y}$, which is interpreted as the chosen option. For example, $\boldsymbol{y}$ in Problem 1 is "Let the . . . = 3/2 . ⟨EOR⟩ B ⟨EOS⟩", whereas in Problem 2 it is "Let s be . . . Answer is C ⟨EOR⟩ C ⟨EOS⟩", where "⟨EOS⟩" is the end of sentence symbol and "⟨EOR⟩" is the end of rationale symbol.

i	x	z	v	r
1	From	Id("Let")	*Let*	y_1
2	a	Id("s")	*s*	y_2
3	pack	Id("be")	*be*	y_3
4	of	Id("the")	*the*	y_4
5	52	Id("sample")	*sample*	y_5
6	cards	Id("space")	*space*	y_6
7	,	Id(".")	*.*	y_7
8	two	Id("\n")	*\n*	y_8
9	cards	Id("Then")	*Then*	y_9
10	are	Id("n")	*n*	y_{10}
11	drawn	Id("(")	*(*	y_{11}
12	together	Id("s")	*s*	y_{12}
13	at	Id(")")	*)*	y_{13}
14	random	Id("=")	*=*	y_{14}
15	.	Str_to_Float(x_5)	**52**	m_1
16	What	Float_to_Str(m_1)	*52*	y_{15}
17	is	Id("C")	*C*	y_{16}
18	the	Id("2")	*2*	y_{17}
19	probability	Id("=")	*=*	y_{18}
20	of	Str_to_Float(y_{17})	**2**	m_2
21	both	Choose(m_1,m_2)	**1326**	m_3
22	cards	Float_to_Str(m_3)	*1326*	y_{19}
23	being	Id("E")	*E*	y_{20}
24	kings	Id("=")	*=*	y_{21}
25	?	Id("event")	*event*	y_{22}
26	<O>	Id("of")	*of*	y_{23}
27	A)	Id("getting")	*getting*	y_{24}
28	2/1223	Id("2")	*2*	y_{25}
29	<O>	Id("kings")	*kings*	y_{26}
30	B)	Id("out")	*out*	y_{27}
31	1/122	Id("of")	*of*	y_{28}
...	...	...	...	...
$\lvert z \rvert$		Id("⟨EOS⟩")	*⟨EOS⟩*	$y_{\lvert y \rvert}$

Table 2: Example of a program z that would generate the output y. In v, *italics* indicates string types; **bold** indicates float types. Refer to §3.3 for description of variable names.

3.2 Generating Programs to Generate Rationales

We wish to generate a latent sequence of **program instructions**, $z = z_1, \ldots, z_{\lvert z \rvert}$, with length $\lvert z \rvert$, that will generate y when executed.

We express z as a program that can access x, y, and the memory buffer m. Upon finishing execution we expect that the sequence of output tokens to be placed in the output vector y.

Table 2 illustrates an example of a sequence of instructions that would generate an excerpt from Problem 2, where columns x, z, v, and r denote the input sequence, the instruction sequence (program), the values of executing the instruction, and where each value v_i is written (i.e., either to the output or to the memory). In this example, instructions from indexes 1 to 14 simply fill each position with the observed output $y_1, \ldots, y_{14}$ with a string, where the Id operation simply returns its parame-

ter without applying any operation. As such, running this operation is analogous to generating a word by sampling from a softmax over a vocabulary. However, instruction z_{15} reads the input word x_5, 52, and applies the operation Str_to_Float, which converts the word 52 into a floating point number, and the same is done for instruction z_{20}, which reads a previously generated output word y_{17}. Unlike, instructions $z_1, \ldots, z_{14}$, these operations write to the external memory m, which stores intermediate values. A more sophisticated instruction—which shows some of the power of our model—is $z_{21} = \text{Choose}(m_1, m_2) \to m_3$ which evaluates $\binom{m_1}{m_2}$ and stores the result in m_3. This process repeats until the model generates the end-of-sentence symbol. The last token of the program as said previously must generate the correct option value, from "A" to "E".

By training a model to generate instructions that can manipulate existing tokens, the model benefits from the additional expressiveness needed to solve math problems within the generation process. In total we define 22 different operations, 13 of which are frequently used operations when solving math problems. These are: Id, Add, Subtract, Multiply, Divide, Power, Log, Sqrt, Sine, Cosine, Tangent, Factorial, and Choose (number of combinations). We also provide 2 operations to convert between Radians and Degrees, as these are needed for the sine, cosine and tangent operations. There are 6 operations that convert floating point numbers into strings and vice-versa. These include the Str_to_Float and Float_to_Str operations described previously, as well as operations which convert between floating point numbers and fractions, since in many math problems the answers are in the form "3/4". For the same reason, an operation to convert between a floating point number and number grouped in thousands is also used (e.g. 1000000 to "1,000,000" or "1.000.000"). Finally, we define an operation (Check) that given the input string, searches through the list of options and returns a string with the option index in {"A", "B", "C", "D", "E"}. If the input value does not match any of the options, or more than one option contains that value, it cannot be applied. For instance, in Problem 2, once the correct probability "1/221" is generated, by applying the check operation to this number we can obtain correct option "C".
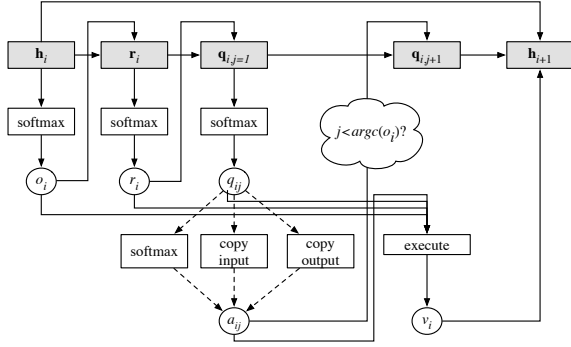
Figure 3: Illustration of the generation process of a single instruction tuple at timestamp i.

3.3 Generating and Executing Instructions

In our model, programs consist of sequences of instructions, z. We turn now to how we model each z_i, conditional on the text program specification, and the program's history. The instruction z_i is a tuple consisting of an operation (o_i), an ordered sequence of its arguments ($\boldsymbol{a}_i$), and a decision about where its results will be placed (r_i) (is it appended in the output $\boldsymbol{y}$ or in a memory buffer $\boldsymbol{m}$?), and the result of applying the operation to its arguments (v_i). That is, $z_i = (o_i, r_i, \boldsymbol{a}_i, v_i)$.

Formally, o_i is an element of the pre-specified set of operations $\mathcal{O}$, which contains, for example add, div, Str_to_Float, etc. The number of arguments required by o_i is given by $argc(o_i)$, e.g., $argc(\text{add}) = 2$ and $argc(\text{log}) = 1$. The arguments are $\boldsymbol{a}_i = a_{i,1}, \ldots, a_{i,argc(o_i)}$. An instruction will generate a return value v_i upon execution, which will either be placed in the output $\boldsymbol{y}$ or hidden. This decision is controlled by r_i. We define the instruction probability as:

$$p(o_i, \boldsymbol{a}_i, r_i, v_i \mid \boldsymbol{z}_{<i}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{m}) =$$
$$p(o_i \mid \boldsymbol{z}_{<i}, \boldsymbol{x}) \times p(r_i \mid \boldsymbol{z}_{<i}, \boldsymbol{x}, o_i) \times$$
$$\prod_{j=1}^{argc(o_i)} p(a_{i,j} \mid \boldsymbol{z}_{<i}, \boldsymbol{x}, o_i, \boldsymbol{m}, \boldsymbol{y}) \times$$
$$[v_i = \text{apply}(o_i, \boldsymbol{a})],$$

where $[p]$ evaluates to 1 if p is true and 0 otherwise, and apply($f, \boldsymbol{x}$) evaluates the operation f with arguments $\boldsymbol{x}$. Note that the apply function is not learned, but pre-defined.

The network used to generate an instruction at a given timestamp i is illustrated in Figure 3. We first use the recurrent state $\mathbf{h}_i$ to generate $p(o_i \mid \boldsymbol{z}_{<i}, \boldsymbol{x}) = \text{softmax}_{o_i \in \mathcal{O}}(\mathbf{h}_i)$, using a softmax over the set of available operations $\mathcal{O}$.

In order to predict r_i, we generate a new hidden state $\mathbf{r}_i$, which is a function of the current program context $\mathbf{h}_i$, and an embedding of the current predicted operation, o_i. As the output can either be placed in the memory $\boldsymbol{m}$ or the output $\boldsymbol{y}$, we compute the probability $p(r_i = \text{OUTPUT} \mid \boldsymbol{z}_{<i}, \boldsymbol{x}, o_i) = \sigma(\mathbf{r}_i \cdot \mathbf{w}_r + b_r)$, where σ is the logistic sigmoid function. If $r_i = \text{OUTPUT}$, v_i is appended to the output $\boldsymbol{y}$; otherwise it is appended to the memory $\boldsymbol{m}$.

Once we generate r_i, we must predict $\boldsymbol{a}_i$, the $argc(o_i)$-length sequence of arguments that operation o_i requires. The jth argument $a_{i,j}$ can be either generated from a softmax over the vocabulary, copied from the input vector $\boldsymbol{x}$, or copied from previously generated values in the output $\boldsymbol{y}$ or memory $\boldsymbol{m}$. This decision is modeled using a latent predictor network (Ling et al., 2016), where the control over which method used to generate $a_{i,j}$ is governed by a latent variable $q_{i,j} \in \{\text{SOFTMAX}, \text{COPY-INPUT}, \text{COPY-OUTPUT}\}$. Similar to when predicting r_i, in order to make this choice, we also generate a new hidden state for each argument slot j, denoted by $\mathbf{q}_{i,j}$ with an LSTM, feeding the previous argument in at each time step, and initializing it with $\mathbf{r}_i$ and by reading the predicted value of the output r_i.

- If $q_{i,j} = \text{SOFTMAX}$, $a_{i,j}$ is generated by sampling from a softmax over the vocabulary $\mathcal{Y}$,

$$p(a_{i,j} \mid q_{i,j} = \text{SOFTMAX}) = \text{softmax}_{a_{i,j} \in \mathcal{Y}}(\mathbf{q}_{i,j}).$$

This corresponds to a case where a string is used as argument (e.g. y_1="Let").

- If $q_{i,j} = \text{COPY-INPUT}$, $a_{i,j}$ is obtained by copying an element from the input vector with a pointer network (Vinyals et al., 2015) over input words $x_1, \ldots, x_{|\boldsymbol{x}|}$, represented by their encoder LSTM state $\mathbf{u}_1, \ldots, \mathbf{u}_{|\boldsymbol{x}|}$. As such, we compute the probability distribution over input words as:

$$p(a_{i,j} \mid q_{i,j} = \text{COPY-INPUT}) = \qquad (1)$$
$$\text{softmax}_{a_{i,j} \in x_1, \ldots, x_{|\boldsymbol{x}|}} \left(f(\mathbf{u}_{a_{i,j}}, \mathbf{q}_{i,j}) \right)$$

Function f computes the affinity of each token $x_{a_{i,j}}$ and the current output context $\mathbf{q}_{i,j}$. A common implementation of f, which we follow, is to apply a linear projection from $[\mathbf{u}_{a_{i,j}}; \mathbf{q}_{i,j}]$ into a fixed size vector (where $[\mathbf{u}; \mathbf{v}]$ is vector concatenation), followed by a tanh and a linear projection into a single value.

162

- If $q_{i,j} = $ COPY-OUTPUT, the model copies from either the output $\boldsymbol{y}$ or the memory $\boldsymbol{m}$. This is equivalent to finding the instruction z_i, where the value was generated. Once again, we define a pointer network that points to the output instructions and define the distribution over previously generated instructions as:

$$p(a_{i,j} \mid q_{i,j} = \text{COPY-OUTPUT}) = \\ \operatorname*{softmax}_{a_{i,j} \in z_1, \ldots, z_{i-1}} \big(f(\mathbf{h}_{a_{i,j}}, \mathbf{q}_{i,j})\big)$$

Here, the affinity is computed using the decoder state $\mathbf{h}_{a_{i,j}}$ and the current state $\mathbf{q}_{i,j}$.

Finally, we embed the argument $a_{i,j}$[1] and the state $\mathbf{q}_{i,j}$ to generate the next state $\mathbf{q}_{i,j+1}$. Once all arguments for o_i are generated, the operation is executed to obtain v_i. Then, the embedding of v_i, the final state of the instruction $\mathbf{q}_{i,|\boldsymbol{a}_i|}$ and the previous state $\mathbf{h}_i$ are used to generate the state at the next timestamp $\mathbf{h}_{i+1}$.

4 Inducing Programs while Learning

The set of instructions $\boldsymbol{z}$ that will generate $\boldsymbol{y}$ is unobserved. Thus, given $\boldsymbol{x}$ we optimize the marginal probability function:

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = \sum_{\boldsymbol{z} \in \mathcal{Z}} p(\boldsymbol{y} \mid \boldsymbol{z}) p(\boldsymbol{z} \mid \boldsymbol{x}) = \sum_{\boldsymbol{z} \in \mathcal{Z}(y)} p(\boldsymbol{z} \mid \boldsymbol{x}),$$

where $p(\boldsymbol{y} \mid \boldsymbol{z})$ is the Kronecker delta function $\delta_{e(\boldsymbol{z}), \boldsymbol{y}}$, which is 1 if the execution of $\boldsymbol{z}$, denoted as $e(\boldsymbol{z})$, generates y and 0 otherwise. Thus, we can redefine $p(\boldsymbol{y}|\boldsymbol{x})$, the marginal over all programs $\mathcal{Z}$, as a marginal over programs that would generate $\boldsymbol{y}$, defined as $\mathcal{Z}(\boldsymbol{y})$. As marginalizing over $\boldsymbol{z} \in \mathcal{Z}(\boldsymbol{y})$ is intractable, we approximate the marginal by generating samples from our model. Denote the set of samples that are generated by $\hat{\mathcal{Z}}(\boldsymbol{y})$. We maximize $\sum \boldsymbol{z} \in \hat{\mathcal{Z}}(\boldsymbol{y}) p(\boldsymbol{z}|\boldsymbol{x})$.

However, generating programs that generate $\boldsymbol{y}$ is not trivial, as randomly sampling from the RNN distribution over instructions at each timestamp is unlikely to generate a sequence $\boldsymbol{z} \in \mathcal{Z}(\boldsymbol{y})$.

This is analogous to the question answering work in Liang et al. (2016), where the query that

---

[1] The embeddings of a given argument $a_{i,j}$ and the return value v_i are obtained with a lookup table embedding and two flags indicating whether it is a string and whether it is a float. Furthermore, if the the value is a float we also add its numeric value as a feature.

generates the correct answer must be found during inference, and training proved to be difficult without supervision. In Roy and Roth (2015) this problem is also addressed by adding prior knowledge to constrain the exponential space.

In our work, we leverage the fact that we are generating rationales, where there is a sense of progression within the rationale. That is, we assume that the rationale solves the problem step by step. For instance, in Problem 2, the rationale first describes the number of combinations of two cards in a deck of 52 cards, then describes the number of combinations of two kings, and finally computes the probability of drawing two kings. Thus, while generating the final answer without the rationale requires a long sequence of latent instructions, generating each of the tokens of the rationale requires far less operations.

More formally, given the sequence $z_1, \ldots, z_{i-1}$ generated so far, and the possible values for z_i given by the network, denoted $\mathcal{Z}_i$, we wish to filter $\mathcal{Z}_i$ to $\mathcal{Z}_i(y_k)$, which denotes a set of possible options that contain at least one path capable of generating the next token at index k. Finding the set $\mathcal{Z}_i(y_k)$ is achieved by testing all combinations of instructions that are possible with at most one level of indirection, and keeping those that can generate y_k. This means that the model can only generate one intermediate value in memory (not including the operations that convert strings into floating point values and vice-versa).

Decoding. During decoding we find the most likely sequence of instructions $\boldsymbol{z}$ given $\boldsymbol{x}$, which can be performed with a stack-based decoder. However, it is important to refer that each generated instruction $z_i = (o_i, r_i, a_{i,1}, \ldots, a_{i,|a_i|}, v_i)$ must be executed to obtain v_i. To avoid generating unexecutable code—e.g., log(0)—each hypothesis instruction is executed and removed if an error occurs. Finally, once the "$\langle \text{EOR} \rangle$" tag is generated, we only allow instructions that would generate one of the option "A" to "E" to be generated, which guarantees that one of the options is chosen.

5 Staged Back-propagation

As it is shown in Figure 2, math rationales with more than 200 tokens are not uncommon, and with additional intermediate instructions, the size $\boldsymbol{z}$ can easily exceed 400. This poses a practical challenge for training the model.

For both the attention and copy mechanisms,

for each instruction z_i, the model needs to compute the probability distribution between all the attendable units c conditioned on the previous state $\mathbf{h}_{i-1}$. For the attention model and input copy mechanisms, $c = \mathbf{x}_{0,i-1}$ and for the output copy mechanism $c = \mathbf{z}$. These operations generally involve an exponential number of matrix multiplications as the size of c and z grows. For instance, during the computation of the probabilities for the input copy mechanism in Equation 1, the affinity function f between the current context $\mathbf{q}$ and a given input $\mathbf{u}_k$ is generally implemented by projecting $\mathbf{u}$ and $\mathbf{q}$ into a single vector followed by a non-linearity, which is projected into a single affinity value. Thus, for each possible input $\mathbf{u}$, 3 matrix multiplications must be performed. Furthermore, for RNN unrolling, parameters and intermediate outputs for these operations must be replicated for each timestamp. Thus, as $\mathbf{z}$ becomes larger the attention and copy mechanisms quickly become a memory bottleneck as the computation graph becomes too large to fit on the GPU. In contrast, the sequence-to-sequence model proposed in (Sutskever et al., 2014), does not suffer from these issues as each timestamp is dependent only on the previous state $\mathbf{h}_{i-1}$.

To deal with this, we use a training method we call **staged back-propagation** which saves memory by considering slices of K tokens in $\mathbf{z}$, rather than the full sequence. That is, to train on a mini-batch where $|z| = 300$ with $K = 100$, we would actually train on 3 mini-batches, where the first batch would optimize for the first $\mathbf{z}_{1:100}$, the second for $\mathbf{z}_{101:200}$ and the third for $\mathbf{z}_{201:300}$. The advantage of this method is that memory intensive operations, such as attention and the copy mechanism, only need to be unrolled for K steps, and K can be adjusted so that the computation graph fits in memory.

However, unlike truncated back-propagation for language modeling, where context outside the scope of K is ignored, sequence-to-sequence models require global context. Thus, the sequence of states $\mathbf{h}$ is still built for the whole sequence $\mathbf{z}$. Afterwards, we obtain a slice $\mathbf{h}_{j:j+K}$, and compute the attention vector.[2] Finally, the prediction of the instruction is conditioned on the LSTM state and the attention vector.

---

[2]This modeling strategy is sometimes known as late fusion, as the attention vector is not used for state propagation, it is incorporated "later".

6   Experiments

We apply our model to the task of generating rationales for solutions to math problems, evaluating it on both the quality of the rationale and the ability of the model to obtain correct answers.

6.1   Baselines

As the baseline we use the attention-based sequence to sequence model proposed by Bahdanau et al. (2014), and proposed augmentations, allowing it to copy from the input (Ling et al., 2016) and from the output (Merity et al., 2016).

6.2   Hyperparameters

We used a two-layer LSTM with a hidden size of $H = 200$, and word embeddings with size 200. The number of levels that the graph $\mathcal{G}$ is expanded during sampling D is set to 5. Decoding is performed with a beam of 200. As for the vocabulary of the softmax and embeddings, we keep the most frequent 20,000 word types, and replace the rest of the words with an unknown token. During training, the model only learns to predict a word as an unknown token, when there is no other alternative to generate the word.

6.3   Evaluation Metrics

The evaluation of the rationales is performed with average sentence level perplexity and BLEU-4 (Papineni et al., 2002). When a model cannot generate a token for perplexity computation, we predict unknown token. This benefits the baselines as they are less expressive. As the perplexity of our model is dependent on the latent program that is generated, we force decode our model to generate the rationale, while maximizing the probability of the program. This is analogous to the method used to obtain sample programs described in Section 4, but we choose the most likely instructions at each timestamp instead of sampling. Finally, the correctness of the answer is evaluated by computing the percentage of the questions, where the chosen option matches the correct one.

6.4   Results

The test set results, evaluated on perplexity, BLEU, and accuracy, are presented in Table 3.

Perplexity. In terms of perplexity, we observe that the regular sequence to sequence model fares poorly on this dataset, as the model requires the generation of many values that tend to be

Model	Perplexity	BLEU	Accuracy
Seq2Seq	524.7	8.57	20.8
+Copy Input	46.8	21.3	20.4
+Copy Output	45.9	20.6	20.2
Our Model	**28.5**	**27.2**	**36.4**

Table 3: Results over the test set measured in Perplexity, BLEU and Accuracy.

sparse. Adding an input copy mechanism greatly improves the perplexity as it allows the generation process to use values that were mentioned in the question. The output copying mechanism improves perplexity slightly over the input copy mechanism, as many values are repeated after their first occurrence. For instance, in Problem 2, the value "1326" is used twice, so even though the model cannot generate it easily in the first occurrence, the second one can simply be generated by copying the first one. We can observe that our model yields significant improvements over the baselines, demonstrating that the ability to generate new values by algebraic manipulation is essential in this task. An example of a program that is inferred is shown in Figure 4. The graph was generated by finding the most likely program z that generates y. Each node isolates a value in x, m, or y, where arrows indicate an operation executed with the outgoing nodes as arguments and incoming node as the return of the operation. For simplicity, operations that copy or convert values (e.g. from string to float) were not included, but nodes that were copied/converted share the same color. Examples of tokens where our model can obtain the perplexity reduction are the values "0.025", "0.023", "0.002" and finally the answer "E" , as these cannot be copied from the input or output.

BLEU. We observe that the regular sequence to sequence model achieves a low BLEU score. In fact, due to the high perplexities the model generates very short rationales, which frequently consist of segments similar to "Answer should be D", as most rationales end with similar statements. By applying the copy mechanism the BLEU score improves substantially, as the model can define the variables that are used in the rationale. Interestingly, the output copy mechanism adds no further improvement in the perplexity evaluation. This is because during decoding all values that can be copied from the output are values that could

have been generated by the model either from the softmax or the input copy mechanism. As such, adding an output copying mechanism adds little to the expressiveness of the model during decoding.

Finally, our model can achieve the highest BLEU score as it has the mechanism to generate the intermediate and final values in the rationale.

Accuracy. In terms of accuracy, we see that all baseline models obtain values close to chance (20%), indicating that they are completely unable to solve the problem. In contrast, we see that our model can solve problems at a rate that is significantly higher than chance, demonstrating the value of our program-driven approach, and its ability to learn to generate programs.

In general, the problems we solve correctly correspond to simple problems that can be solved in one or two operations. Examples include questions such as "Billy cut up each cake into 10 slices, and ended up with 120 slices altogether. How many cakes did she cut up? A) 9 B) 7 C) 12 D) 14 E) 16", which can be solved in a single step. In this case, our model predicts "120 / 10 = 12 cakes. Answer is C" as the rationale, which is reasonable.

6.5 Discussion.

While we show that our model can outperform the models built up to date, generating complex rationales as those shown in Figure 1 correctly is still an unsolved problem, as each additional step adds complexity to the problem both during inference and decoding. Yet, this is the first result showing that it is possible to solve math problems in such a manner, and we believe this modeling approach and dataset will drive work on this problem.

7 Related Work

Extensive efforts have been made in the domain of math problem solving (Hosseini et al., 2014; Kushman et al., 2014; Roy and Roth, 2015), which aim at obtaining the correct answer to a given math problem. Other work has focused on learning to map math expressions into formal languages (Roy et al., 2016). We aim to generate natural language rationales, where the bindings between variables and the problem solving approach are mixed into a single generative model that attempts to solve the problem while explaining the approach taken.

Our approach is strongly tied with the work on sequence to sequence transduction using the encoder-decoder paradigm (Sutskever et al., 2014;
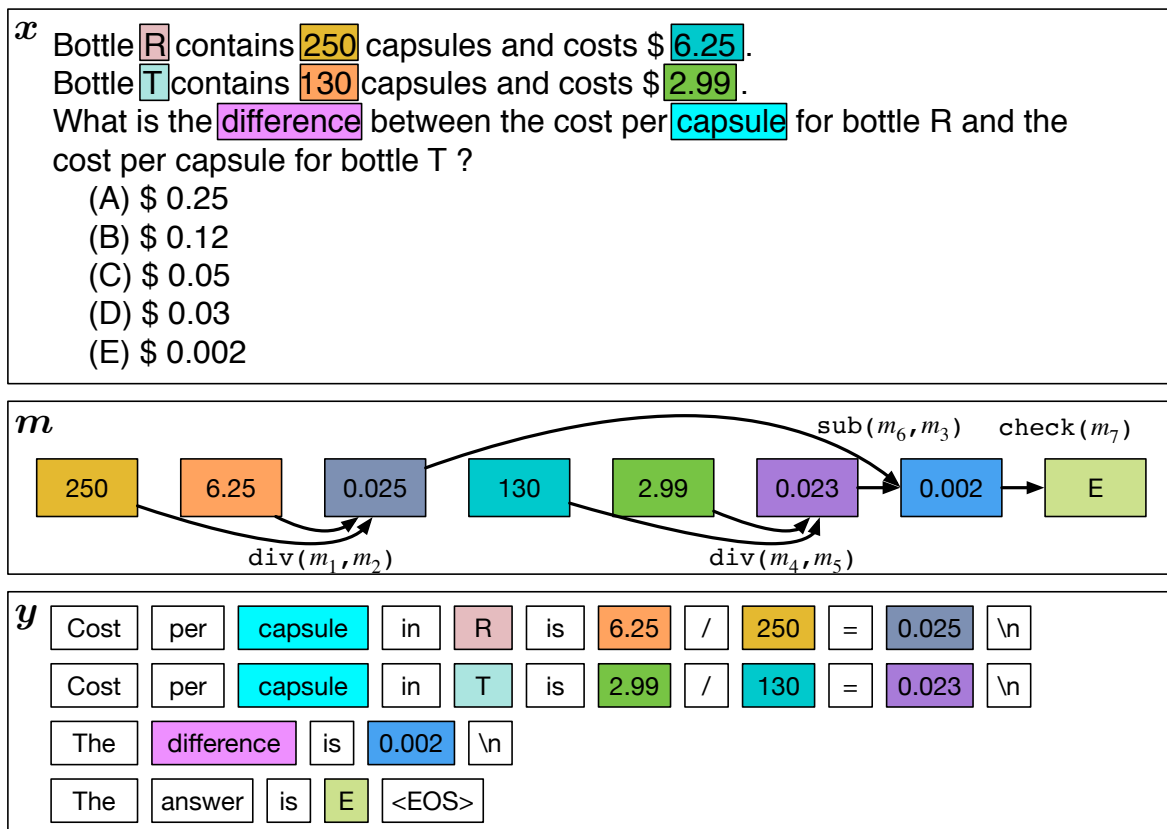
Figure 4: Illustration of the most likely latent program inferred by our algorithm to explain a held-out question-rationale pair.

Bahdanau et al., 2014; Kalchbrenner and Blunsom, 2013), and inherits ideas from the extensive literature on semantic parsing (Jones et al., 2012; Berant et al., 2013; Andreas et al., 2013; Quirk et al., 2015; Liang et al., 2016; Neelakantan et al., 2016) and program generation (Reed and de Freitas, 2016; Graves et al., 2016), namely, the usage of an external memory, the application of different operators over values in the memory and the copying of stored values into the output sequence.

Providing textual explanations for classification decisions has begun to receive attention, as part of increased interest in creating models whose decisions can be interpreted. Lei et al. (2016), jointly modeled both a classification decision, and the selection of the most relevant subsection of a document for making the classification decision. Hendricks et al. (2016) generate textual explanations for visual classification problems, but in contrast to our model, they first generate an answer, and then, conditional on the answer, generate an explanation. This effectively creates a post-hoc justification for a classification decision rather than a program for deducing an answer. These papers,

like ours, have jointly modeled rationales and answer predictions; however, we are the first to use rationales to guide program induction.

8 Conclusion

In this work, we addressed the problem of generating rationales for math problems, where the task is to not only obtain the correct answer of the problem, but also generate a description of the method used to solve the problem. To this end, we collect 100,000 question and rationale pairs, and propose a model that can generate natural language and perform arithmetic operations in the same decoding process. Experiments show that our method outperforms existing neural models, in both the fluency of the rationales that are generated and the ability to solve the problem.

References

Jacob Andreas, Andreas Vlachos, and Stephen Clark. 2013. Semantic parsing as machine translation. In *Proc. of ACL.*

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Ben-

gio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv* 1409.0473.

Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. Semantic parsing on freebase from question-answer pairs. In *Proc. of EMNLP*.

Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwiska, Sergio Gmez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, Adri Puigdomnech Badia, Karl Moritz Hermann, Yori Zwols, Georg Ostrovski, Adam Cain, Helen King, Christopher Summerfield, Phil Blunsom, Koray Kavukcuoglu, and Demis Hassabis. 2016. Hybrid computing using a neural network with dynamic external memory. *Nature* 538(7626):471–476.

Brent Harrison, Upol Ehsan, and Mark O. Riedl. 2017. Rationalization: A neural machine translation approach to generating natural language explanations. *CoRR* abs/1702.07826.

Lisa Anne Hendricks, Zeynep Akata, Marcus Rohrbach, Jeff Donahue, Bernt Schiele, and Trevor Darrell. 2016. Generating visual explanations. In *Proc. ECCV*.

Mohammad Javad Hosseini, Hannaneh Hajishirzi, Oren Etzioni, and Nate Kushman. 2014. Learning to solve arithmetic word problems with verb categorization. In *Proc. of EMNLP*.

Bevan Keeley Jones, Mark Johnson, and Sharon Goldwater. 2012. Semantic parsing with bayesian tree transducers. In *Proc. of ACL*.

Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent continuous translation models. In *Proc. of EMNLP*.

Nate Kushman, Yoav Artzi, Luke Zettlemoyer, and Regina Barzilay. 2014. Learning to automatically solve algebra word problems. In *Proc. of ACL*.

Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2016. Rationalizing neural predictions. In *Proc. of EMNLP*.

Chen Liang, Jonathan Berant, Quoc Le, Kenneth D. Forbus, and Ni Lao. 2016. Neural symbolic machines: Learning semantic parsers on freebase with weak supervision. *arXiv* 1611.00020.

Wang Ling, Edward Grefenstette, Karl Moritz Hermann, Tomás Kociský, Andrew Senior, Fumin Wang, and Phil Blunsom. 2016. Latent predictor networks for code generation. In *Proc. of ACL*.

Stephen Merity, Caiming Xiong, James Bradbury, and Richard Socher. 2016. Pointer sentinel mixture models. *arXiv* 1609.07843.

Arvind Neelakantan, Quoc V. Le, and Ilya Sutskever. 2016. Neural programmer: Inducing latent programs with gradient descent. In *Proc. ICLR*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: A method for automatic evaluation of machine translation. In *Proc. of ACL*.

Chris Quirk, Raymond Mooney, and Michel Galley. 2015. Language to code: Learning semantic parsers for if-this-then-that recipes. In *Proc. of ACL*.

Scott E. Reed and Nando de Freitas. 2016. Neural programmer-interpreters. In *Proc. of ICLR*.

Subhro Roy and Dan Roth. 2015. Solving general arithmetic word problems. In *Proc. of EMNLP*.

Subhro Roy, Shyam Upadhyay, and Dan Roth. 2016. Equation parsing: Mapping sentences to grounded equations. In *Proc. of EMNLP*.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. *arXiv* 1409.3215.

Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In *Proc. of NIPS*.