Neural Mathematical Solver with Enhanced Formula Structure

Zhenya Huang[1], Qi Liu[1], Weibo Gao[1], Jinze Wu[1], Yu Yin[1], Hao Wang[1], Enhong Chen[1,*]
[1]Anhui Province Key Laboratory of Big Data Analysis and Application, School of Computer Science and Technology, University of Science and Technology of China,
{huangzhy,weibogao,hxwjz,yxonic,wanghao3}@mail.ustc.edu.cn; {qiliuql,cheneh}@ustc.edu.cn

Reporter: Weibo Gao

Outline

Background

- Automatically answering math problems
  - A crucial and challenging task in AI
  - Requirements
    - **Linguistic** understanding ability
      - Semantic understanding
      - Operator extraction
    - **Mathematical** comprehension ability
      - Understand formulas with free-text format

Problem: Robin was making baggies of cookies with 6 cookies in each bag. If she had 23 chocolate cookies and 25 oatmeal cookies, how many baggies could she make?
Expression: $x = (23 + 25) \div 6$
Answer: 8

Question: Let $f(x) = -x^3 - x^2$. Let $g(x) = -2x$. solve $f(g(x))$
Answer: $8x^3 - 4x^2$

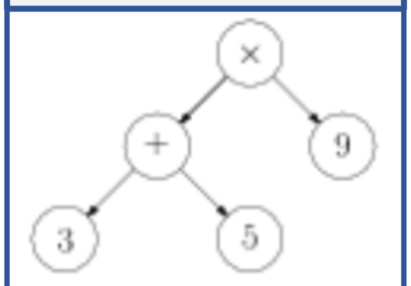
Related work

- Math word problem
  - Elementary problem (primary school level)
  - Translate questions text into expression forms for answers
  - Existing methods
    - Rules-schemes-matching methods
    - Statistical learning
      - E.g., template-based, tree-based
    - Seq2seq deep learning

Just consist of natural language content

Math word problem	expression
Gwen was organizing her book case making sure each of the shelves had exactly 9 books on it. She has 2 types of books - mystery books and picture books. If she had 3 shelves of mystery books and 5 shelves of picture books. How many books did she have total?	$(3+5)\times 9 = 72$

expression tree



Background

- Math word Problem
  - Elementary problem (primary school level)
    - Linguistic learning for natural language content
    - Operator extraction (+)
    - Semantic understanding
- Mathematical problem
  - Complex problem (high school level)
    - Language content
    - Specific but informative **formulas**
  - Requirement
    - Linguistic understanding
    - **Mathematical comprehension**

Problem: Robin was making baggies of cookies with 6 cookies in each bag. If she had 23 chocolate cookies and 25 oatmeal cookies, how many baggies could she make?
Expression: $x = (23 + 25) \div 6$
Answer: 8

Math word problem

Question: Let $f(x) = -x^3 - x^2$. Let $g(x) = -2x$. solve $f(g(x))$
Answer: $8x^3 - 4x^2$

Mathematical problem

Background

- **Challenges:** How to represent formula-enriched problem?
  - How to to understand formulas with their free-text format?

$\sin \sqrt{x}/_2$

sin, √, x, /, 2 word-level

s, i, n, √, x, /, 2 character-level

\sin, \sqrt, \frac ?

  - How to design a unified architecture to incorporate **linguistic** and **mathematical** information?

Question: Let $f(x) = -x^3 - x^2$. Let $g(x) = -2x$. solve $f(g(x))$
Answer: $8x^3 - 4x^2$

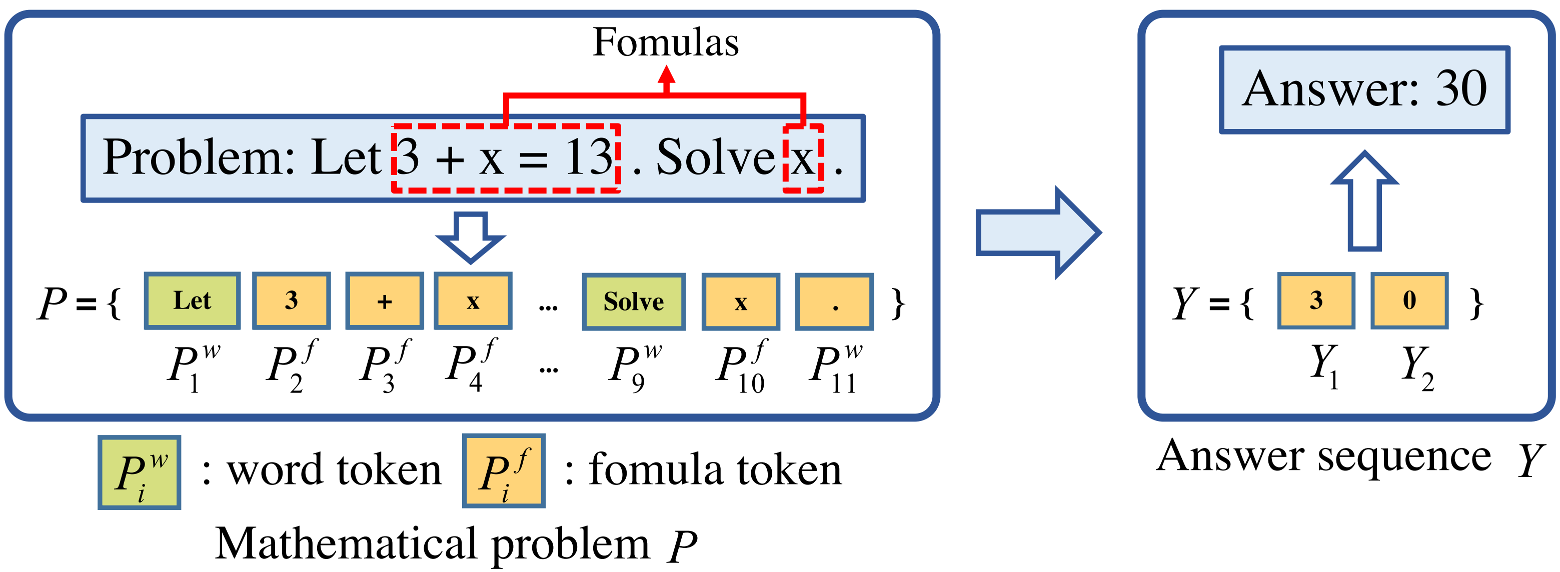
Mathematical information

Linguistic information

Outline

Problem Definition

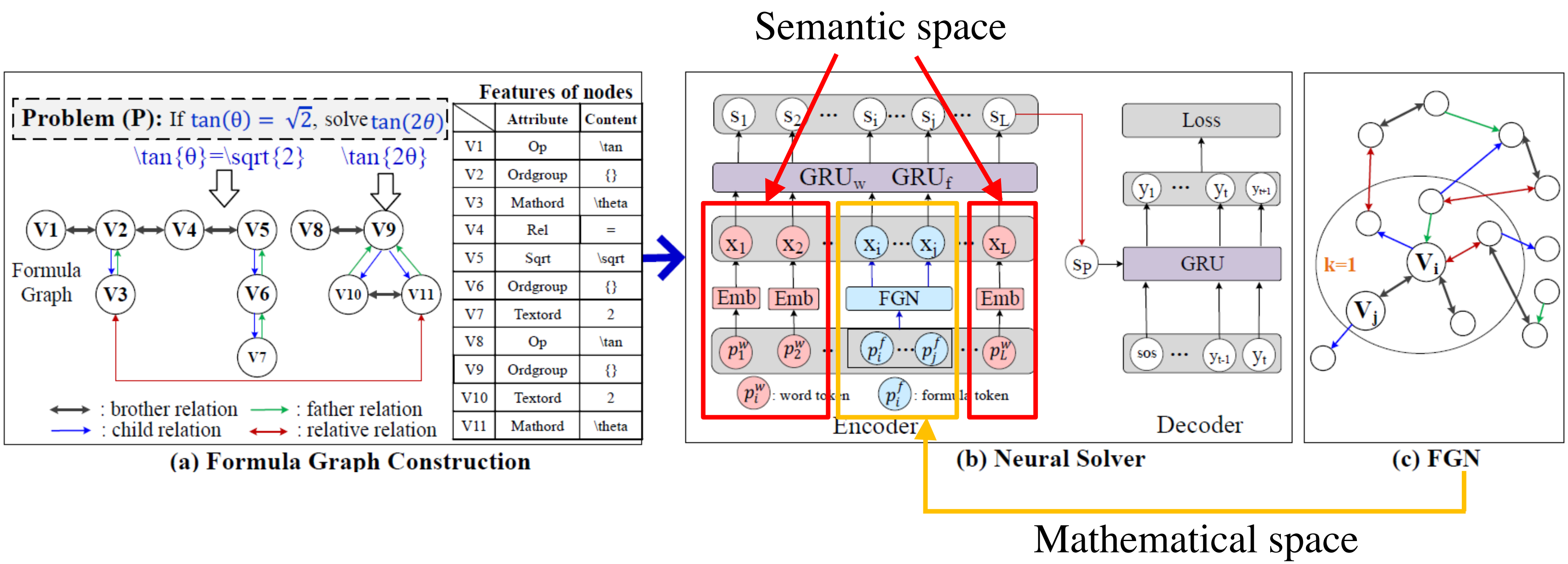
- Given
  - Mathematical problem: $P = \{p_1, p_2, \cdots, p_L\}$
    - Token: p_i is a word token or formula token (e.g., quantities, symbols)
- Goal
  - Read tokens from P
  - Gnerate answer sequence: $Y = \{y_1, y_2, \cdots, y_T\}$



Outline

NMS Framework

- NMS framework
  - Formula Graph Construction
    - Develop an assistant tool to construct formula dependency graph
  - Neural Solver
    - FGN: Formula graph network
    - Sequence model: Encoder-Decoder architecture



(a) Formula Graph Construction (b) Neural Solver (c) FGN

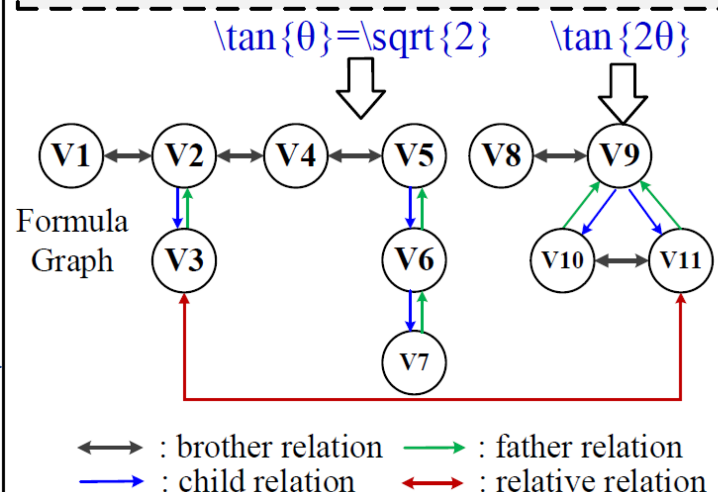
NMS Framework

- Formula graph construction
  - **Goal**: present formulas in a structural way
  - Develop a **TeX-based formula-dependent graph tool**
  - Nodes
    - Variables: θ
    - Numbers: 2
    - Operators: \tan
  - Edges (four relasions)
    - Brother, father, child
    - **Relative**
  - Features
    - Attribute, content

Advantages

- Reduce redundant
- Keep structure information
- Enhance semantic information

Problem (P): If $\tan(\theta) = \sqrt{2}$, solve $\tan(2\theta)$



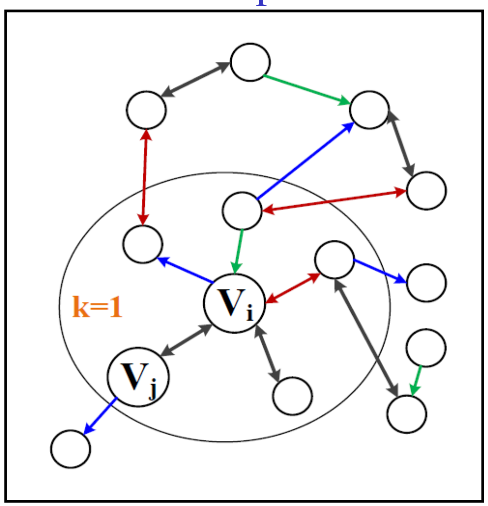
Features of nodes

	Attribute	Content
V1	Op	\tan
V2	Ordgroup	{}
V3	Mathord	\theta
V4	Rel	=
V5	Sqrt	\sqrt
V6	Ordgroup	{}
V7	Textord	2
V8	Op	\tan
V9	Ordgroup	{}
V10	Textord	2
V11	Mathord	\theta

NMS Framework

- Neural solver
  - FGN: capture fomula structure information
  - Sequence model: incorporate **semantic** and **structural** information

Formula Graph Network

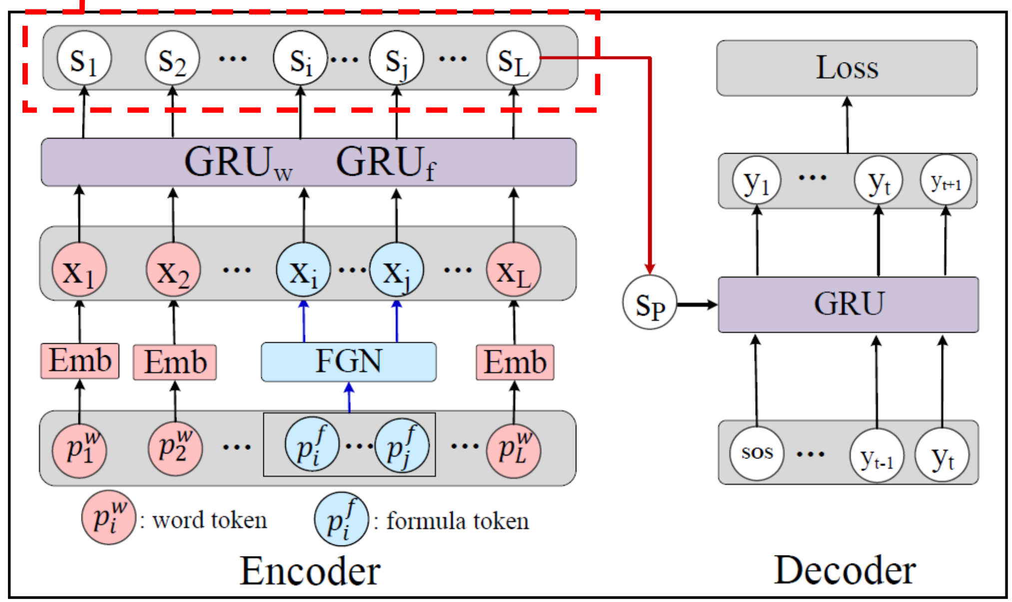


$$h_{N_i}^{k+1} = \sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{|N_i^r|} W_r^k h_j^k$$

$$h_i^{k+1} = \sigma(W^k h_i^k \oplus h_{N_i}^{k+1})$$

: brother relation
: child relation
: father relation
: relative relation

$$s_l = \begin{cases} \mathrm{GRU}_w(x_l, s_{l-1}; \theta_w) & \text{if} \quad x_l \text{ from word space,} \\ \mathrm{GRU}_f(x_l, s_{l-1}; \theta_f) & \text{if} \quad x_l \text{ from formula space.} \end{cases}$$



Neural solver

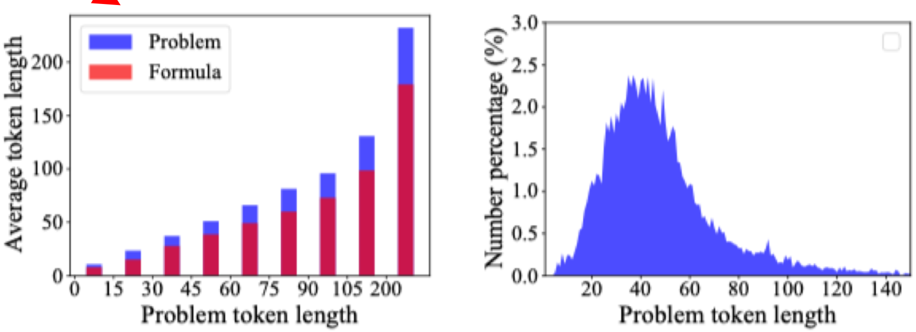
Outline

Experiment

- Dataset
  - MATH dataset (high school level)
- Data analysis
  - Formula tokens take large portions
    - 69% on average
    - Larger portions in shorter problems
- Baseline methods (seq2seq)
  - GRU
  - BiGRU
  - RMC
  - Attention
  - Transformer
- Evaluation metrics
  - ACC, BLEU, ROUGE

Table 1: The statistics of the dataset.

Num. problem	Avg. problem Length	Avg. answer Length	Avg. formula Number	Avg. formula Length
31,500	48.47	8.87	3.46	35.91



(a) Token Length (b) Problem Distribution

Experiment

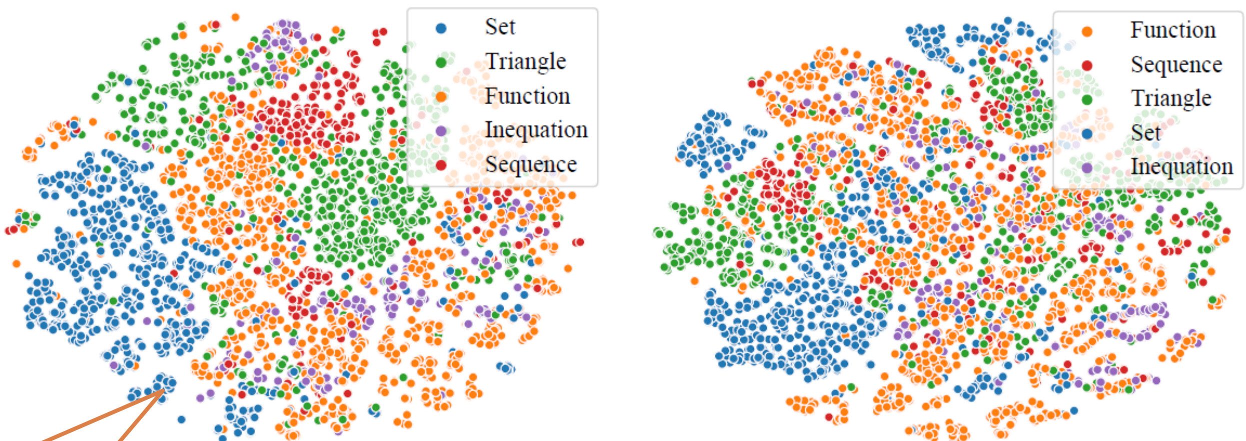
- Experiment
  - Task: solving mathematical problems
- Observations
  - **NMS** performs the best
    - Capture mathematical relations effectively
  - Transformer and Seq2Seq-BiGRU perform better than other baselines
    - Design sophisticated encoders
  - RMC performs not very well
    - Probably because it requires many parameters

Table 2: The overall performance of problem solving.

Training/test ratio	60%/40%			70%/30%			80%/20%			90%/10%		
Metric	ACC	BLEU	ROUGE	ACC	BLEU	ROUGE	ACC	BLEU	ROUGE	ACC	BLEU	ROUGE
Seq2Seq-GRU	27.94%	27.78	52.10	28.72%	28.43	52.89	29.38%	29.68	52.99	30.25%	30.50	55.37
Seq2Seq-BiGRU	30.40%	31.78	55.83	30.85%	32.09	56.94	30.47%	31.89	57.05	32.87%	33.85	57.70
Seq2Seq-RMC	26.38%	26.32	49.06	26.50%	27.01	49.81	26.82%	26.98	50.53	27.51%	27.66	51.47
Seq2Seq-Attn	29.59%	30.16	54.95	30.58%	31.48	55.15	30.94%	32.07	56.42	31.13%	31.91	55.59
Transformer	30.71%	32.15	55.81	31.31%	32.90	55.85	32.20%	33.52	56.12	32.32%	34.82	57.38
NMS-F	29.94%	31.24	53.71	30.78%	32.99	55.91	31.86%	33.54	56.96	32.16%	33.62	57.98
NMS	**31.85%**	**33.09**	**55.61**	**32.14%**	**33.88**	**57.22**	**33.65%**	**34.08**	**57.55**	**34.21%**	**36.67**	**59.93**

Experiment

- Visualization
  - Task: project problems embeddings into 2D space by t-SNE
- Observations
  - Problems with same concepts learned are easier to be grouped
    - They are closer in the hidden space
  - Problems with simple formula structures cluster nearly
    - E.g., “Set” problems
  - Many types of formulas cause different patterns
    - E.g., “Function” problems



More reasonable

(a) NMS

(b) Seq2Seq-BiGRU

Outline

Conclusion & Future work

- Overall results
  - Develop a TeX-based formula-dependent graph tool to maintain the structural information of each problem.
  - Design FGN to capture mathematical relations.
  - Design a neural solver to incorporate semantic infomation and structural infomation.

- Future work
  - Seek ways to predict quantities effectively
    - ½ vs. $^{111}/_{222}$
  - Design different graph networks for learning formula structure
    - Reasoning on different problem types
  - Consider more specific structures of more complex problems
    - "geometry" problem: containing figures

Thanks for your listening!

huangzhy@mail.ustc.edu.cn