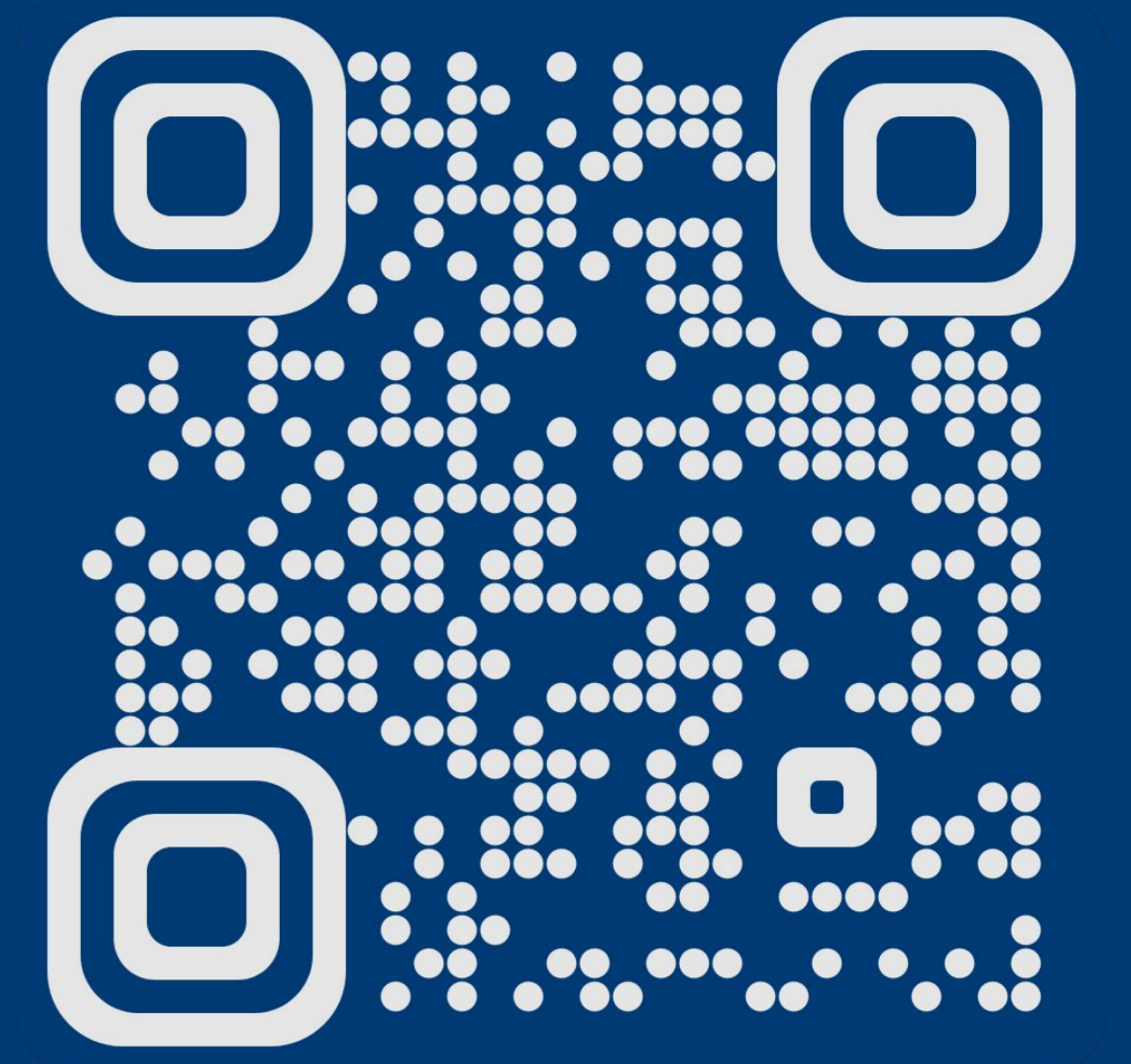


ArMATH: a Dataset for Solving Arabic Math Word Problems

Reem Alghamdi¹, Zhenwen Liang², Xiangliang Zhang¹²
¹ King Abdullah University of Science and Technology (KAUST)
² University of Notre Dame
reem.alghamdi@kaust.edu.sa, {zliang6, xzhang33}@nd.edu



Introduction

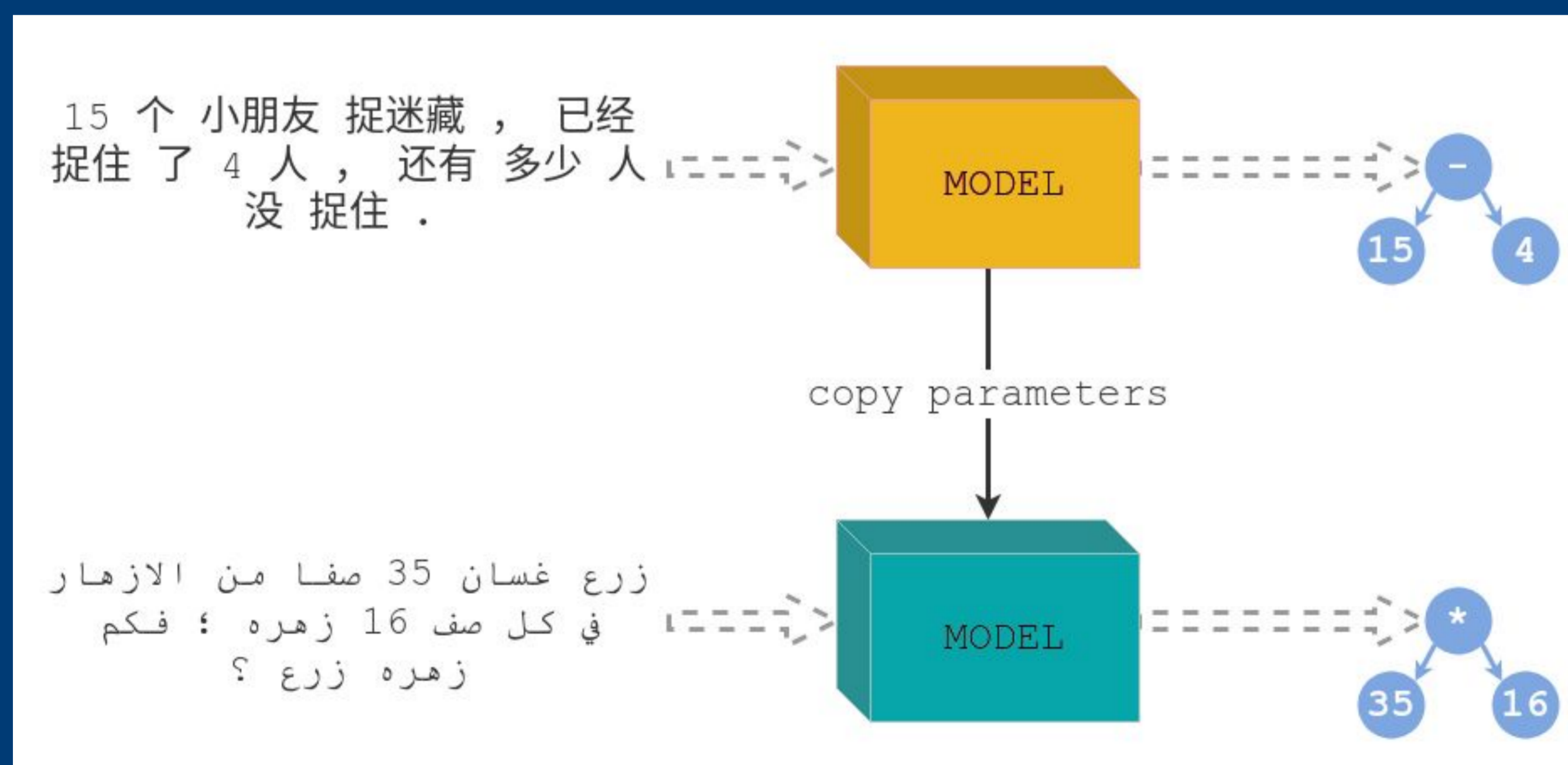
A Math Word Problem (MWP) is a textual description of a mathematical problem, which should be solved by deriving a math equation to reach the answer.

Question (English)	Ali ate 3 apples from the basket and his brother ate 2. If the basket originally had 8 apples, how many are left?
Question (Arabic)	أكل علي 3 تفاحات من سلة فواكه، وتناول أخوه تفاحتين؛ فكم تفاحة تبقى في السلة إذا كان فيها 8 تفاحات في البداية؟
Equation	$x = 8 - 3 - 2$
Template	$x = N2 - N0 - N1$
Answer	3

An example of a math word problem (MWP)

Solvers

- Global Tree Structure (GTS) proposed by (Xie and Sun, 2019b) was used as a base model.
- Pretrained word2vec models (aravec, fasttext) were used.
- GTS was first trained using the chinese MATH23K dataset. Then, the knowledge was transferred to the arabic solver over ArMATH.



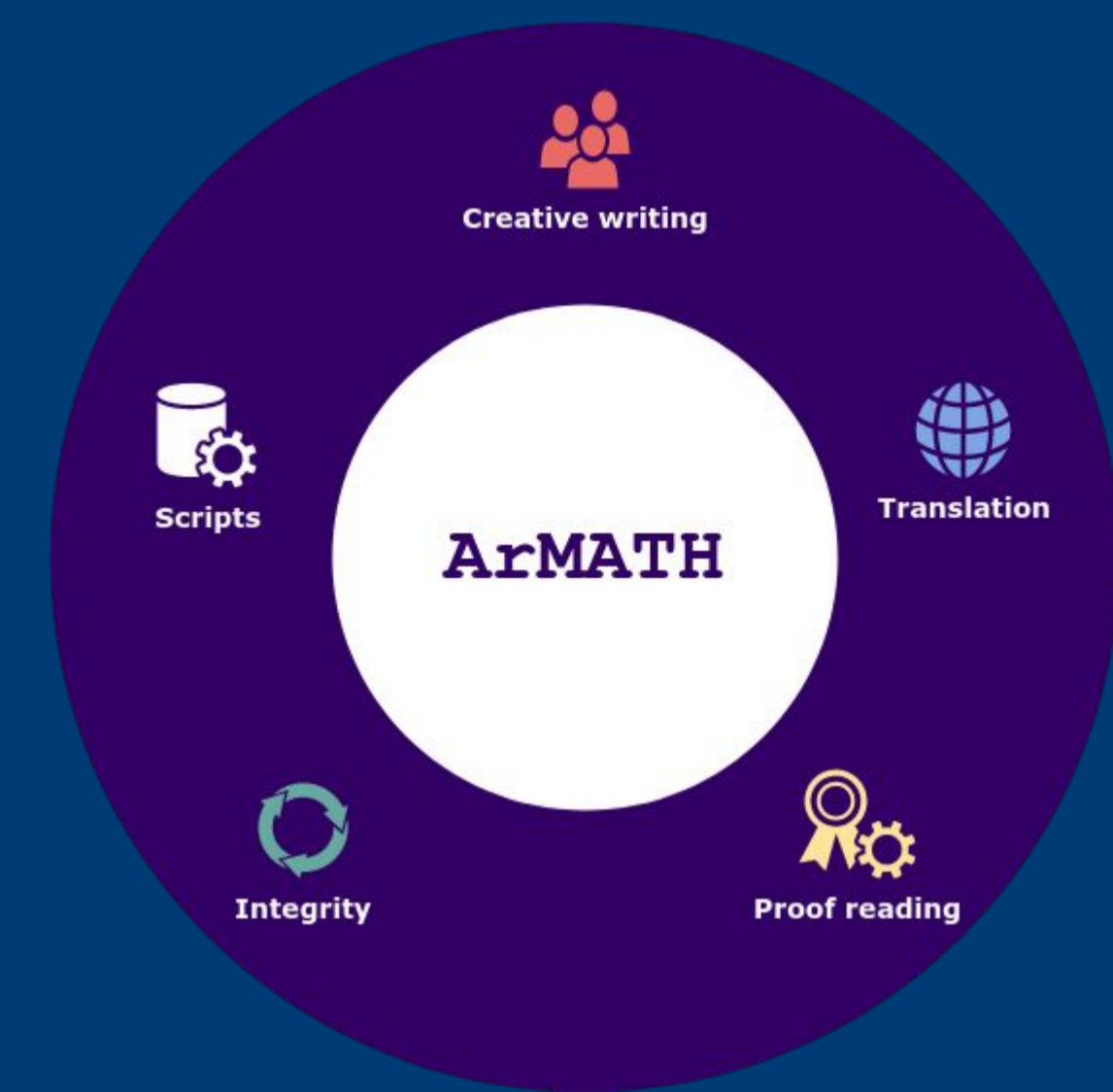
Model demonstration

MWP datasets

- Existing MWP datasets differ in terms of the natural language of the questions, questions types, question level, and type of intermediate information given.
- The most famous datasets are MAWPS dataset in English and MATH23K dataset in Chinese.
- Both are one-variable, primary-school level questions.

ArMATH Dataset

- 6,000 samples (question-equation pairs): 3,533 samples from creative writing, and 2,467 MATH23K-inspired samples.
- The dataset covers one-variable, primary-school level questions.
- preprocessing was done for easier usage



Data making

Dataset	MATH23K	MAWPS	ArMATH
Language	Chinese	English	Arabic
# Questions	23,160	3,320	6,000
# Templates	2,187	311	883
# Sentences	70.1K	6.3K	11.2K
# Words	822K	-	8.5K

MWP datasets comparison

Evaluation & Results

- The transfer learning enhanced solver has an accuracy of 74.15%, 3% higher than the solver without using transfer learning.
- Transfer learning was especially helpful for templates with few samples representing them.

