TATIANA SHAVRINA

R&D TEAM, NLP, CDS

Sberbank
Artificial Intelligence

HOW WE RUN A FULL-FLEDGED ROBOT COLLEGE STUDENT TEST

AI JOURNEY 2019, Russian state exam

Bringing together

- reasoning
- knowledge
- text generation
- transformers and universal LMs

AI Journey is a ML competition, aiming to develop Artificial General Intelligence (AGI)

During the competition this year, participants were offered a Robot College Student Test - to pass State Exam in Russian language using various sources of school knowledge, commonsense knowledge, previous exam data with answers.

Solutions are evaluated both automatically (test part) and with the help of teachers (essay part)

Exam Structure

26 tasks - questions with open and closed answer options (59% of the grade)

27th task - essay on the text (41% of the grade)

Types of tasks:

- spelling
- logics
- semantics
- punctuation
- orthoepy (stress)
- morphology and syntax
- text compilation / generation

TASK EXAMPLE

№ 9. Indicate the answer options in which in all words of one row an stressed vowel in a root is missing. Record the response numbers.

1) lo..k, pl..nt, adj..ctive
2) sp..ral, lin..n, comf..rt
3) b..ige, f..rmat, h..rden
4) prel..minary, r..dside, n..tice
5) instru..mental, li..tning, swi..mer

Answer: 3,4

AI Journey in numbers

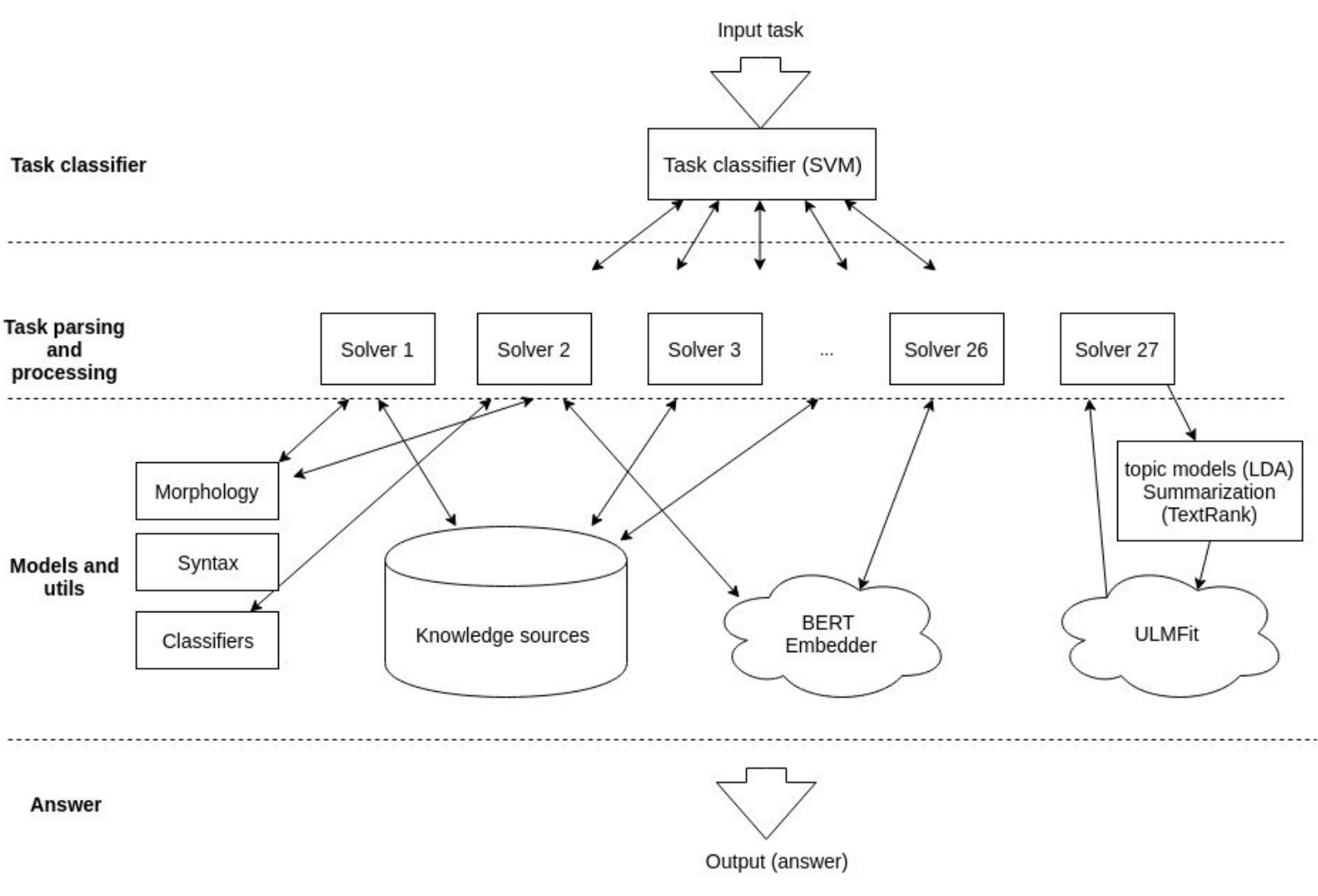
- 98 teams
  The most active participants sent their solution 206 times!
- one of the most difficult baselines. The authors are 11 people from Sberbank and Huawei, and contains 4000+ lines of code.
- 88 teams took advantage of the baseline (30 points)
- competition for everyone - the technical level of decisions of the participants is very different. There are company and student teams in the top 10.

contest.ai-journey.ru/ru/leaderboard

#	Команда	Результат решений			Последнее решение	Попыток
		Тест (max 34)	Сочинение (max 24)	Итого (max 100)		
1	qbic	19,93	14,25	59,77	01 ноября 2019, 23:56	48
2	Bilbo Bagging	20,90	12,67	58,47	01 ноября 2019, 18:09	170
3	Magic City	14,50	16,33	55,63	01 ноября 2019, 23:47	115
4	borsden	15,77	13,33	53,40	29 октября 2019, 05:54	74
5	Ololosh AI	16,70	10,58	50,93	30 октября 2019, 21:06	150
6	nice	12,63	13,83	49,70	01 ноября 2019, 16:56	40
7	Niw	21,20	4,75	49,20	01 ноября 2019, 19:45	111
OOC	CDS_team	17,17	5,42	45,23	01 ноября 2019, 18:02	21
8	stickman	17,07	5,00	43,77	01 ноября 2019, 14:57	70
8	Orcs	13,00	8,92	43,77	01 ноября 2019, 23:58	19

Eventually
Russian AI ...
got 69 out of 100
for the exam!

BASELINE


Input task
Task classifier
Task classifier (SVM)
Task parsing and processing
Solver 1
Solver 2
Solver 3
...
Solver 26
Solver 27
Models and utils
Morphology
Syntax
Classifiers
Knowledge sources
BERT Embedder
topic models (LDA)
Summarization (TextRank)
ULMFit
Answer
Output (answer)

['My position is {}'.format(problem_thesis_dict[theme])]

HOW PARTICIPANTS DID IT

pain =”

```
if "most" in x.lower():
    y_pred =
solver_most_strong(x)
```

TASK QUALITY

task	1st score	task	1st score
task_1	0,70	task_15	0,83
task_2	0,53	task_16	0,90
task_3	0,70	task_17	0,90
task_4	0,97	task_18	0,56
task_5	0,66	task_19	0,86
task_6	0,43	task_20	0,90
task_7	0,90	task_21	0,50
task_8	0,86	task_22	0,63
task_9	0,83	task_23	0,30
task_10	0,97	task_24	0,33
task_11	0,90	task_25	0,73
task_12	0,77	task_26	0,75
task_13	1,00	task_27	0,68
task_14	0,83		

Completely Solved Tasks

100% - task 13, spelling - choosing type of negation for a verb
Solution: using BERT for binary classification.

97% - task 4, orthoepy
Solution: dictionary lookup is combined with memorizing correct and erroneous word stresses and improving the rules to choose the right option; using dictionaries and rules to score candidates.

97% - task 10, spelling of prefixes
Solution: morphological analysis and dictionary lookup; a version of the baseline approach with more complex rules; a version of the baseline approach with a custom knowledge base.

Completely Solved Tasks

90% - task 7, errors in word formation
Solution: scoring candidates with n-gram models and morphological analysis; a sophisticated system of rules that includes custom dictionaries; an improved version of the baseline that includes dictionary lookup.

90% - task 11, spelling suffixes
Solution: a complex component-based approach with custom dictionaries, morphological analysis and
rules; a logistic regression fit on word features to predict the missing letter.

The Hardest Tasks

40-50% - task 6, elimination of speech redundancy
Solution: pairwise comparison of fasttext embeddings of all nouns and verbs with exception of stop words; a dictionary lookup approach with a fallback to word2vec and cosine similarity in case the lookup is failed.

40-50% - task 21, punctuation rules
Solution: pymorphy2 and rule-based approach; an LGBM classifier fit on TF-IDF and morphological features from pymorphy2.

40-50% - task 18, placement of commas + reasoning
Solution: replacing each placeholder with a [MASK] token, using BERT's output to decide if this placeholder replaces a comma, carefully chosen probability thresholds (individual for each task).

The Hardest Tasks

30% - task 23, narrative-reasoning-description
Solution: scoring the candidates upon the cosine similarity of word2vec embeddings.

33% - task 24, the lexical meaning of the word
Solution: synonyms and antonyms are found with fasttext, idioms are extracted by means of dictionary lookup, and in all other cases, the system simply returns the least frequent word in the text that is also not a proper name. They also include a component-based approach that combines rules, morphological analysis by Mystem 7, Word2Vec and dictionary lookup.

The Essay

Self-written templates win :(
few-shot classification of themes and problems
(BERT embeddings)
Search for an author’s position in the text using BERT

Teachers experiencing Uncanny valley with generative models

experiments:

- GPT2
- TextRank,
- NER by DeepPavlov
- cutting ready-made essays into sentences
  and composing a new text
- periphrases of pieces of text using vector patterns

SberCloud

ARTIFICIAL INTELLIGENCE JOURNEY

Проверь себя в ЕГЭ. Кто круче? Ты или машина?

AGI | ТЕСТОВЫЕ ЗАДАНИЯ | ПОЛНЫЙ ВАРИАНТ | СОЧИНЕНИЕ

Искусственный интеллект сдает экзамен ЕГЭ по русскому языку

Высокий уровень сложности

специфика русского языка
низкая вероятность cheating'а

Результ AGI - 69 баллов

уровень среднего школьника на ЕГЭ

26 вопросов

с закрытыми и открытыми вариантами ответа в тестовой части

100-балльная система оценки
36 баллов – проходной балл в университет

Baseline-соревнования

4000 строк кода
первое решение подобной задачи

Профессиональные учителя

проверяют сочинения по прочитанному тексту

NEXT STEPS

1. In general, the score is good - 69 points
   This is a good mark, and an “excellent” mark starts with 72
2. Open best solution on github
3. Permanent leaderboard, solution development
4. Demo on the site

TO DOs:

5. Long text generation
   - controlled generation
   - discursive features generation
6. Automatic error correction (spelling, speech standards), punctuation placement - components of the Grammarly analog

7. New subjects: biology, history, social studies


THANK YOU!
Artificial Inte
1 000 000 ₽
lamoda
lamoda
MACHINE HEAD