

Automatic Answer Grading Using transformers

Mário Harvan, Patrik Tiszai, Juraj Martiček

Overview

- Automatic hand written answer grading
- Optical character recognition
- Sentence bert
- OpenAI API for fine tuning language models

Dataset creation

- OCR scans of SUI exams
- Templates for answer separation
- Page order corrections
- Dataset validation

```
00320d24-94ad-4e9c-b584-d4e6c69426c0.xml:
```

```
  answers:
```

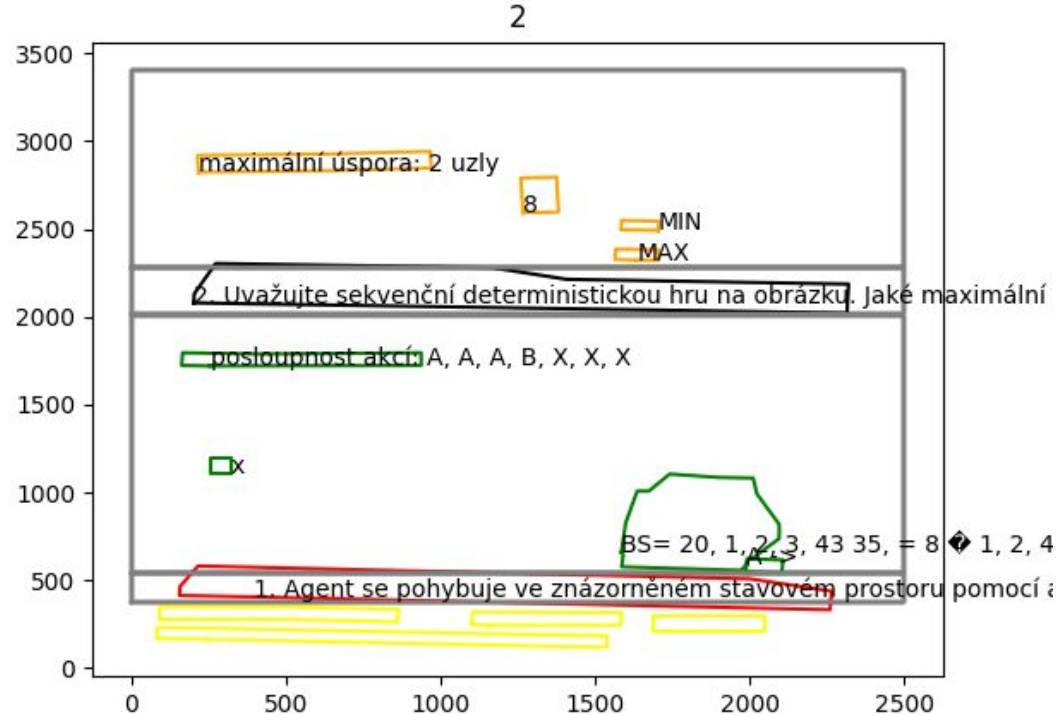
```
    6: PROTOŽE KAŽDÝ POCHOD LZE ZAPSAT V LINEÁRNÍM TVARU  $\frac{1}{3}x^3 + bx^2 + cx + d$   
      n Snažíme se natrénovat polynom N - stupně, ODHADUJEME PARAMETRY  $a, b, c, d$  URČUJÍCÍ  
      SECI A POJUM KŘIVKY;- SNAŽÍME SE MÍT CO NEJMENŠÍ POČET PARAMETRŮ PRO ATIONÁLNÍ  
      F-CI
```

```
    7: per; $P(X|w) = P(w) = P(w|x) - P(X);P(x|w)$ 
```

```
  login: 50568a87cf2aab80a607af9913d3482c
```

```
  points: &id002
```

Sorting and separating answers



SBERT similarity grading

- Cosine similarity for comparing answers
- SBert pre-trained multilingual model
- Answer classification to grade groups

OpenAI model fine-tuning

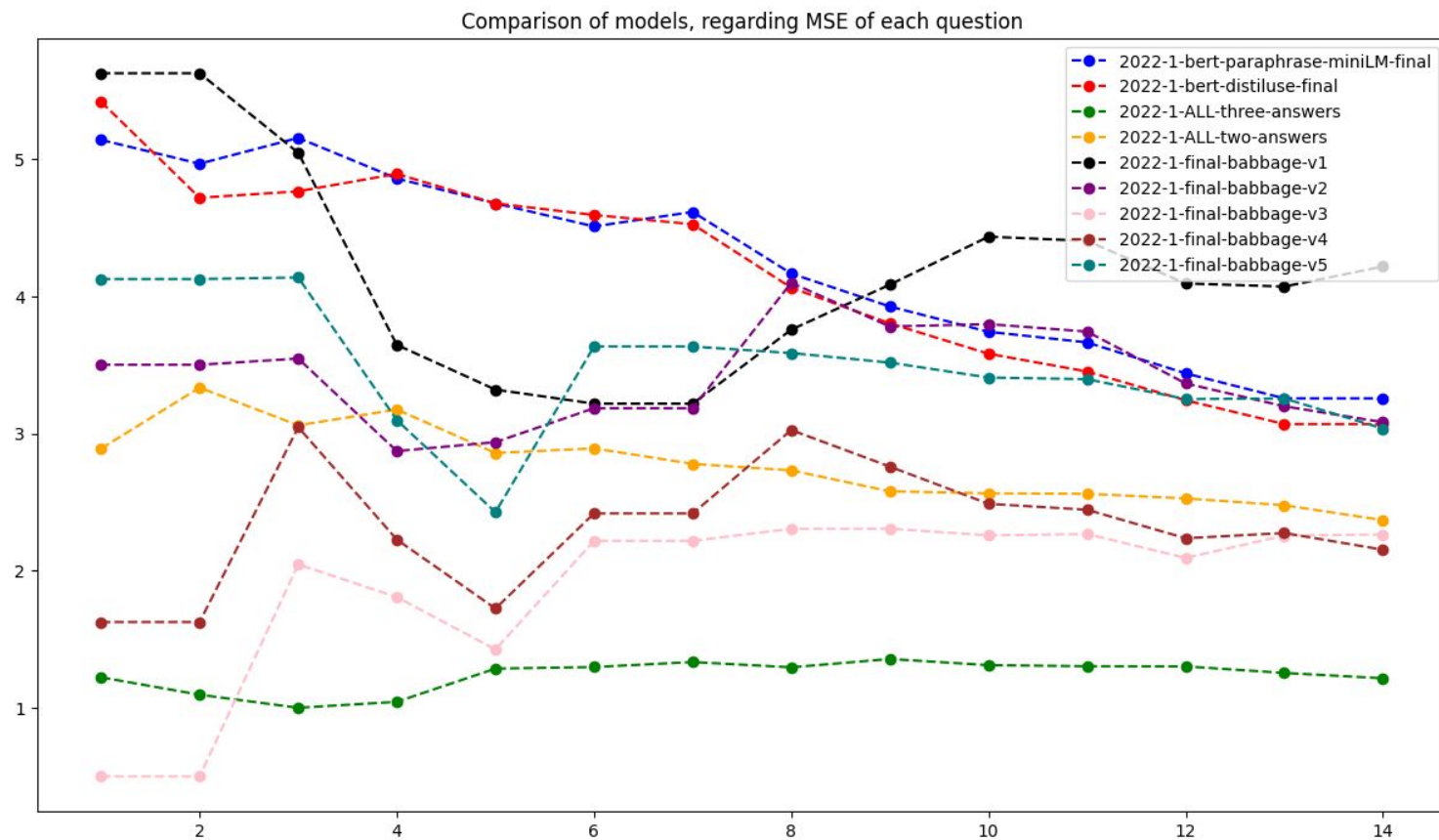
- OpenAI API for fine-tuning
- 5 fine-tuned models
- Ada and Babbage engines

```
{  
  "prompt": "For question: 6. Uvedte dva příklady diskriminativně trénovaných modelů pro klasifikaci,  
  které umožňují vyhodnocení aposteriorních pravděpodobností tříd pro vstupní (testovaný) vzor.  
  Alespoň pro jeden z těchto dvou příkladů uveďte, jak se tyto aposteriorní pravděpodobnosti konkrétně spočítají.  
  V čem se jejich výpočet liší ve srovnání s generativními klasifikačními modely? se modely - diskrim rozsahy ,  
  grade answer (from 0 to 4): 65 to. 9=0, h-h DFS: g = len (death), h = o (neboru o)",  
  "completion": "",  
  "meta": "2022-2-6-00d52da93a2ea09d33d9c34959629374"  
},
```

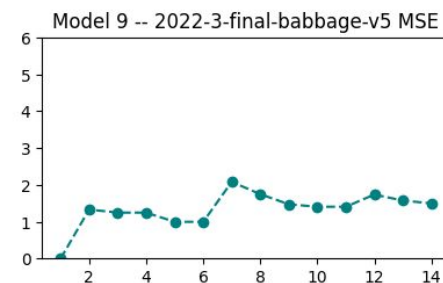
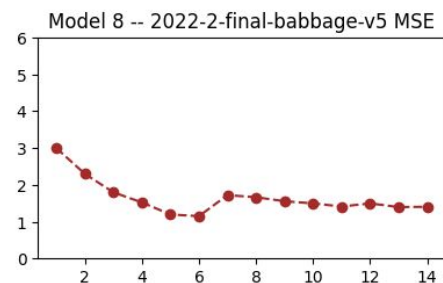
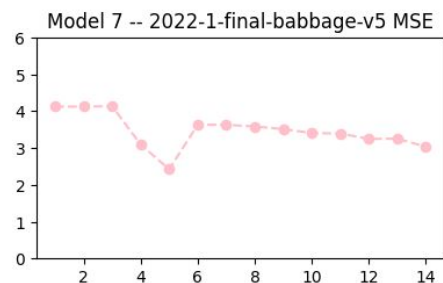
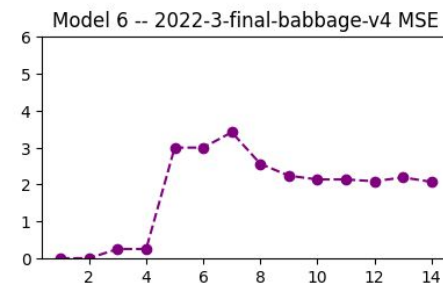
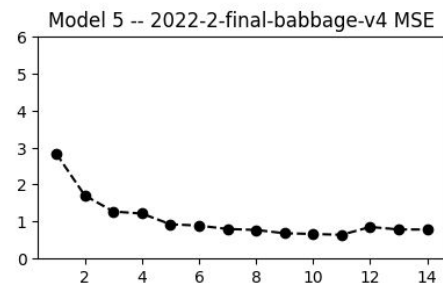
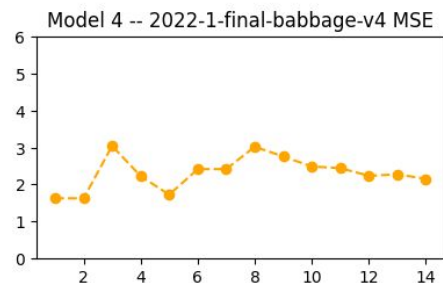
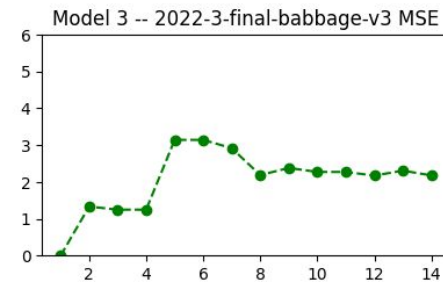
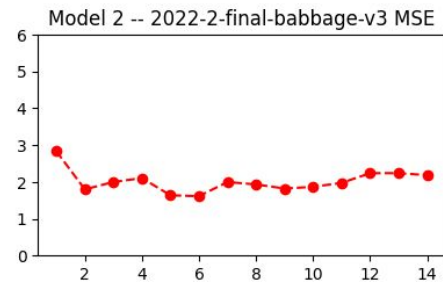
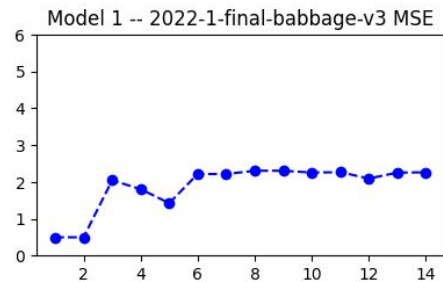
Evaluation

- Mean squared error for each question
- Remove graphical answers
- Comparison of both methods
- Selection of best performing models

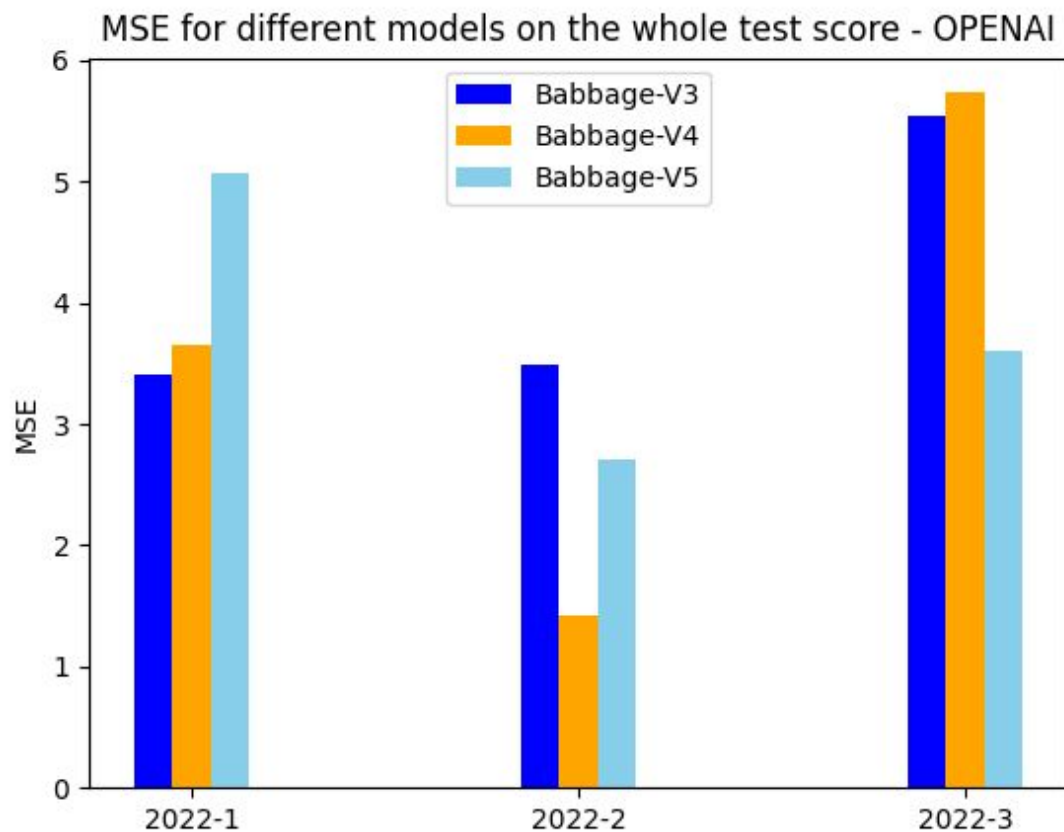
Results



Results



Results



Thanks for your attention