A Brief Introduction to Me and My Research

石雨凌

上海财经大学

2021.7.26

目录

个人简介

- 上海财经大学, 数学与应用数学专业, 年级排名 **13/104**, 本学期成绩排名 **1/104**. TOEFL 102分.
- 主要课程: **数学分析 (3.3), 高等代数 (4.0), 概率论 (4.0), 文本挖掘 (4.0), 深度学习 (4.0), 人工智能 (4.0).**
- 完成多项自然语言处理相关研究项目, 且有一篇计算数学论文已发表.
- 曾获高中全国物理竞赛省一等奖 (实验部分全省第 8 名), 自学大学与研究生阶段物理课程.

Outline

有限元方法简介

Finite element space and grids:

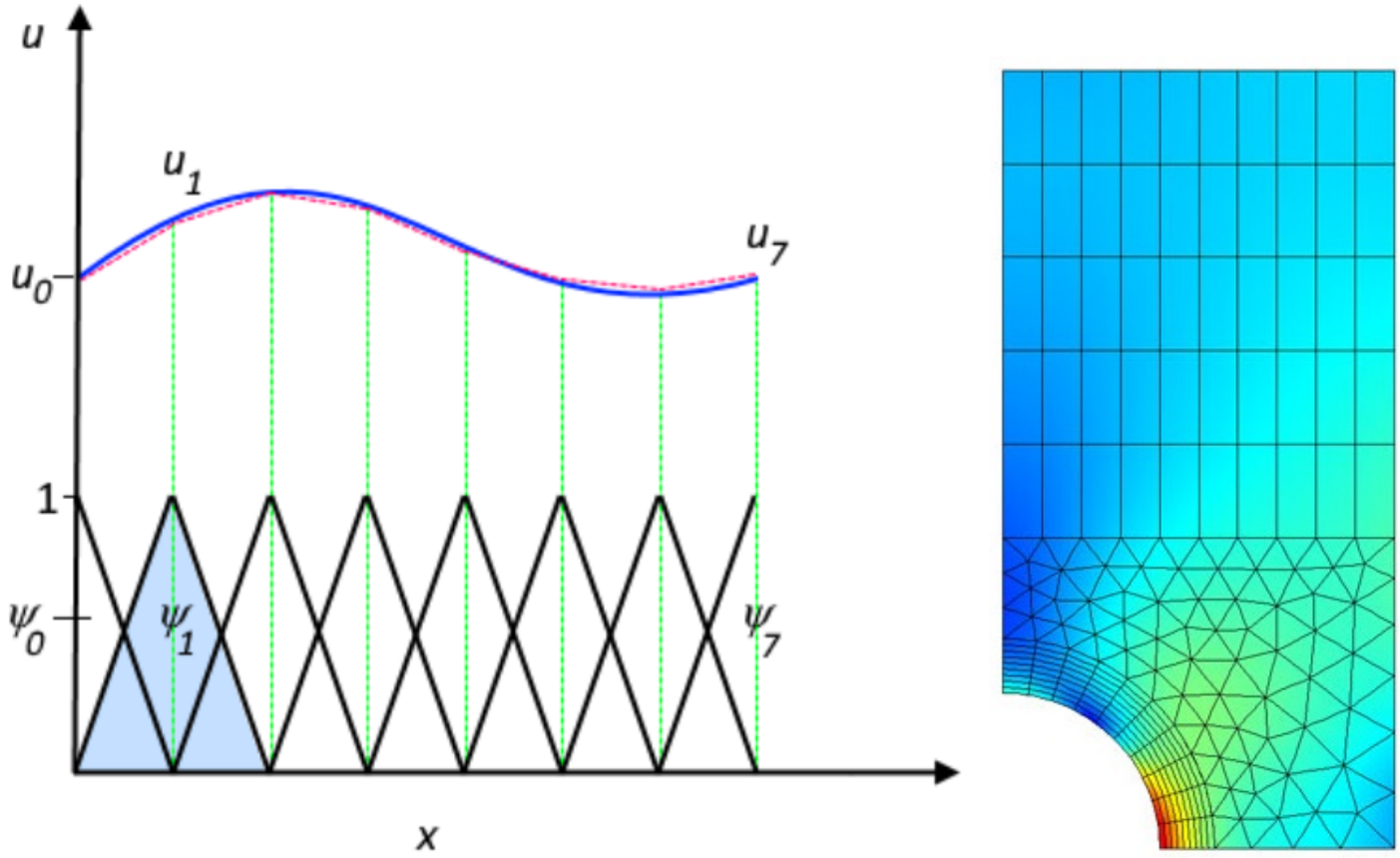


图: Linear basis in 1D; Grids

主要内容

Equation:

$$\begin{cases} \varepsilon^2 \Delta^2 u - \Delta u = f & \text{in } \Omega, \\ u = \partial_n u = 0 & \text{on } \partial\Omega, \end{cases}$$

- Original ways to solve:
  - Conforming elements: computational expensive
  - Non-conforming elements: isn't convergent

主要内容

Our work:

- Modified the right hand side via projection:

$$\begin{aligned} (\nabla w_h, \nabla \chi_h) &= (f, \chi_h) && \forall \chi_h \in W_h \\ \varepsilon^2 a_h(u_{h0}, v_h) + b_h(u_{h0}, v_h) &= (\nabla w_h, \nabla_h v_h) && \forall v_h \in V_{h0} \end{aligned}$$

- Decoupled the left hand side into four simple equations:

$$\begin{aligned} (\mathrm{curl}_h z_h, \mathrm{curl}_h v_h) &= (\nabla w_h, \nabla_h v_h) && \forall v_h \in V_{h0} \\ (\phi_h, \psi_h) + \varepsilon^2 (\nabla_h \phi_h, \nabla_h \psi_h) + (\mathrm{div}_h \psi_h, p_h) &= (\mathrm{curl}_h z_h, \psi_h) && \forall \psi_h \in V_{h0}^{CR} \\ (\mathrm{div}_h \phi_h, q_h) &= 0 && \forall q_h \in \mathcal{Q}_h \\ (\mathrm{curl}_h u_{h0}, \mathrm{curl}_h \chi_h) &= (\phi_h, \mathrm{curl}_h \chi_h) && \forall \chi_h \in V_{h0} \end{aligned}$$

实验结果

- Can be solved efficiently with the simplest Morley element.
- Final paper was accepted by *Journal of Scientific Computing*.

h	#dofs	Eq.(5.1)	Eq.(5.7a)	Eq.(5.7b)-(5.7c)	Eq.(5.7d)
		steps	steps	steps	steps
2^{-1}	24	1	1	16	1
2^{-2}	112	1	4	27	3
2^{-3}	480	4	5	34	5
2^{-4}	1984	6	7	34	7
2^{-5}	8064	6	9	41	9
2^{-6}	32512	7	11	43	11
2^{-7}	130560	7	14	44	14
2^{-8}	523264	9	17	46	17
2^{-9}	2095104	9	20	50	21
2^{-10}	8384512	12	27	55	27

图: Robust iteration steps when solving

Outline

研究背景

Existing methods

- Gradient based methods: gradient, dot product with embeddings, integrated gradient ...
- Perturbation based methods: input reduction, adversarial perturbations ...

Trying to explain what the model is learning during the training process.

$$Loss\left(x_1,\ldots,x_d\right)=\sum_{n_1=0}^{\infty}\cdots\sum_{n_d=0}^{\infty}\frac{\left(x_1-a_1\right)^{n_1}\cdots\left(x_d-a_d\right)^{n_d}}{n_1!\cdots n_d!}\left(\frac{\partial^{n_1+\cdots+n_d}f}{\partial x_1^{n_1}\cdots\partial x_d^{n_d}}\right)\left(a_1,\ldots,a_d\right)$$

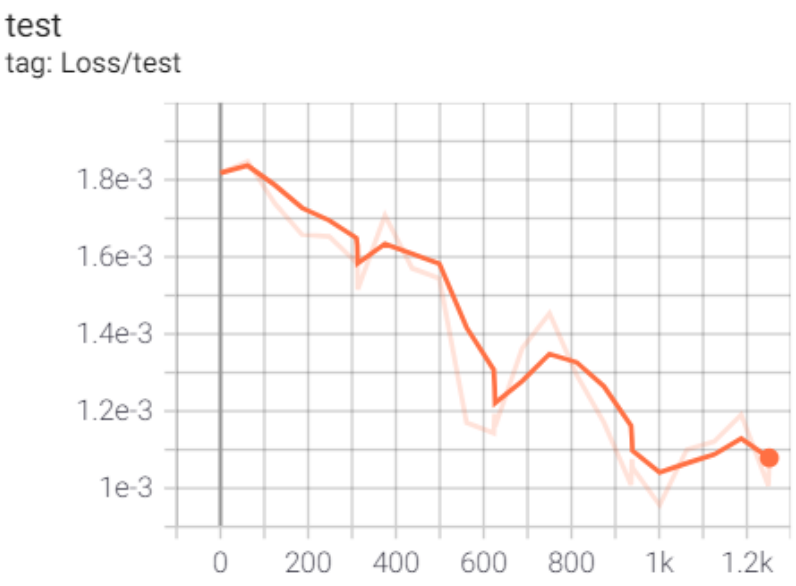
$$=f\left(a_1,\ldots,a_d\right)+\sum_{j=1}^{d}\frac{\partial f\left(a_1,\ldots,a_d\right)}{\partial x_j}\left(x_j-a_j\right)+\frac{1}{2!}\sum_{j=1}^{d}\sum_{k=1}^{d}\frac{\partial^2 f\left(a_1,\ldots,a_d\right)}{\partial x_j\partial x_k}\left(x_j-a_j\right)\left(x_k-a_k\right)$$

$$+\frac{1}{3!}\sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{l=1}^{d}\frac{\partial^3 f\left(a_1,\ldots,a_d\right)}{\partial x_j\partial x_k\partial x_l}\left(x_j-a_j\right)\left(x_k-a_k\right)\left(x_l-a_l\right)+\cdots$$

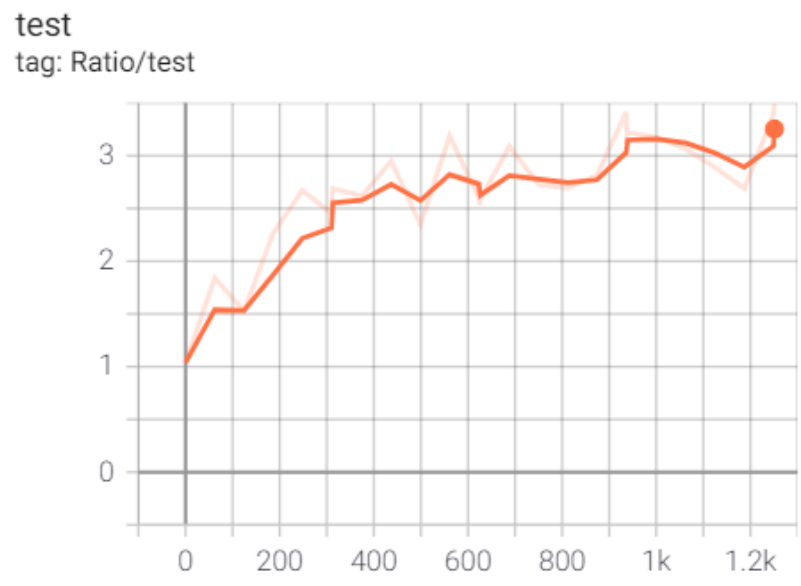
实验结果

- Model learning relations between "essential" words?
- $\frac{\partial Loss}{\partial E_i E_j}$ durning training BERT-Base:

Loss



Ratio



实验结果

- Sentences:
  1. Two young boys of opposing teams play football, while wearing full protection uniforms and helmets.
  2. Boys play football.

	word	most relavent	score
0	[SEP]	[SEP]	0.044089
2	[CLS]	[SEP]	0.025706
4	[CLS]	[SEP]	0.024833
6	[SEP]	.	0.014351
8	football1	[SEP]	0.014056
10	.	[SEP]	0.013824
12	[SEP]	football2	0.011118
14	[SEP]	football1	0.010987
16	football2	[SEP]	0.010924
18	[SEP]	helmets	0.010910

图: At the beginning of training

实验结果

- Sentences:
  1. Two young boys of opposing teams play football, while wearing full protection uniforms and helmets.
  2. Boys play football.

	word	most relavent	score
0	football1	football2	0.066224
2	and	football2	0.053644
4	football2	protection	0.034782
6	boys	football2	0.025395
8	football2	play	0.021575
10	uniforms	football2	0.018982
12	[CLS]	football2	0.018318
14	[SEP]	football2	0.013509
16	helmets	football2	0.012761
18	football1	protection	0.012189

图: Trained for 200 batches

Outline

项目介绍

数据集介绍

- Used the Natural Questions (NQ) (**Kwiatkowski et al., 2019**) dataset from Google AI.
- Each example is comprised of a google query and a corresponding Wikipedia page.

主要方案

- Fine-tuning on SQuAD 2.0
- Mixed Precision Training
- Hard Negative Sampling
- Sifting candidates

实验结果

Model	Public F1	Private F1
Kaggle Best	0.713	0.717
BERT Base baseline	0.516	0.482
BERT Base (Hard Negative Sampling)	0.579	0.574
BERT Sifted* $\rightarrow$ ALBERT xlarge	**0.640**	**0.659**
Sifted $\rightarrow$ BERT Base+ALBERT$_{\text{(ensemble)}}$	0.665	0.666
Sifted $\rightarrow$ BERT Large+ALBERT$_{\text{(ensemble)}}$	**0.738**	**0.718**

$*$ Sifted here stands for first using BERT base to sift candidates

Outline

主要内容

数据集介绍

- Complaints from citizens. The task is to predict label (200 classes in total).

主要方案

- Pre-trained models on similar dataset THUCNews.
- Adopted **focal loss** for the long-tailed distributed labels.
- Designed a **auxiliary sentence pair task**.
- Also tried adversarial training, data augmentation, pesudo labels, adding other layers after BERT, using RoBERTa-wwm, ERNIE, etc.

辅助任务设计

- The original classification task failed to utilize information in the labels.
- An auxiliary sentence pair task is then designed.

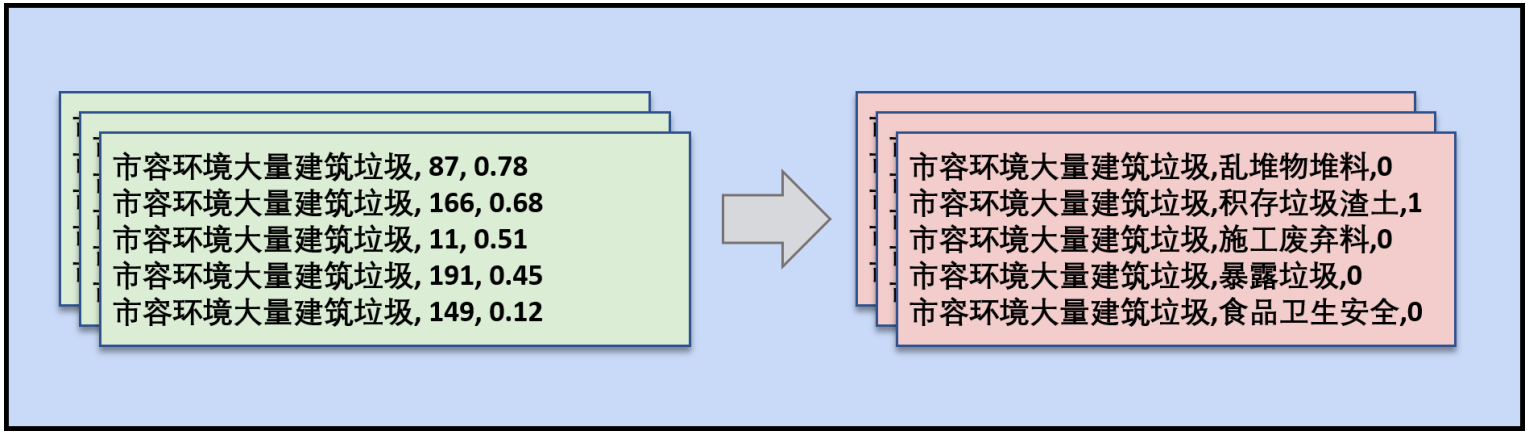


图: Generating pair data

实验结果

- Highest score in class (led by 0.4%).

表: Selected Experiment Results

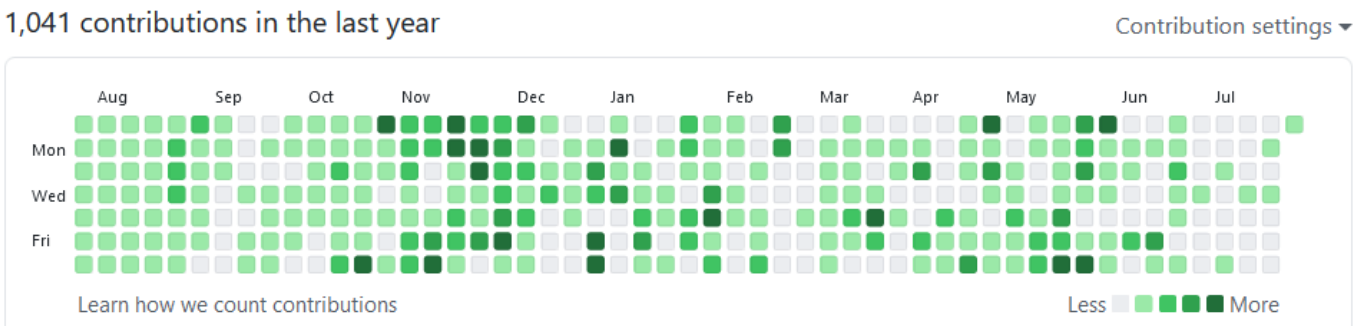
Model	Public score	Private score
ERNIE[1]	0.7981	0.8000
ERNIE[2]	0.7995	0.8030
ERNIE[3]	0.8049	0.8010

[1] Original Task: text classification
[2] Focal Loss
[3] Auxiliary Task: sentence pair classification

Outline

总结

- 具有良好的数学基础和丰富 NLP 项目经验.
- 积极乐观, 热爱研究并能主动探索.
- 希望有机会能在中文信息处理实验室继续提升自己.



Thanks for your attention!