

Информация о статье

Что? SCARF: Self-Supervised Contrastive Learning using Random Feature Corruption
и Кто? Dara Bahri, Heinrich Jiang, Yi Tay, Donald Metzler
Где? ICLR 2022 (spotlight presentation)
Когда? 29 Sept 2021 (modified: 23 Nov 2021)

Авторы

- Dara Bahri [16]

Работа(ы) Research Scientist в Google Research.

Ранее работал над recommender systems в Twitter, его работа привела к существенному росту активности пользователей (по его словам). Еще раньше делал internet-scale computer vision classification platform для маленького стартапа.

Образование UC Berkeley in Mathematics and Electrical Engineering and Computer Science (EECS).

Статьи На Google Scholar находится 37 работ.[17]

Много работ, посвященных трансформерам, attention.

Самые цитируемые:

- Berkeley advanced reconstruction toolbox, 2015 (248 цитирований)
- Efficient transformers: A survey, 2020 (203 цитирования)
- Synthesizer: Rethinking self-attention for transformer models, 2021 (108 цитирований)

- Heinrich Jiang [18]

Работа(ы) Research Scientist в Google Research.

Ранее работал Research Scientist в A*STAR Artificial Intelligence

Образование Princeton University, A. B. Mathematics

Статьи На Google Scholar находится 39 работ.[19]

Много работ посвящено кластеризации (DBSCAN) и классификации.

Самые цитируемые:

- To trust or not to trust a classifier, 2018 (239 цитирований)
- Identifying and correcting label bias in machine learning, 2020 (91 цитирование)
- Two-player games for efficient non-convex constrained optimization, 2019 (66 цитирований)

- Yi Tay [20]

Работа(ы) Research Scientist в Google Research.

Ранее работал trader в компании KCG Holdings. До этого был Algorithmic Trader Intern в Jump Trading и KCG Holdings.

Образование PhD - Computer Science (NTU), Bachelors of Engineering - Computer Science (NTU)

Статьи На Google Scholar находится 86 работ.[21]

Много работ посвящено рекомендациям, трансформерам.

Самые цитируемые:

- Deep learning based recommender system: A survey and new perspectives, 2017 (1522 цитирований)
- Latent Relational Metric Learning via Memory-based Attention for Collaborative Ranking, 2018 (212 цитирований)
- Efficient transformers: A survey, 2020 (204 цитирования)

- Donald Metzler[22]

Работа(ы) Senior Staff Software Engineer at Google.

Ранее работал Research Assistant Professor в University of Southern California (USC) и Senior Research Scientist в Yahoo!.

Образование PhD - Computer Science (University of Massachusetts Amherst), B. S. - Computer Science and Mathematics (Rose-Hulman Institute of Technology)

Статьи На Google Scholar находится 138 работ.[23]

Много работ посвящено markov random fields.

Самые цитируемые:

- Search engines: Information retrieval in practice, 2010 (1725 цитирований)
- A markov random field model for term dependencies, 2005 (1046 цитирований)
- Expected reciprocal rank for graded relevance, 2009 (843 цитирования)

Совместные работы

Рассмотрим всевозможные подмножества авторов (как минимум 2 автора). (Рис. 1)

Bahri	Jiang	Tay	Metzler	#papers
				2
				2
				21
				2
				2
				24
				2
				2
				21
				25
				14

Рис. 1: Совместные работы авторов. Клетка зеленая - если автор работал над статьей (если белая, то мог работать, а мог и не работать)

Выводы о совместной работе:

- Есть только 2 работы, над которыми трудились все 4 автора (включая данную). Вторая работа - Label Smoothed Embedding Hypothesis for Out-of-Distribution Detection
- Jiang работал с Tay и/или Metzler только вместе с Bahri (над SCARF и работой из предыдущего пункта)
- Bahri, Tay, Metzler много работали вместе
- Jiang работал отдельно с Bahri (то есть Bahri - «связующее звено»)

Кого цитирует

В данной статье 94 цитирования.

«Самоцитирование» (авторы цитируют собственные работы) только одно: Deep k-nn for noisy labels (D Bahri, H Jiang, M Gupta)

Модель из данной статьи используется в качестве одного из бейзлайнов.

Авторы утверждают, что techniques to generate views or corruptions очень domain-specific. Они приводят в пример подход color distortion[6] и cropping[7] в CV и token masking[8] в NLP

При этом утверждается, что существует мало статей про self-supervised contrastive representation learning on tabular datasets

Однако упоминается Self-supervised learning for large-scale item recommendations[5] со схожим подходом. (о ней подробнее в секции Схожие статьи)

Кто цитирует

Данную статью цитирует только 1 работа: Deep Neural Networks and Tabular Data: A Survey[4]

Эта работа - an overview of state-of-the-art deep learning methods for tabular data. Там просто рассказывается о новых моделях.

Схожие статьи

- Self-supervised learning for large-scale item recommendations[5] (Эта статья цитируется) Естественно, различия присутствуют: Yao at al. proposes masking random features in a correlated manner and applying a dropout for categorical features, while our approach involves randomizing random features based on the features' respective marginal training distribution [1]
- SubTab: Subsetting Features of Tabular Data for Self-Supervised Representation Learning[9] (На мой взгляд, это главный конкурент) Подход данной статьи: learning from tabular data into a multi-view representation learning problem by dividing the input features to multiple subsets[9] Отмечу, что в одном из рецензии на данную работу просили сравнить результаты с результатами SCARF. Авторы не смогли, так как кода SCARF не было в доступе и результаты были приведены не для accuracy, а для «win» метрики.

Возможная дальнейшая судьба

Применение Предложенную модель можно использовать тогда, когда ведется работа с tabular data в задаче классификации, когда мало размеченных данных или есть шум в лейблах.

Направления дальнейшей работы

- Подсчитать общепринятые метрики (accuracy)
- Сравнить с другими моделями (SubTab)
- Возможно, можно как-нибудь использовать лейблы и в схеме предобучения. (Понятно, что метод self-supervised, но есть этап supervised fine-tuning и, может, fine-tune старой схемой с каким-либо «навотором» для учета лейблов тоже будет полезен)

Еще больше «схожих» работ

Быстрый поиск в интернете выдал следующие работы, «схожие» с данной:

- CoaT [10]
- SKIRT [11]
- NAT [12]
- GloVe [13]

- SHOE [14]
- UMBRELLA [15]

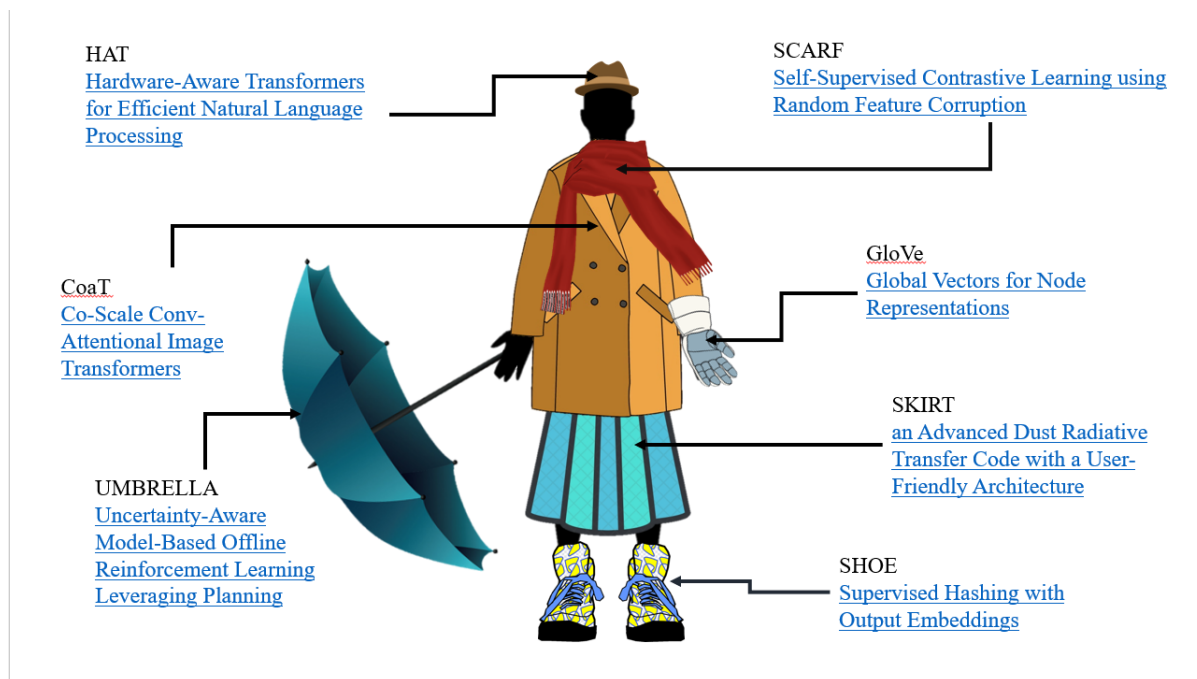


Рис. 2: «Схожие» работы

Список литературы

- [1] Dara Bahri, Heinrich Jiang, Yi Tay, Donald Metzler. SCARF: Self-Supervised Contrastive Learning using Random Feature Corruption
- [2] Dara Bahri, Heinrich Jiang, Yi Tay, Donald Metzler. Label Smoothed Embedding Hypothesis for Out-of-Distribution Detection
- [3] Dara Bahri, Heinrich Jiang, Maya Gupta. Deep k-NN for Noisy Labels
- [4] Vadim Borisov, Tobias Leemann, Kathrin Seßler, Johannes Haug, Martin Pawelczyk, Gjergji Kasneci Deep Neural Networks and Tabular Data: A Survey
- [5] Tiansheng Yao, Xinyang Yi, Derek Zhiyuan Cheng, Felix Yu, Ting Chen, Aditya Menon, Lichan Hong, Ed H. Chi, Steve Tjoa, Jieqi Kang, Evan Ettinger. Self-supervised Learning for Large-scale Item Recommendations
- [6] Richard Zhang, Phillip Isola, Alexei A. Efros. Colorful Image Colorization
- [7] Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations
- [8] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, Tie-Yan Liu. MPNet: Masked and Permuted Pre-training for Language Understanding
- [9] Talip Ucar, Ehsan Hajiramezanali, Lindsay Edwards. SubTab: Subsetting Features of Tabular Data for Self-Supervised Representation Learning
- [10] Weijian Xu, Yifan Xu, Tyler Chang, Zhuowen Tu. Co-Scale Conv-Attentional Image Transformers
- [11] Peter Camps, Maarten Baes. SKIRT: an Advanced Dust Radiative Transfer Code with a User-Friendly Architecture
- [12] Hanrui Wang, Zhanghao Wu, Zhijian Liu, Han Cai, Ligeng Zhu, Chuang Gan, Song Han. HAT: Hardware-Aware Transformers for Efficient Natural Language Processing
- [13] Robin Brochier, Adrien Guille, Julien Velcin. Global Vectors for Node Representations
- [14] Sravanthi Bondugula, Varun Manjunatha, Larry S. Davis, David Doermann. SHOE: Supervised Hashing with Output Embeddings
- [15] Christopher Diehl, Timo Sievernich, Martin Krüger, Frank Hoffmann, Torsten Bertram. UMBRELLA: Uncertainty-Aware Model-Based Offline Reinforcement Learning Leveraging Planning
- [16] <https://dara.run/index.html>
- [17] <https://scholar.google.com/citations?hl=en&user=j5PpT0wAAAAJ>

- [18] <https://www.linkedin.com/in/heinrich-jiang-a5593652>
- [19] <https://scholar.google.com/citations?user=RiDdF2YAAAAJ&hl=en>
- [20] <https://vanzytay.github.io/bio/>
- [21] https://scholar.google.com/citations?user=VBclY_cAAAAJ&hl=en
- [22] <https://www.linkedin.com/in/donmetzler>
- [23] <https://scholar.google.com/citations?user=bmXpOd8AAAAJ&hl=en>
- [24] <https://openreview.net/forum?id=vrhNQ7aYSdr¬eId=Rtv1hCHf90>