Friedrich-Alexander-Universität
Technische Fakultät

Chair for Computer Science 6 (Data Management)

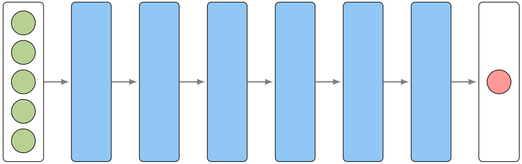
Deep Transfer Learning Model Selection

In A Time Series Modeling Context

Melanie Bianca Sigl
melanie.sigl@fau.de

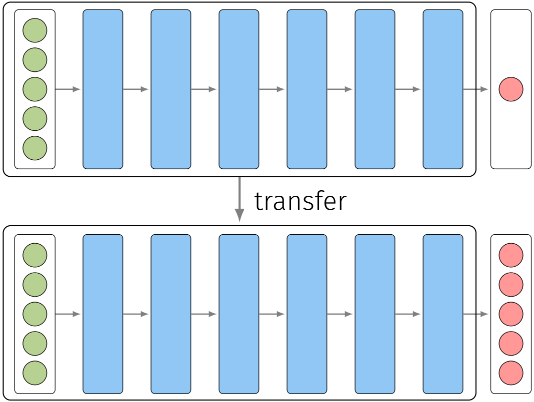
July 5, 2022

Deep Transfer Learning as State-of-the-Art



Deep Transfer Learning as State-of-the-Art

- **Reduce time to select a hyperparameter space**
- **Reduces training time**
- Reduces amount of models to train
- **Avoids overfitting**
- Even obtain **better results**



Features of a DL-Model are Transferable

Transferability is negatively affected by:

- Specialization of higher layer neurons $\rightarrow$ specialization towards target task[1]
- Increasing dataset and/or task dissimilarity[2]

Transfer learning is powerful:

- Initializing a network with transferred features results in a generalization boost that lingers after fine-tuning
- When target training dataset is significantly small
- Training of a large DL model without overfitting

[1] Jason Yosinski et al. "How transferable are features in deep neural networks?" In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., 2014, pp. 3320–3328.

[2] Karl R. Weiss, Taghi M. Khoshgoftaar, and Dingding Wang. "A survey of transfer learning". In: *J. Big Data* 3 (2016), p. 9
Hassan Ismail Fawaz et al. "Transfer learning for time series classification". In: *IEEE International Conference on Big Data, Big Data 2018, Seattle, WA, USA, December 10-13, 2018*. Ed. by Naoki Abe et al. IEEE, 2018, pp. 1367–1376

Abundance of Available DL Models: Which DL Model to Choose?

Numerous ”model repositories” exist, such as:

TensorFlow Hub | Keras | ONNX | ModelDepot | ModelZoo

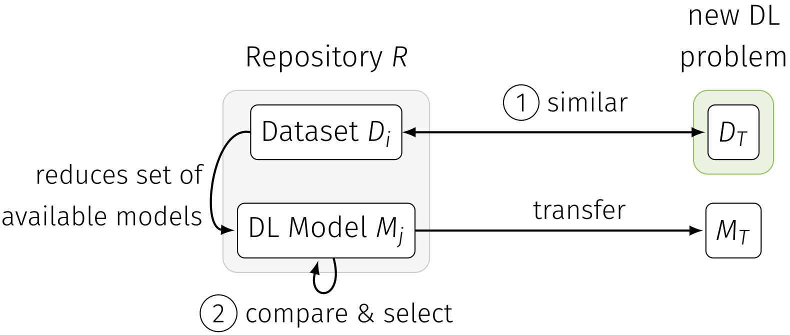
Problem:

An abundance of (freely) available pre-trained DL models complicates selection of an *appropriate* DL model for successful Transfer Learning.

Design and build a **repository** to support the **search and selection** of one deep learning model for **transfer learning** such that, for a given new dataset and its task, the obtained model receives a **better performance measure** than training such a model from scratch.

> **Leading Hypothesis**
> If the training data of a stored model is similar to those of the new dataset, then this stored model is suitable for Transfer Learning.

Transfer Learning Framework



Transfer Learning Rarely Used in a Time Series Modeling Context

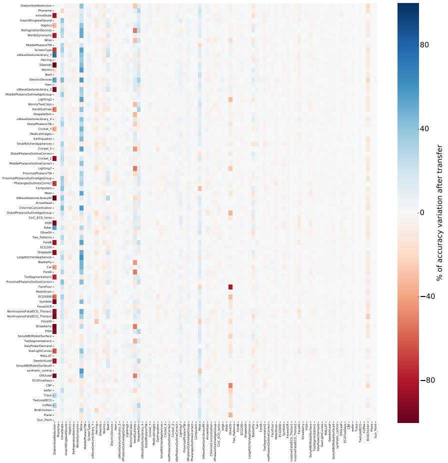
Transfer Learning for *Univariate* Time Series

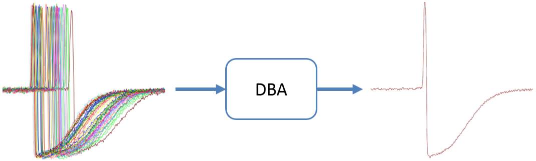
- Ye and Dai, 2018: Time series forecasting. Extract characteristics from univariate time series with extreme learning machine with kernels, train multiple models on these information, create ensemble models.[3]
- He, Pang, and Si: (Financial) time series forecasting. Train DL model on two source datasets.[4]
- Ye and Dai, 2021: Time series forecasting. Find similar time series with DTW and Jensen-Shannon divergence, train CNN on similar time series and then fine-tune on both query and similar time series.[5]

---

[3] Rui Ye and Qun Dai. "A novel transfer learning framework for time series forecasting". In: *Knowl. Based Syst.* 156 (2018), pp. 74–99.
[4] Qi-Qiao He, Patrick Cheong-Iao Pang, and Yain-Whar Si. "Transfer Learning for Financial Time Series Forecasting". In: *PRICAI 2019: Trends in Artificial Intelligence - 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, August 26-30, 2019, Proceedings, Part II.* ed. by Abhaya C. Nayak and Alok Sharma. Vol. 11671. Lecture Notes in Computer Science. Springer, 2019, pp. 24–36.
[5] Rui Ye and Qun Dai. "Implementing transfer learning across different datasets for time series forecasting". In: *Pattern Recognit.* 109 (2021), p. 107617.

Fawaz et al.: Transfer Learning for Time Series Classification[6]





- 85 univariate time series + CNN
- Brute force Transfer Learning
- Replace output layer with new softmax layer and fine-tune
- 71/85 better with similarity
- Nearest neighbor not always best, sometimes 2nd nearest neighbor

[6] Hassan Ismail Fawaz et al. "Transfer learning for time series classification". In: *IEEE International Conference on Big Data, Big Data 2018, Seattle, WA, USA, December 10-13, 2018*. Ed. by Naoki Abe et al. IEEE, 2018, pp. 1367–1376.

Searching Datasets for Transfer Learning

- Natural Language Processing (NLP) such as sentiment analysis, part-of-speech tagging, and named entity recognition. [RP17; Akd20; SLE18; LM20; Lin+19; Bä18]
  - [Lin+19]: Extract textual features such as type-token ratio, word overlap, phonological distance; train ensemble of decision-trees; select high-resource language for transfer learning for a low-resource language.
- Image datasets: [Bha+20; Sch+18; BEF19; Pra+19]
  - Compute dataset difficulty of dataset to search for datasets with similar difficulty [Sch+18; Ist+19]
  - [Pra+19]: Propose *dataset similarity ranker* for image datasets. Extract image related features, train several classifiers to predict which feature vector belongs to which dataset, enemble predictions.
- Extracted features depend on type of dataset (text vs. image). Various similarity functions used.

Data Set Characteristics for Similarity

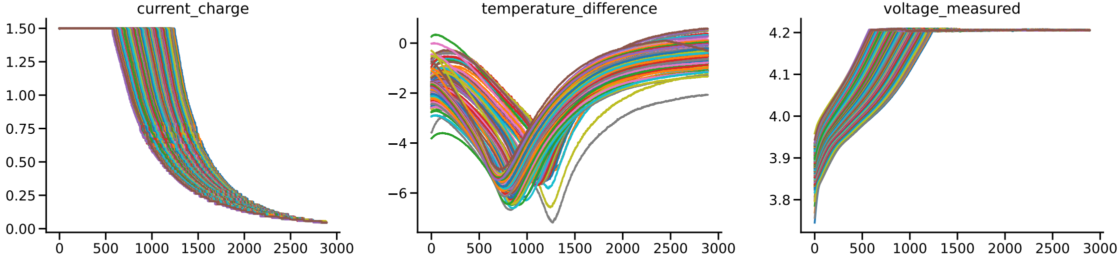
Dataset Used: NASA's Lithium Ion Battery Dataset

- NASA's Li-Ion battery dataset consisting of 34 data sets[7]

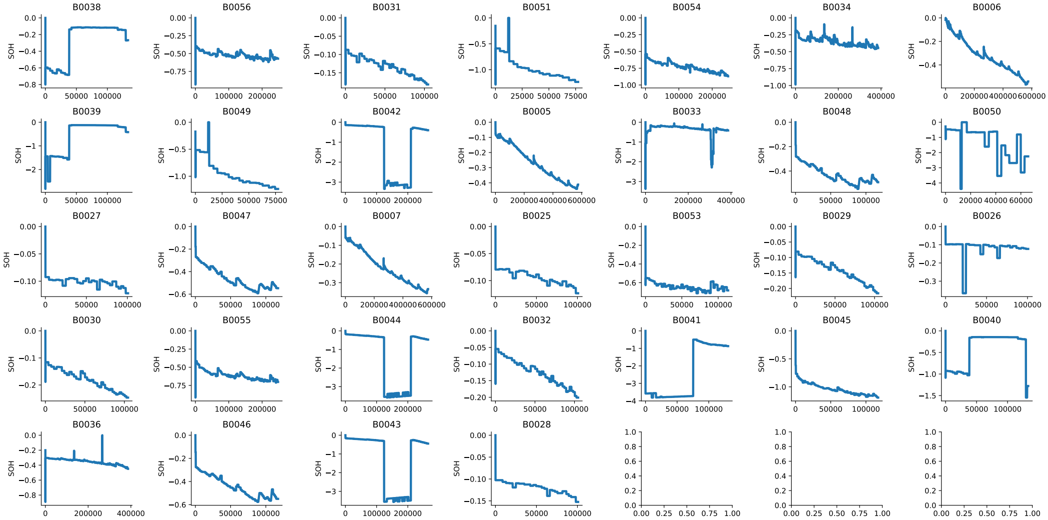
- Battery EOL after losing 30% of its capacity.
- Datasets can be used for Predictive Maintenance and Remaining Useful Life (RUL) prediction.
- *State-of-Health* (SOH) is commonly used for RUL prediction. $\frac{\text{current capacity}}{\text{current capacity}}\%$

[7] Bhaskar Saha and Kai Goebel. *Battery Data Set*. NASA Ames Prognostics Data Repository. NASA Ames Research Center, Moffett Field, CA. 2007. URL: `http://ti.arc.nasa.gov/project/prognostic-data-repository` (visited on 05/09/2022).

Dataset Used: NASA's Lithium Ion Battery Dataset



Lithium Ion Batteries State-of-Health of Each Battery



Naive Approach to Compute Time Series Similarity

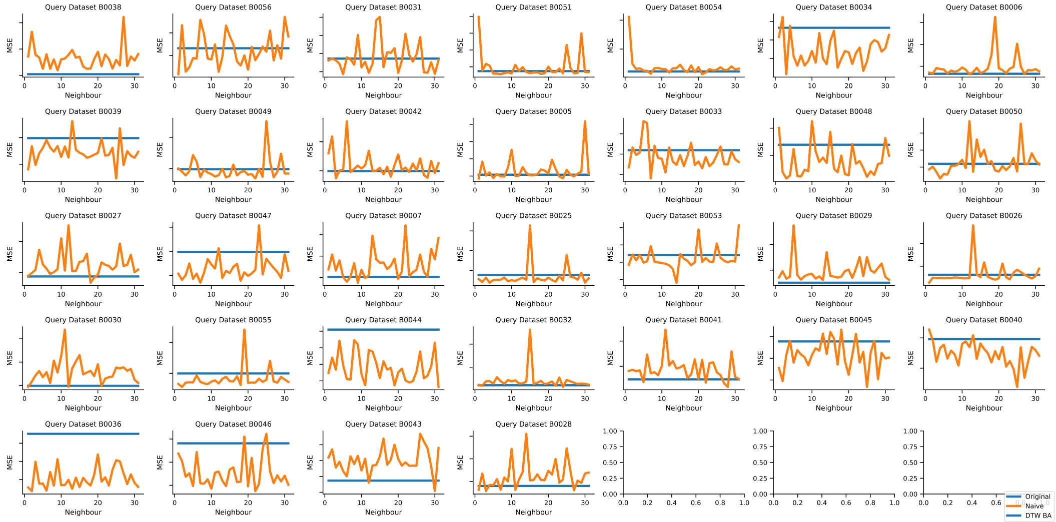
- Descriptive statistics: minimum, maximum, mean, median, variance, standard deviation, quantiles (0.05, 0.25, 0.75, 0.95)
- Use Euclidean distance to order datasets.

Build Transfer Ground-Truth:

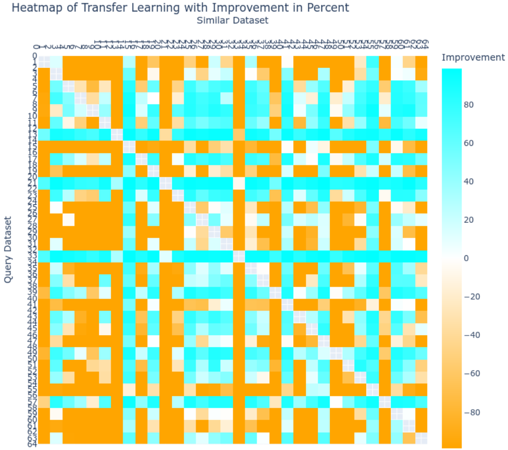
- Train LSTM on each individual battery dataset.
- Perform brute force transfer learning similarly to Fawaz et. al.[8]
- *Transfer DL-Model with Modification:* Re-initialize last layer (output layer) and retrain whole network.

[8] Hassan Ismail Fawaz et al. "Transfer learning for time series classification". In: *IEEE International Conference on Big Data, Big Data 2018, Seattle, WA, USA, December 10-13, 2018*. Ed. by Naoki Abe et al. IEEE, 2018, pp. 1367–1376.

Transfer Results: Naive Approach to Time Series Similarity



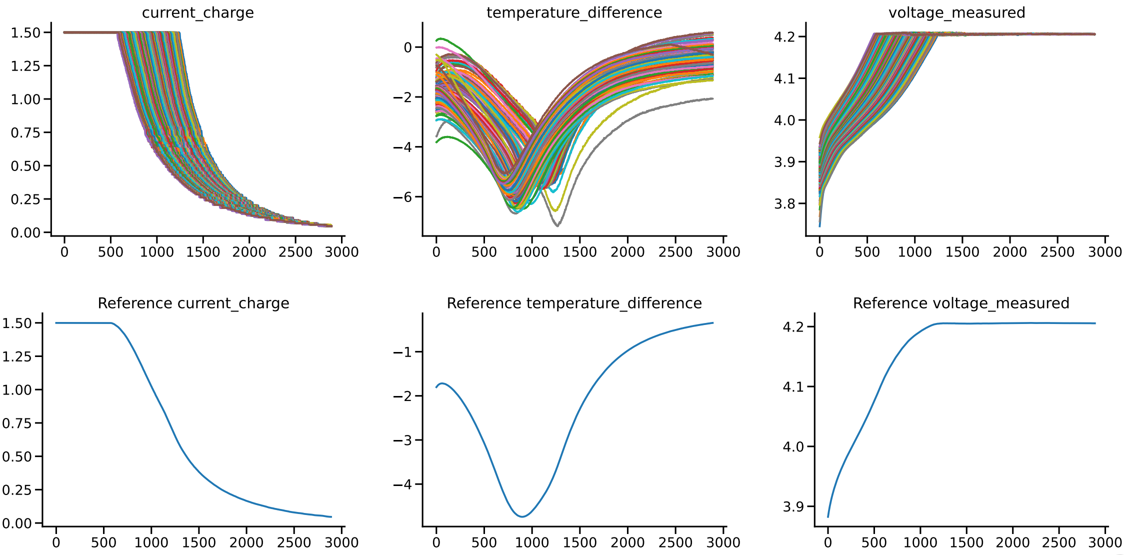
Transfer Results: Naive Approach to Time Series Similarity



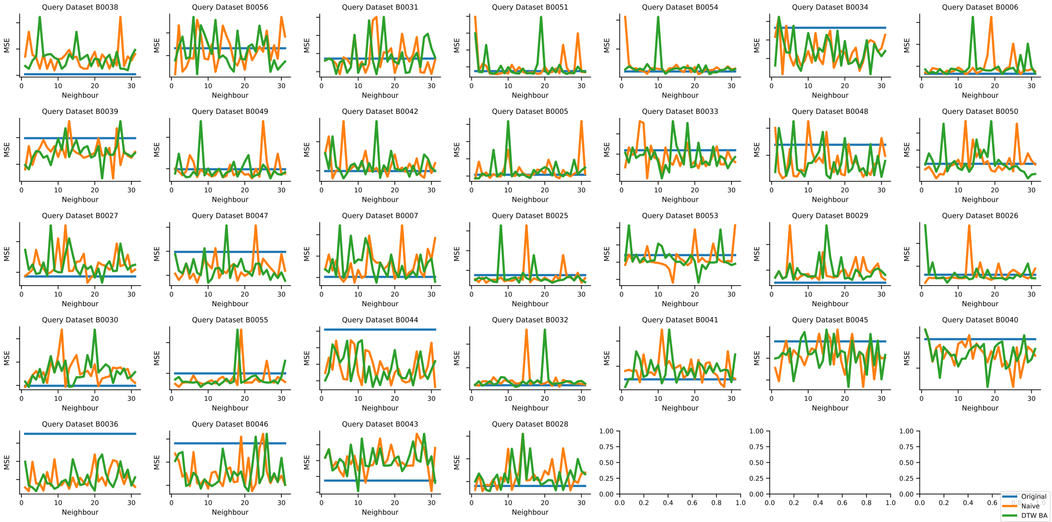
Summary

- Transfer learning shows positive improvement in some cases.
- Other cases show negative transferability.
- First nearest neighbour usually performs poorly.
- Second nearest neigbour also performs poorly.
- Distributional statistics not sufficient.

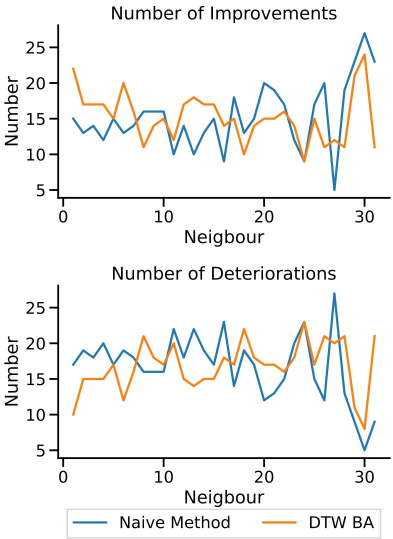
Create Reference Time Series Using DTW with Barycenter Averaging



Transfer Results: Naive Approach vs. DTW Barycenter Averaging



Transfer Results: Naive Approach vs. DTW Barycenter Averaging



Summary

- DTW Barycenter Averaging generally performs better (first five neigbours)
- Interesting: Farthest neigbours of naive approach performs better
- Some datasets notoriously successful, whereas some always yield negative transfer results

Ways Forward to Improve Naive Similarity

- Extract **time-series specific characteristics** such as trend, seasonality, chaos, skewness, and kurtosis. Thus, extracting similar characteristics as Wang, Smith, and Hyndman[9].
- **Extract further characteristics** as Fulcher and Jones, albeit maybe not as much as they did (they extracted over 9,000 features).[10]

---

[9] Xiaozhe Wang, Kate Smith, and Rob Hyndman. "Characteristic-Based Clustering for Time Series Data". In: *Data Mining and Knowledge Discovery* 13.3 (May 2006), pp. 335–364.
[10] Ben D Fulcher and Nick S Jones. "Highly comparative feature-based time-series classification". In: *IEEE Transactions on Knowledge and Data Engineering* 26.12 (2014), pp. 3026–3037.

Ways Forward to Improve DTW Barycenter Averaging Similarity

- Currently: Sum DTW distances of each dataset attribute.
- Find similar dataset similarity based on **individual attribute similarities**.
- DTW Barycenter Averaging method yields a reference time series that may not reflect the variability of a set of time series. Extend it with a **confidence interval**.

Summary and Ways Forward

Summary and Ways Forward

- Improve naive similarity approach by using dedicated time series characteristics
- Improve DTW Barycenter Averaging approach by extending to search for a similar attribute among all available attributes of another dataset; extend it by introducing a confidence interval.
- Use datasets from another domain such as NASA's turbofan engine degradation dataset
- Extend current regression task to time series classification

Friedrich-Alexander-Universität
Technische Fakultät

Chair for Computer Science 6 (Data Management)

Deep Transfer Learning Model Selection

In A Time Series Modeling Context

Melanie Bianca Sigl
melanie.sigl@fau.de

July 5, 2022

Appendix

Bibliography i

[Akd20] Arda Akdemir. “Research on Task Discovery for Transfer Learning in Deep Neural Networks”. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, ACL 2020, Online, July 5-10, 2020*. Ed. by Shruti Rijhwani et al. Association for Computational Linguistics, 2020, pp. 33–41.

[BEF19] Kevin Bascol, Rémi Emonet, and Elisa Fromont. “Improving Domain Adaptation By Source Selection”. In: *ICIP 2019 - IEEE International Conference on Image Processing*. Taïpei, Taiwan: IEEE, Sept. 2019.

[Bha+20] Bishwaranjan Bhattacharjee et al. “P2L: Predicting Transfer Learning for Images and Semantic Relations”. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020, Seattle, WA, USA, June 14-19, 2020*. IEEE, 2020, pp. 3284–3293.

[Bä18] Jesper Bäck. “Domain similarity metrics for predicting transfer learning performance”. PhD thesis. Linköping University, 2018.

[Faw+18] Hassan Ismail Fawaz et al. “Transfer learning for time series classification”. In: *IEEE International Conference on Big Data, Big Data 2018, Seattle, WA, USA, December 10-13, 2018*. Ed. by Naoki Abe et al. IEEE, 2018, pp. 1367–1376.

[FJ14] Ben D Fulcher and Nick S Jones. “Highly comparative feature-based time-series classification”. In: *IEEE Transactions on Knowledge and Data Engineering* 26.12 (2014), pp. 3026–3037.

[HPS19] Qi-Qiao He, Patrick Cheong-Iao Pang, and Yain-Whar Si. “Transfer Learning for Financial Time Series Forecasting”. In: *PRICAI 2019: Trends in Artificial Intelligence - 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, August 26-30, 2019, Proceedings, Part II*. Ed. by Abhaya C. Nayak and Alok Sharma. Vol. 11671. Lecture Notes in Computer Science. Springer, 2019, pp. 24–36.

[Ist+19] Roxana Istrate et al. “TAPAS: Train-Less Accuracy Predictor for Architecture Search”. In: *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019, pp. 3927–3934.

[Lin+19] Yu-Hsiang Lin et al. “Choosing Transfer Languages for Cross-Lingual Learning”. In: *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28-August 2, 2019, Volume 1: Long Papers*. Ed. by Anna Korhonen, David R. Traum, and Lluís Màrquez. Association for Computational Linguistics, 2019, pp. 3125–3135.

[LM20] Samuel Louvan and Bernardo Magnini. “How Far Can We Go with Data Selection? A Case Study on Semantic Sequence Tagging Tasks”. In: *Proceedings of the First Workshop on Insights from Negative Results in NLP, Insights 2020, Online, November 19, 2020*. Ed. by Anna Rogers, João Sedoc, and Anna Rumshisky. Association for Computational Linguistics, 2020, pp. 15–21.

[Pra+19] Ameya Prabhu et al. “”You might also like this model”: Data Driven Approach for Recommending Deep Learning Models for Unknown Image Datasets”. In: *CoRR* abs/1911.11433 (2019). arXiv: `1911.11433`.

[RP17] Sebastian Ruder and Barbara Plank. “Learning to select data for transfer learning with Bayesian Optimization”. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*. Ed. by Martha Palmer, Rebecca Hwa, and Sebastian Riedel. Association for Computational Linguistics, 2017, pp. 372–382.

[SG07] Bhaskar Saha and Kai Goebel. *Battery Data Set*. NASA Ames Prognostics Data Repository. NASA Ames Research Center, Moffett Field, CA. 2007. URL: `http://ti.arc.nasa.gov/project/prognostic-data-repository` (visited on 05/09/2022).

[Sch+18] Florian Scheidegger et al. "Efficient Image Dataset Classification Difficulty Estimation for Predicting Deep-Learning Accuracy". In: *CoRR* abs/1803.09588 (2018). arXiv: `1803.09588`.

[SLE18] Lex Razoux Schultz, Marco Loog, and Peyman Mohajerin Esfahani. "Distance Based Source Domain Selection for Sentiment Classification". In: *CoRR* abs/1808.09271 (2018). arXiv: `1808.09271`.

[WSH06] Xiaozhe Wang, Kate Smith, and Rob Hyndman. "Characteristic-Based Clustering for Time Series Data". In: *Data Mining and Knowledge Discovery* 13.3 (May 2006), pp. 335–364.

[WKW16] Karl R. Weiss, Taghi M. Khoshgoftaar, and Dingding Wang. "A survey of transfer learning". In: *J. Big Data* 3 (2016), p. 9.

[YD18] Rui Ye and Qun Dai. "A novel transfer learning framework for time series forecasting". In: *Knowl. Based Syst.* 156 (2018), pp. 74–99.

[YD21] Rui Ye and Qun Dai. “Implementing transfer learning across different datasets for time series forecasting”. In: *Pattern Recognit.* 109 (2021), p. 107617.

[Yos+14] Jason Yosinski et al. “How transferable are features in deep neural networks?” In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., 2014, pp. 3320–3328.