Identifying Suitable Tasks for Inductive Transfer Through the Analysis of Feature Attributions

Alexander J. Hepburn and Richard McCreadie

School of Computing Science, University of Glasgow, Glasgow, G12 8QQ, United Kingdom

Introduction

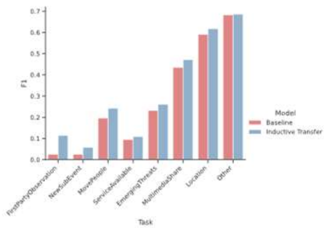
- **Transfer learning** approaches have shown to significantly improve performance on downstream tasks, however, finding effective task combinations often requires extensive brute-force searches.
- Can we, then, predict whether transfer between two tasks will be beneficial **without actually performing the experiment?**

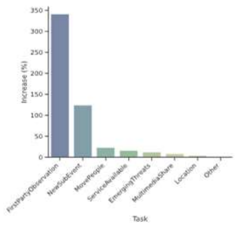
Aim

This work aims to demonstrate that there exists correlation between the shared linguistic properties of task pairs and their combined performance output, and that this prior knowledge can be leveraged to **dramatically reduce the time taken to find effective task combinations**.

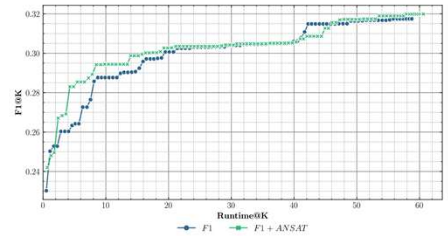
Analysing Cross-task Active Terms to Predict Performance Output

- **8 out of 12** tasks benefit from applying transfer learning.
- We observe a **12.4%** increase in F1-score performance over those 8 tasks.
- The largest increase in performance is **341%** for the **First Party Observation** category.
- Unexpectedly, we also find gains in performance in which the training examples for the target task outnumber those for the source task.



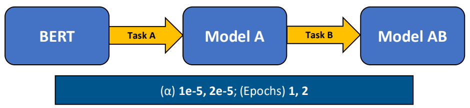


- Training all models takes **60.6 hours** and achieves an F1-score of **0.3199**.
- Using our regression model, we are able to achieve **0.3003** (only **6.12% worse** than our best) at **30 hours** or **50.5%** less training time.
- If we were to accept an F1-score of **0.2859** (**10.78%** reduction), we can further reduce our time to **10 hours** or by **83.5%**.



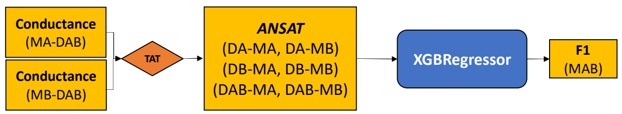
Methods

Improving Performance Through Inductive Transfer



- We decompose a multi-label dataset into **12** binary classification problems.
- Using BERT as a base model, we fine-tune each task with **4** separate hyperparameter combinations, creating **896** models.
- We observe the effects of transfer learning between these task combinations and note performance change.

Optimising Transfer Learning with Feature Attributions



- We calculate post-hoc, attribution-based **conductance** scores on Model A, B using auxiliary models.
- We use different **threshold values (TAT)** to determine *term activity* and calculate the **Average Number of Shared Active Terms (ANSAT)**.
- We train an XGBoost regression model on our activity-based features to predict their **combined F1-score** and use these scores as a ranking of the best-performing task pairs.

Conclusions

- There is **clearly significant scope** for improving performance by leveraging attribution-based techniques.
- Computing conductance has **significant computational overhead** and quickly becomes impractical to implement.
- ANSAT, as a method of dataset representation comparison, requires further work to **increase the accuracy** of its estimations.
- Further work is required to **test the generalisability** of our approach.

References

Pan, S.J., Yang, Q.: A survey on transfer learning. (2010)
Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep networks. (2017)
Dhamdhere, K., Sundararajan, M., Yan, Q: How important is a neuron? (2018)