MOOD CLASSIFICATION USING LISTENING DATA

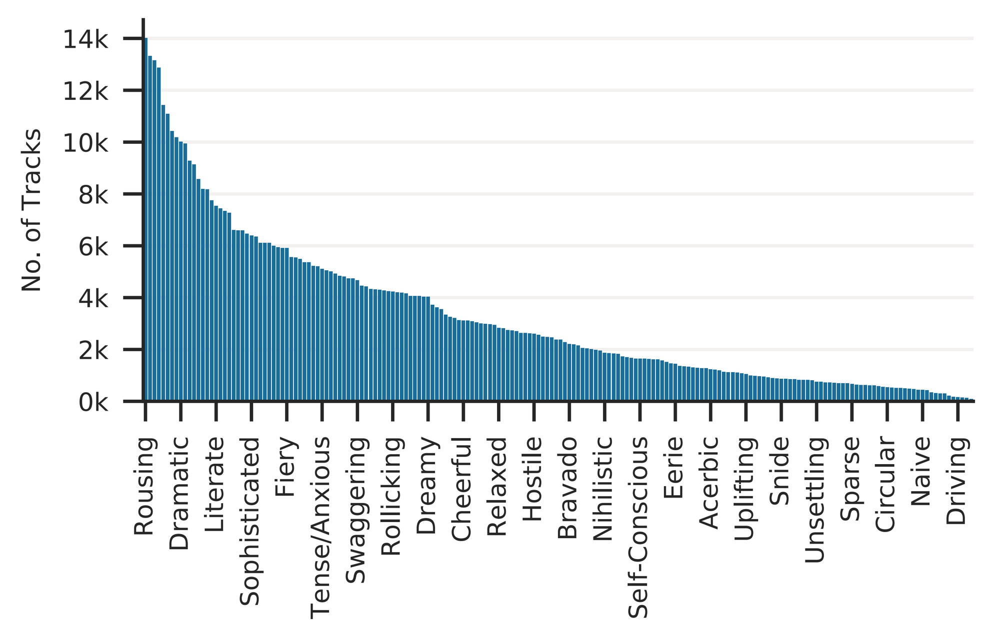
F. Korzeniowski[1], O. Nieto[1], M. McCallum[1], M. Won[2], S. Oramas[1], E. Schmidt[3]
(the first two authors contributed equally to this work)

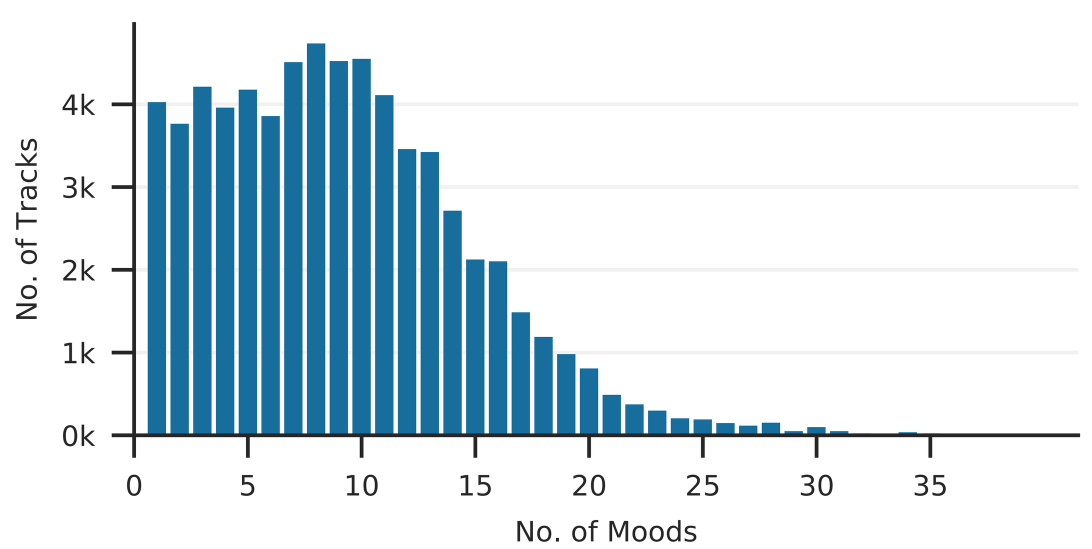
[1]Pandora Media LLC., Oakland, USA
[2]Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[3]Netflix Inc., Los Gatos, USA

AllMusic Mood Subset

A NEW DATASET FOR LARGE-VOCABULARY MOOD CLASSIFICATION

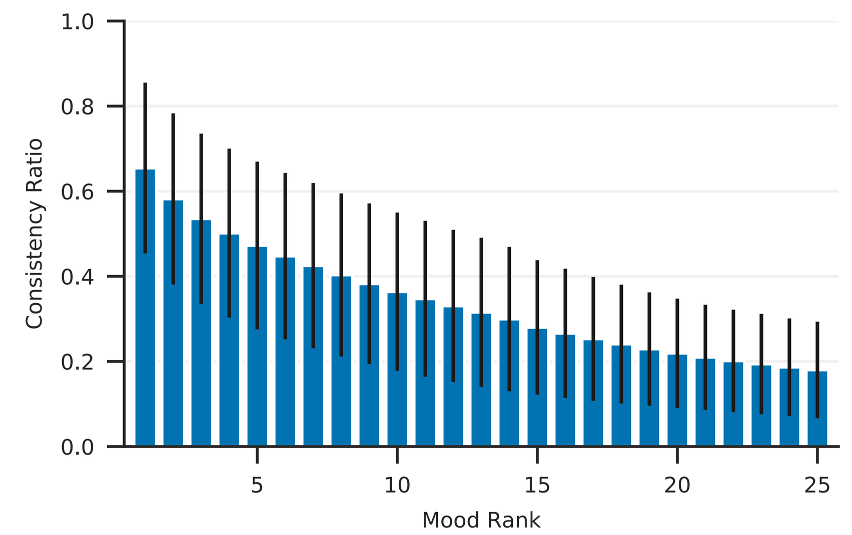
- Links the **Million Song Dataset** to AllMusic moods.
- **66993 tracks** in total.
- **188** different **mood tags.**
- Echo Nest Taste Profile provides **listening data.**
- 7-digital audio previews provide **audio data.**
- AllMusic **mood data** is proprietary, but freely available at *allmusic.com*
- Moods are annotated on an **album level** and propagated down to tracks.





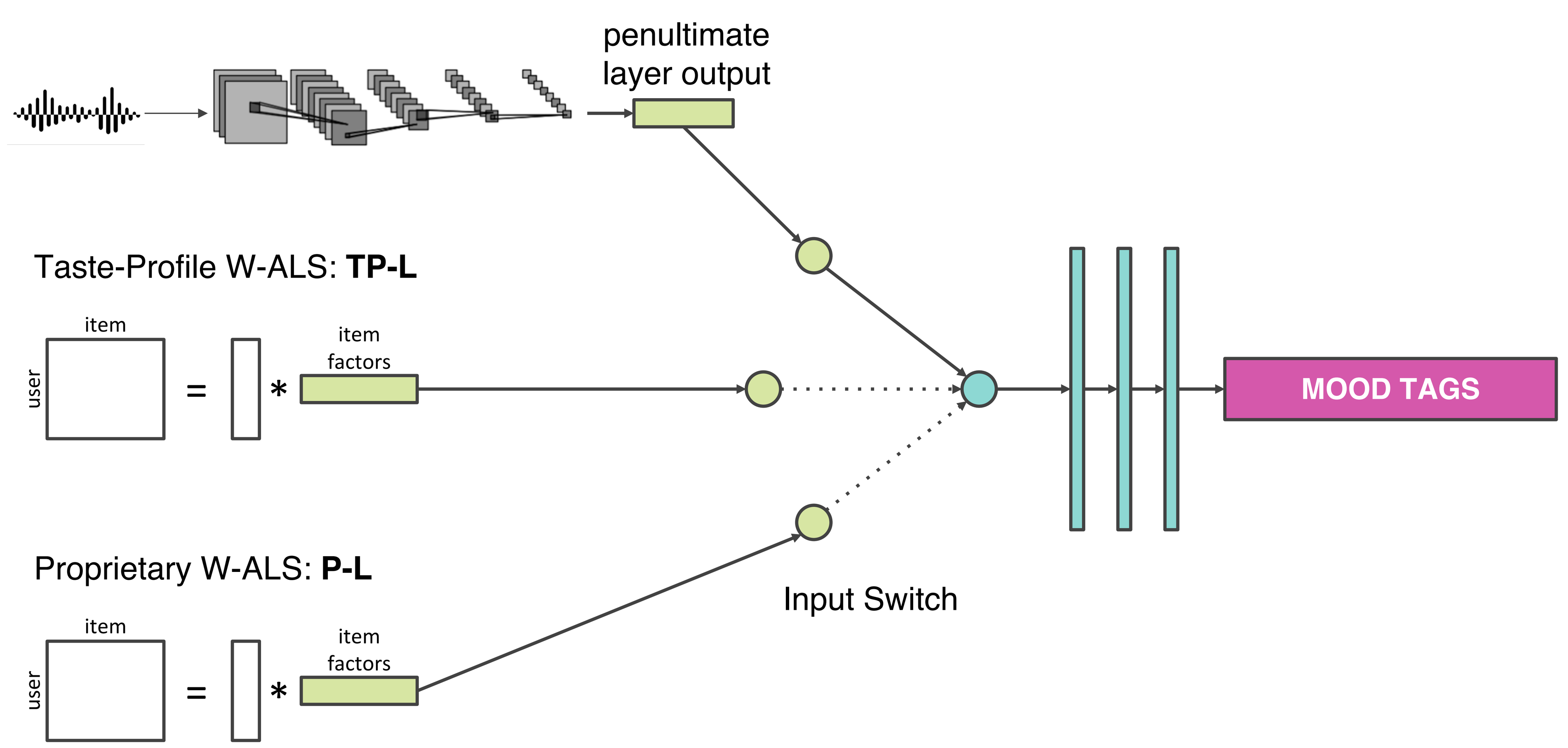
Listening Data for Mood Classification

- The **Echo Nest Taste Profile:**
  - **28M play counts** of **1M users** and **384k tracks**.
- **Observation**: users listen to music **consistent in mood**: **65.4%** of all listened tracks by a user contain the **most popular mood tag** for a user.
- **Thus**: listening-based embeddings could be powerful features for mood classification!

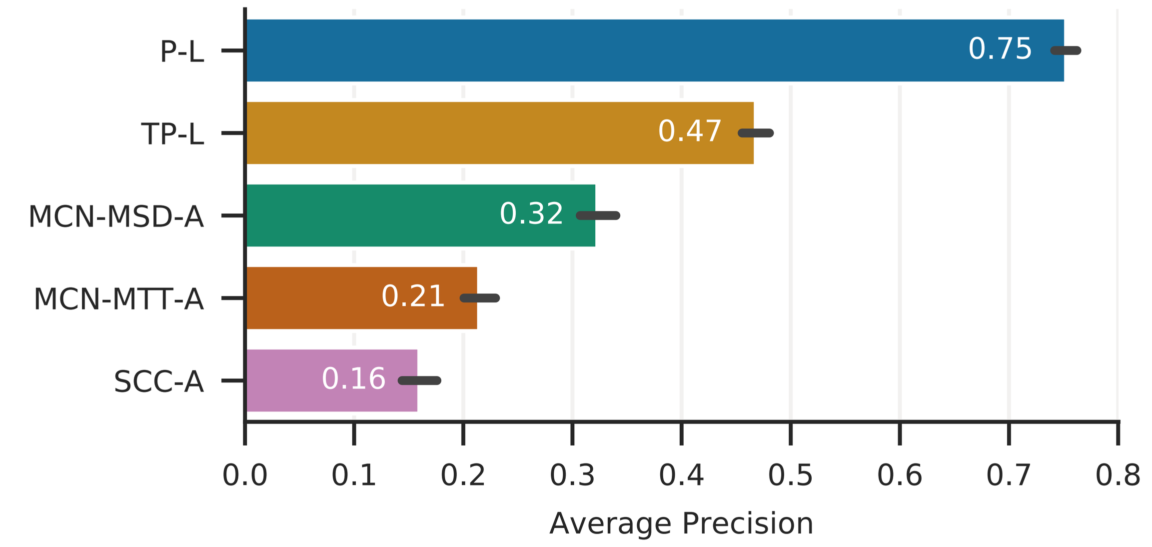


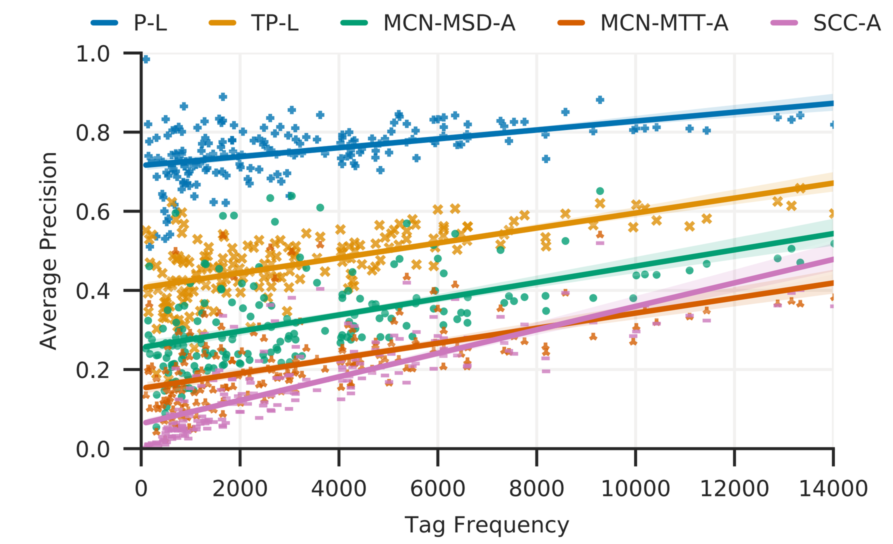
Experiment: Transfer Learning

- **Audio-based embeddings:**
  - **MCN-MSD-A**: Musicnn trained on Million Song dataset
  - **MCN-MTT-A**: Musicnn trained on MagnaTagATune dataset
- **Listening-based embeddings:**
  - **TP-L:** weighted ALS of Taste Profile interaction matrix (implicit feedback)
  - **P-L:** proprietary ALS of in-house explicit feedback (> 100B thumbs)

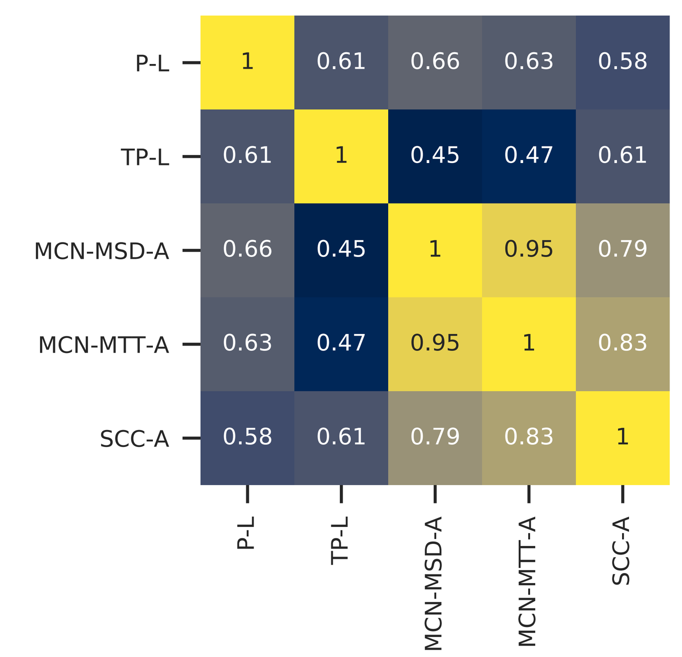


Result: Listening Data Outperforms Audio Data



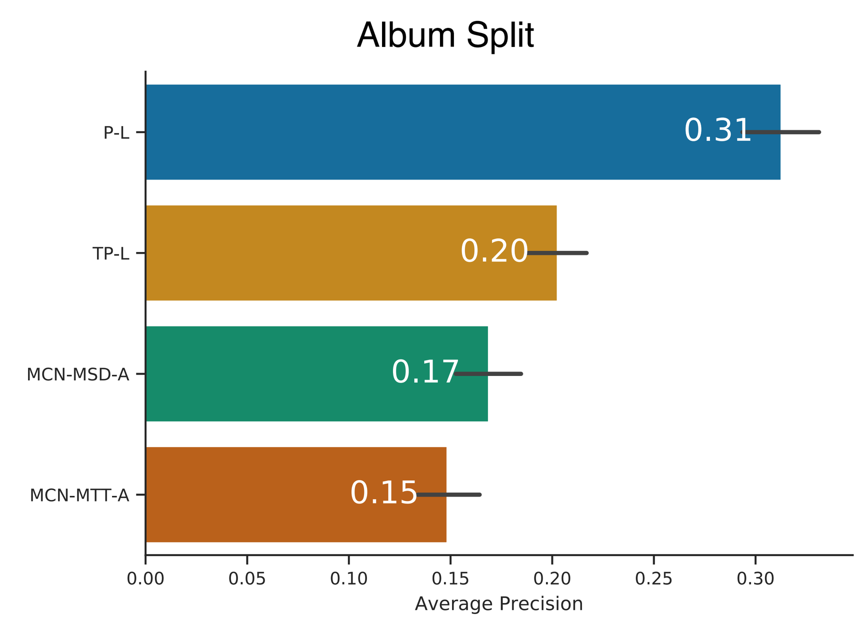


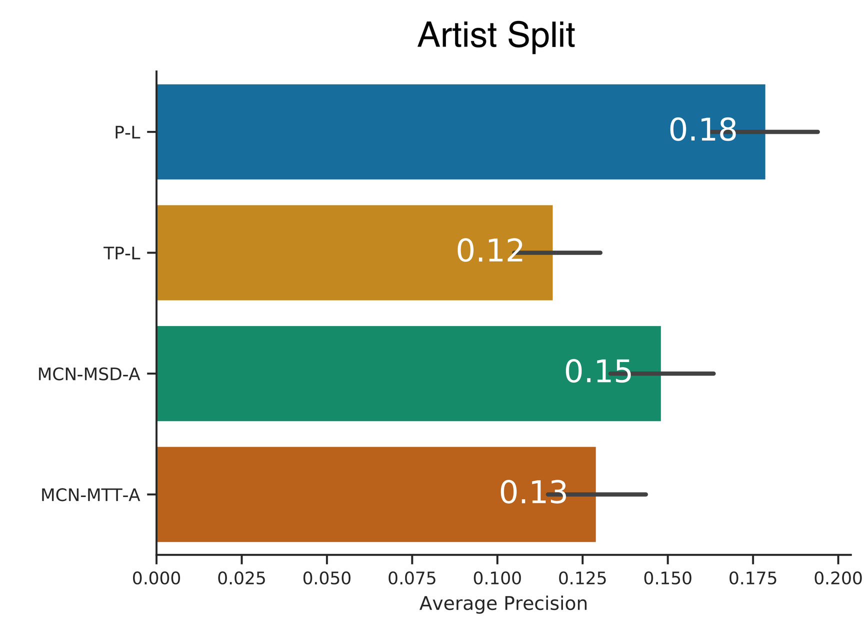
- **Transfer learning** from models trained on big data **helps with rare tags**.
- Massive **explicit feedback better than implicit.**
- **Tag-wise results correlate** between audio-based models, regardless of absolute results.
- **Implicit vs explicit** feedback seem to capture **different listener behavioral data**.



Chasing Confounding Factors: More Splits, More Results

- Split data **by album and by artist** and re-run experiments.





- **Album-split lowers results significantly**, more so for listening-based than for audio-based embeddings.
- **Artist-split further reduces results**, **Taste-Profile** based embeddings **lose their advantage** over audio-based ones.
- **Conclusion:** Taste-Profile seems to take advantage of mood-artist correlation to achieve results; internal **explicit feedback** data **performs best regardless**.

Code and Data: https://github.com/fdlm/listening-moods