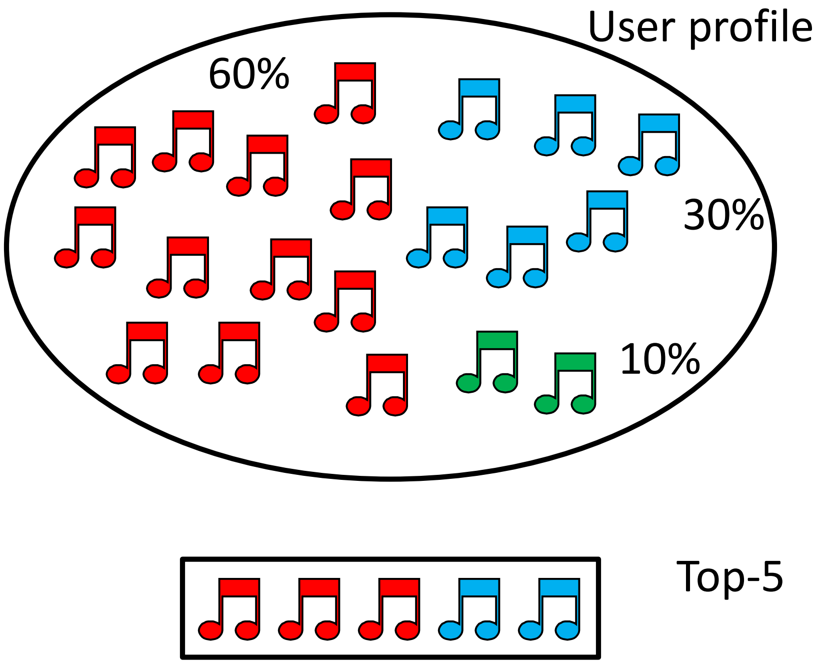
A Comparison of Calibrated and Intent-Aware Recommendations

Mesut Kaya
TU Delft, Netherlands

Derek Bridge,
University College Cork, Ireland

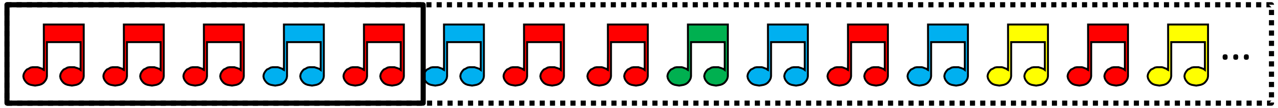
Calibrated Recommendations

- The top-*n* recommendations from a model trained for accuracy may
  - focus on the user's main interests
  - leaving lesser interests under-represented or absent
- Calibrated recommendations [Steck 2018]
  - the goal is to reflect the various interests of the user in the appropriate proportions



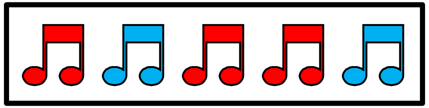
Calibrated Recommendations

- Candidate items, ordered for relevance $s(u, i)$ by a baseline recommender



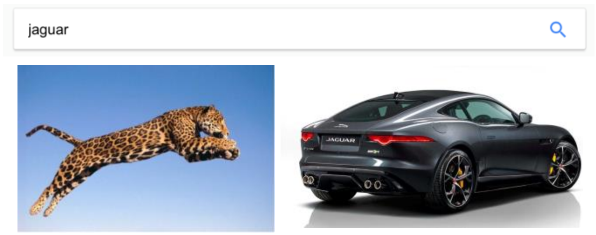
- Greedily re-rank by a linear combination

$$f_{obj}(i, RL) = \overbrace{(1 - \lambda)s(u, i)}^{\text{Relevance}} + \overbrace{\lambda\, cal(i, RL)}^{\text{Calibration}} \qquad 0 \le \lambda \le 1$$



Intent-Aware Diversification

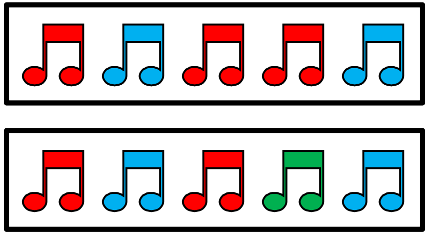
- Early work on diversification
  - greedily re-rank to decrease similarity of items in top-*n* [Carbonell & Goldstein 1998, Ziegler et al 2005]
- Intent-aware diversification from IR [Santos et at. 2010]
  - top-*n* contains results for each interpretation of an ambiguous query



- Intent-aware diversification for recommenders [Vargas 2015]

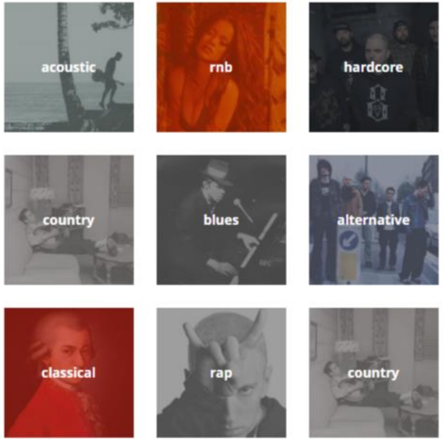
$$f_{obj}(i, RL) = (1-\lambda)s(u,i) + \lambda\, div_{IA}(i, RL)$$

  - top-*n* contains results for each of the user's interests (from her profile)
  - this is like calibration but formulated in a different way that *takes relevance into account* as well



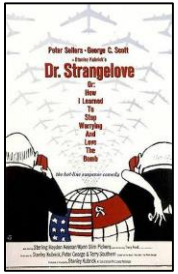
User Interests

- In [Steck 2018] and [Vargas 2015]
  - user interests are defined by item features, e.g. genres, in the user's profile

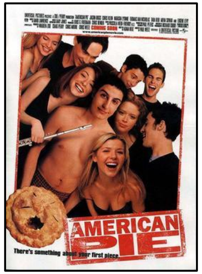


[Image from Last.FM]

- But item features
  - are not available in every domain
  - are often noisy and inconsistently applied
  - may not describe subjective tastes

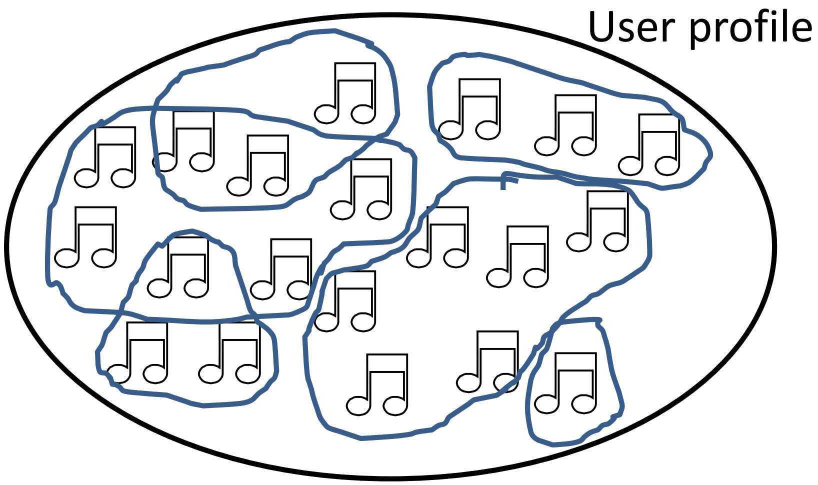


"comedy"



User Interests

- In our work, we define user interests as *subprofiles*



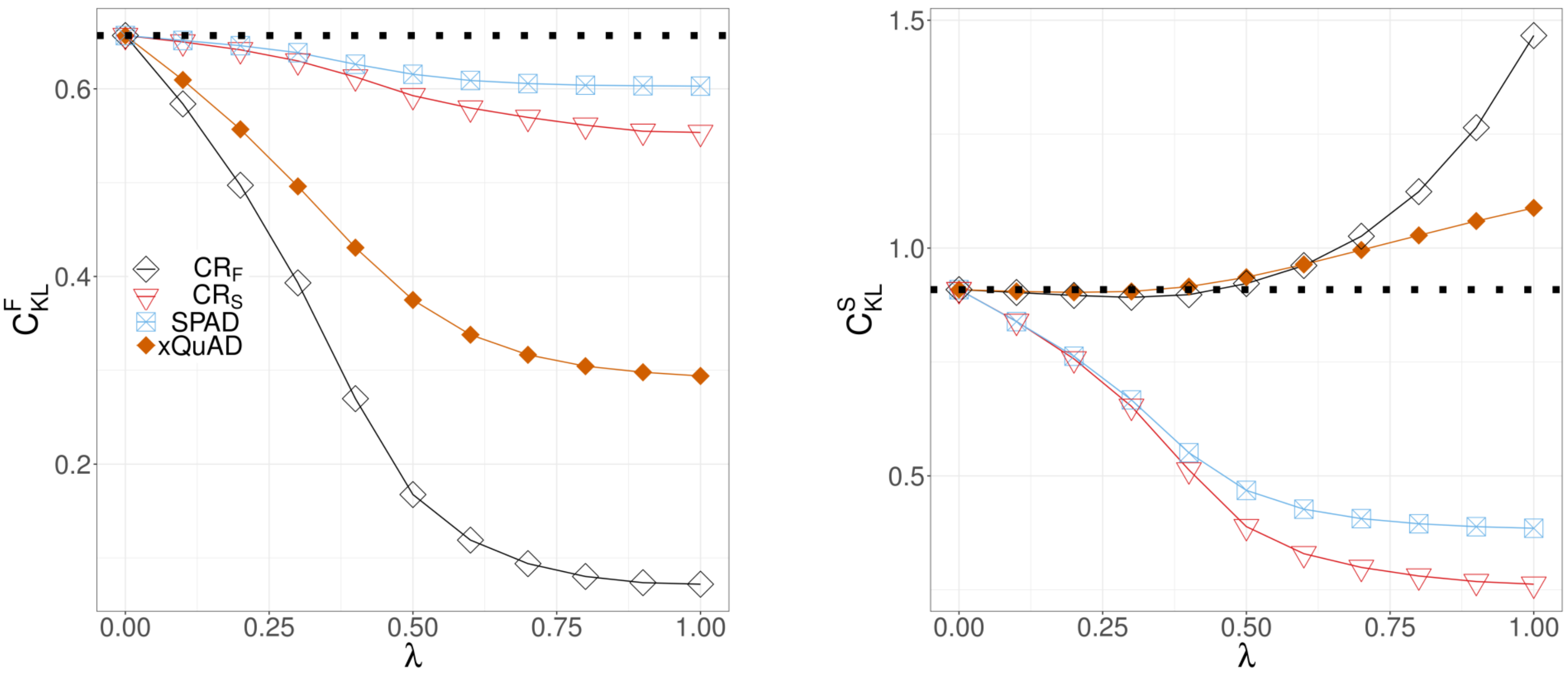
- Members of a subprofile
  - have similar interactions/ratings
  - no use of item features

×		
	×	×
		×
×	×	×
	×	×

Summary So Far

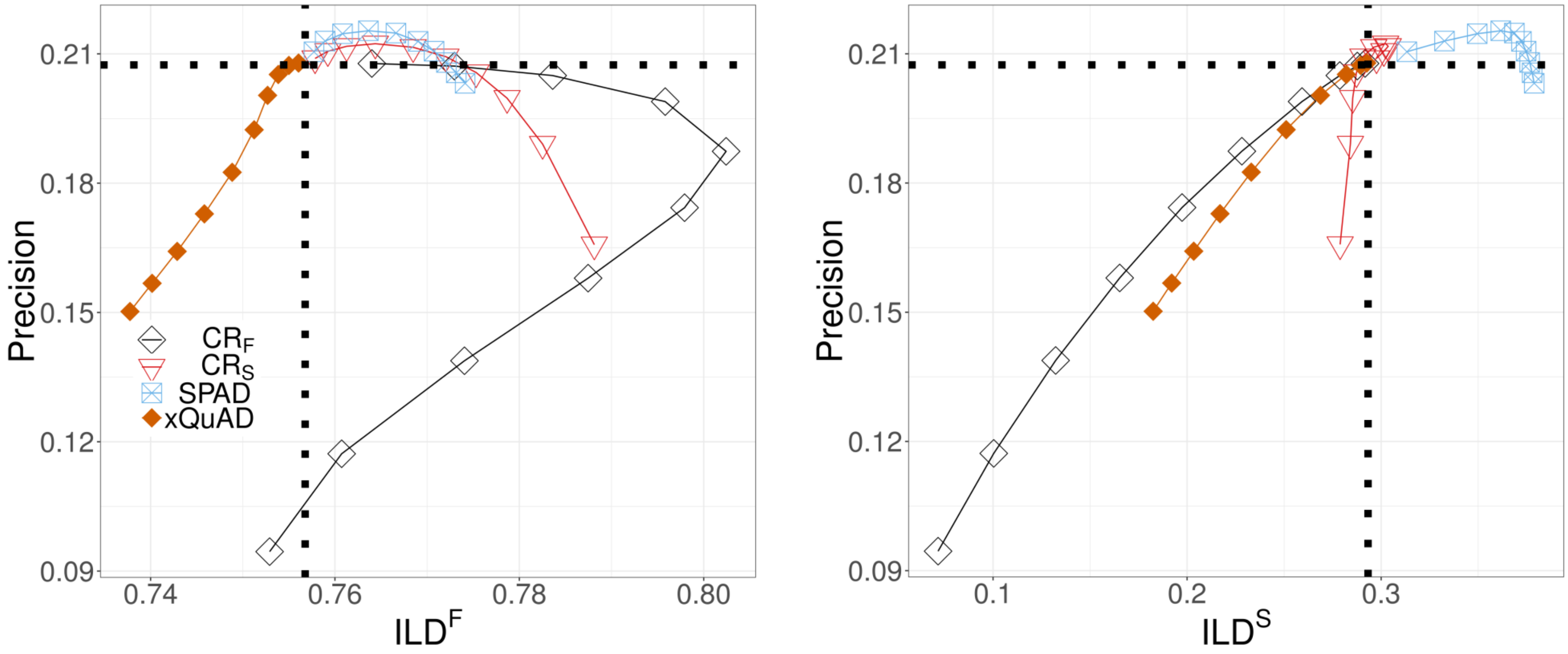
	User interests from item features	User interests from subprofiles
Calibrated Recommenders	CR_F [Steck 2018]	CR_S
Intent-Aware Recommenders	xQuAD [Vargas 2015]	SPAD

Calibration



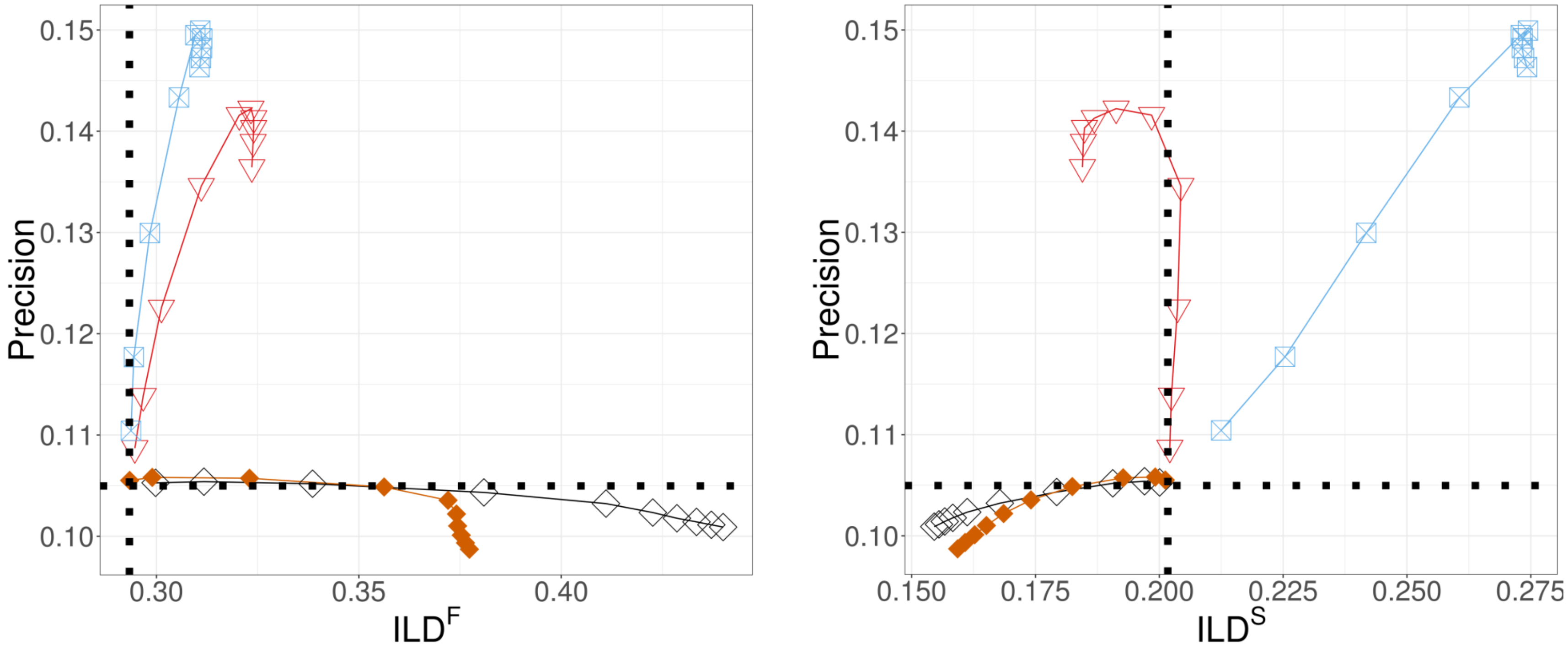
Implicit ratings version of MovieLens 20 Million Dataset

Precision versus Diversity



Implicit ratings version of MovieLens 20 Million Dataset

Precision versus Diversity



Implicit ratings version of TasteProfile Dataset

Concluding Remarks

- On these datasets, the approaches that use subprofiles (CR_S and SPAD) achieve
  - the highest precision
  - better than baseline calibration according to both calibration metrics
  - good diversity according to α-nDCG (see paper)
- SPAD also achieves
  - better than baseline diversity according to both ILD metrics
  - suffers least from the relevance/ diversity trade-off

- Future directions
  - investigate how users perceive calibrated/ diversified recommendations
  - apply subprofiles to tasks other than calibration/ diversification
  - develop the idea that calibration in general (and these approaches in particular) could be used for fairness in recommendations

Thanks

(I’m hiring)