Make the most of ML

without growing the team

Claudiu Musat
Research Director

Data Analytics and AI
Swisscom

swisscom

Things change.

Data rich

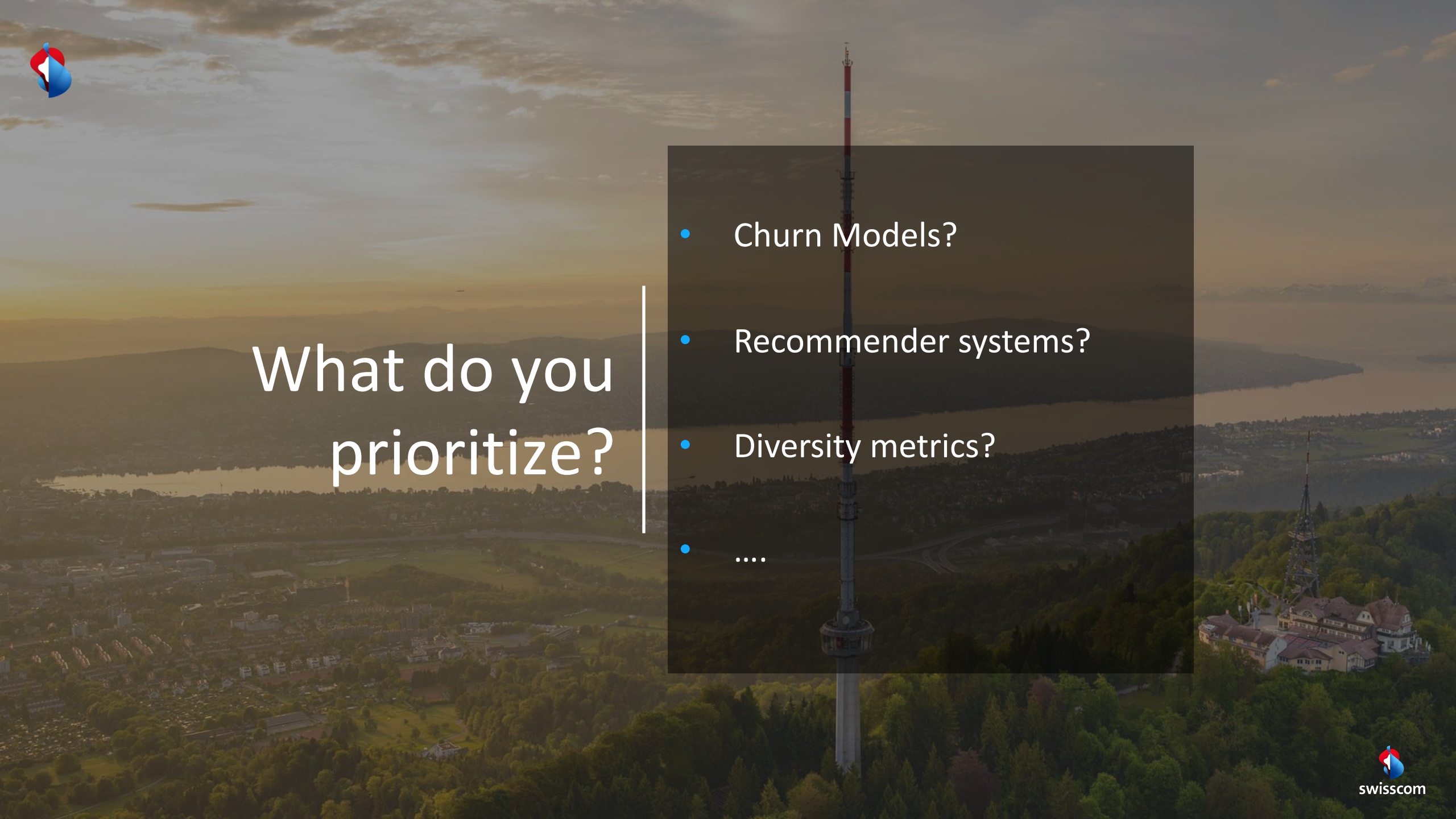
Interaction rich

Small data science team.


back
home
guide
ok
option
vol
txt
radio

Bringing down the silo walls

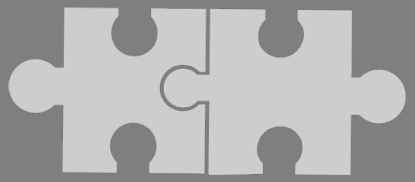
swisscom


What do you prioritize?
Churn Models?
Recommender systems?
Diversity metrics?
….
swisscom

What else is this data good for?

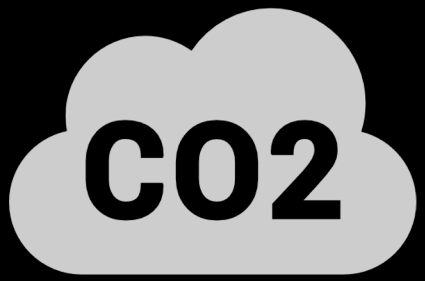
How can I integrate constraints and objectives?

How can I use all the relevant knowledge?



swisscom

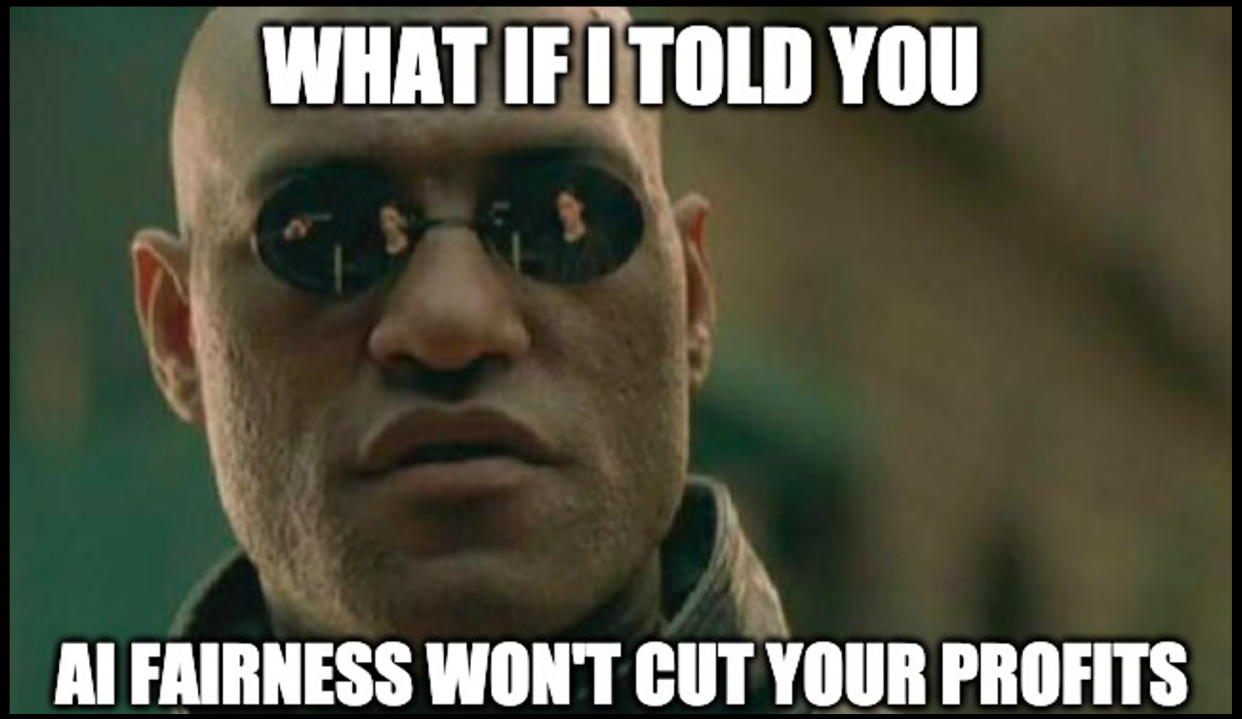
Data Reusability



From anonymized network data to smart cities and CO2 tracking

swisscom

A multitude of objectives



Accuracy

Novelty

Revenue

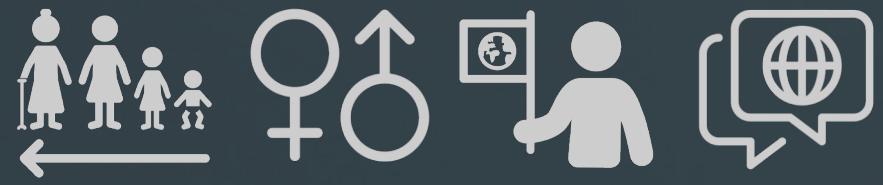
Diversity

Fairness

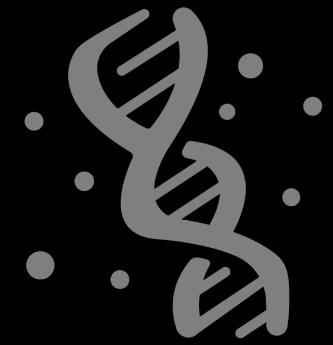
From wish to action

- Strategic goals

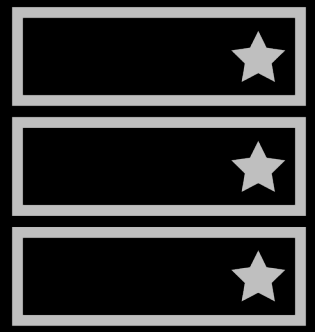
- Short term action



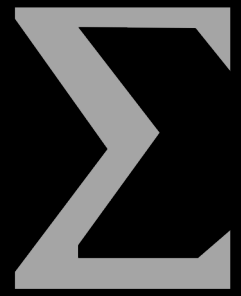
Multi-objective recommender systems



Evolutionary Algorithms



Re-ranking



Weighted Sum

$$\alpha_i \nabla_i$$

Multi-gradient descent

Multi-objective optimization problem

$$\min_{w \in R^D} \boldsymbol{L}(\boldsymbol{w}) = \min_{w \in R^D} \begin{vmatrix} L_1(w) \\ L_2(w) \\ \vdots \\ L_n(w) \end{vmatrix}_{n \times 1}$$

Common descent vector

$$\boldsymbol{\nabla_w L}(\boldsymbol{w}) = \sum_{i=1}^{N} \alpha_i \nabla_w L_i(w)$$

$\alpha_1, \ldots, \alpha_n \geq 0$ and $\sum_{i=1}^{n} \alpha_i = 1$.

Stochastic multi-subgradient descent

- **QCOP**

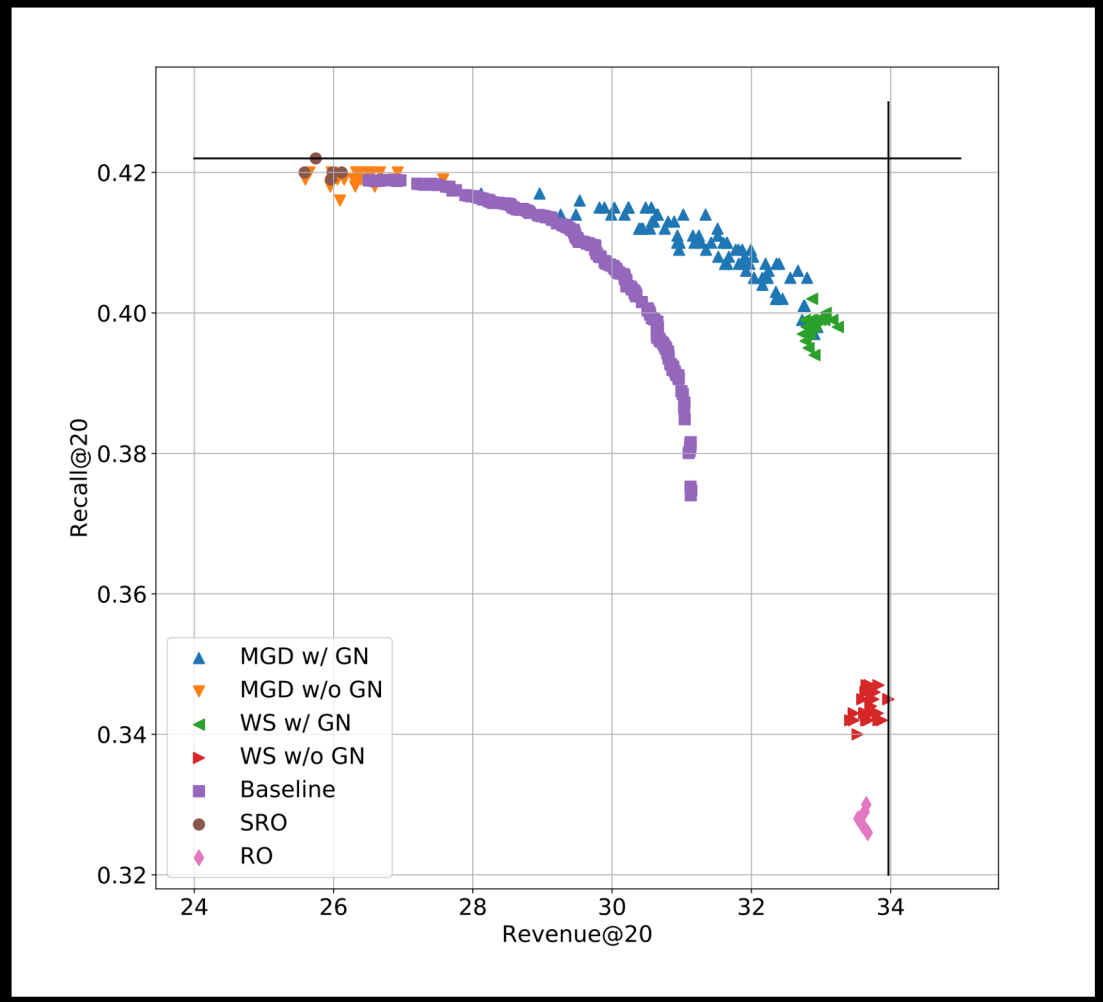
$$\boldsymbol{\nabla_w L(w)} = \min_{\alpha_1,\dots,\alpha_n} \left\{ \left\| \sum_{i=1}^{N} \alpha_i \nabla_w L_i(w) \right\|^2 \;\middle|\; \sum_{i=1}^{N} \alpha_i = 1, \alpha_1, \dots, \alpha_n \geq 0 \right\}$$

- The trick: Gradient normalization

$$\nabla_w \widehat{L_i}(w) = \frac{\nabla_w L_i(w)}{L_i(w_{init})}$$

$$\widehat{L_i}(w) \in [0, 1]$$

Correlated Objectives in Recommenders



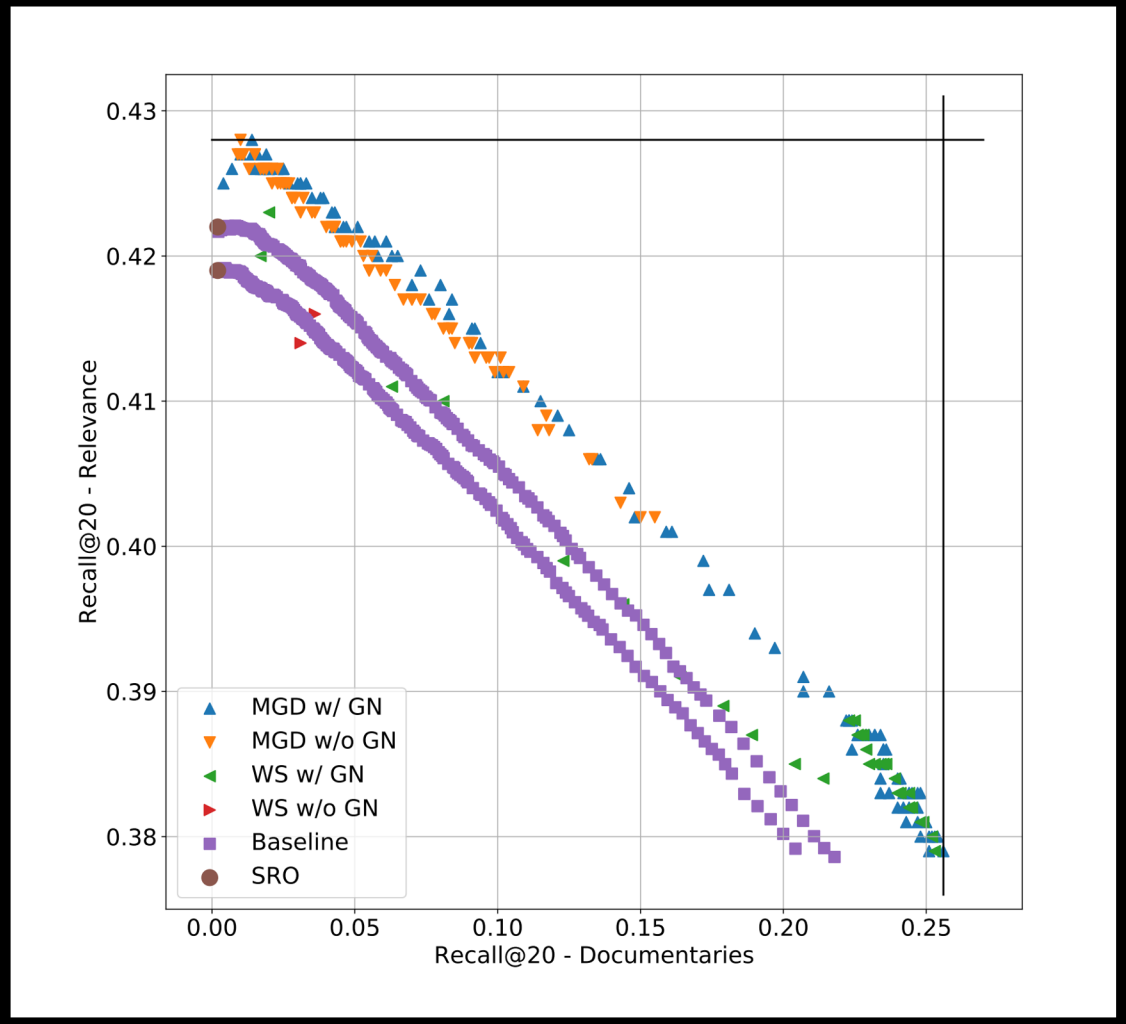
Single Objectives
Revenue
Semantic Relevance

Reranking

Weighted Sum
w/o grad norm
with grad norm

SMSGDA
w/o grad norm
with grad norm

Uncorrelated Objectives in Recommenders



Single Objectives
Revenue
Semantic Relevance

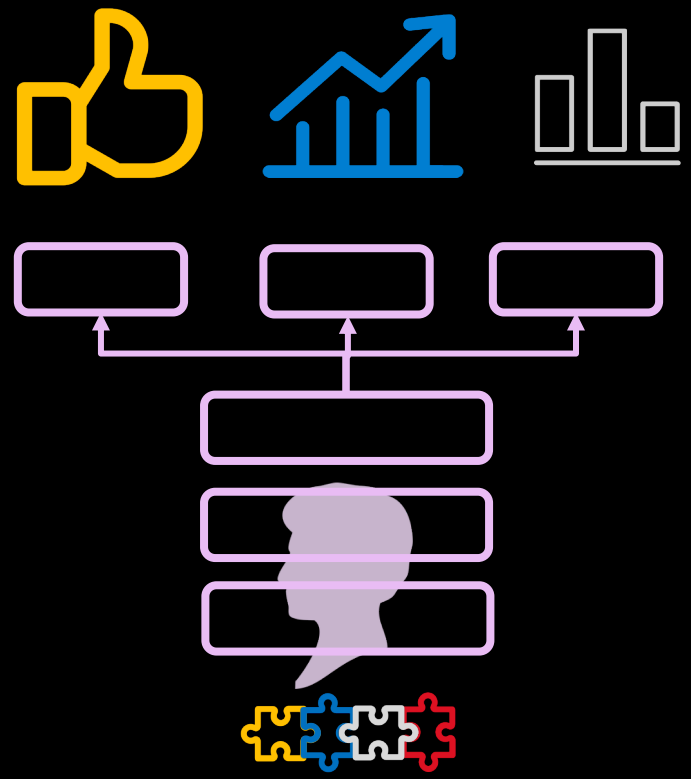
Reranking

Weighted Sum
w/o grad norm
with grad norm

SMSGDA
w/o grad norm
with grad norm

Multi-Gradient Descent for Multi-Objective Recommender Systems. Milojkovik et al. WICRS@AAAI 2020

Knowledge Efficiency



Evaluating the Search Phase of Neural Architecture Search. Yu, Sciuto et al. ICLR 2020

swisscom

@_claudiu_musat

Claudiu Musat

swisscom