Recommendation Independence

Toshihiro Kamishima*, Shotaro Akaho*, Hideki Asoh*, and Jun Sakuma**

www.kamishima.net

**National Institute of Advanced Industrial Science and Technology (AIST), Japan*

***University of Tsukuba; and RIKEN Center for Advanced Intelligence Project, Japan*

1st Conference on Fairness, Accountability, and Transparency (FAT* 2018)

@ New York, USA, Feb. 24, 2018

START

Outline

We overview our series of researches on fairness in recommendation rather then focusing on this paper

Basics of Recommendation

- recommendation tasks, collaborative filtering

Recommendation Independence

- sensitive feature, recommendation independence

Applications of Recommendation Independence

- adherence to laws and regulations, fair treatment of content providers, exclusion of Unwanted Information

Independence-Enhanced Recommendation

- regularization approach, model-based approach

Conclusions & Future work

Basics of Recommendation

Recommender System

[Konstan+ 03]

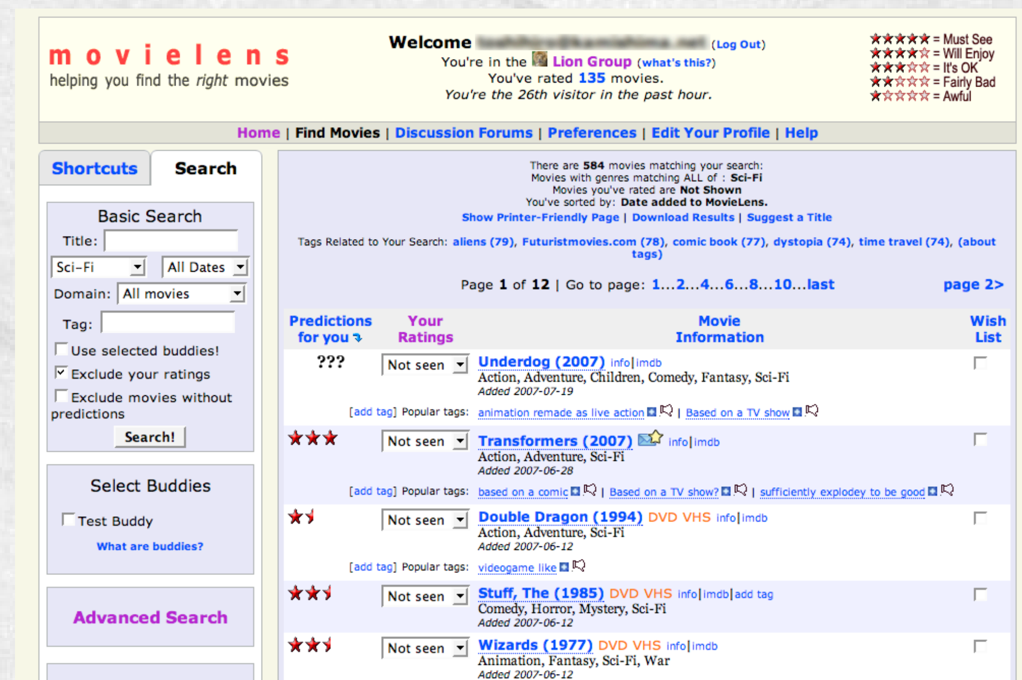
Recommenders: Tools to help identify worthwhile stuff

[Herlocker+ 04, Gunawardana+ 09]

Find Good Items

Ranking items according to users' preference, to help for finding at least one target item

Predicting Ratings



Presenting items with predicted ratings for a user, to help for exploring items

＊ Screen-shots are acquired from Amazon.co.jp and Movielens.org on 2007-07-26

Collaborative Filtering

Collaborative filtering is a major approach for predicting users' preference in a word-of-mouth manner

recommending items liked by those who having similar preferences

✻ There are other approaches: content-based filtering or knowledge-based filtering

Recommendation Independence

Sensitive Feature

As in a case of standard recommendation, we use random variables
X: a user, Y: an item, and R: a recommendation outcome

We adopt a variable required for recommendation independence

S : **sensitive feature**

- This represents information that should be ignored in a recommendation process
- Its values are determined depending on a user and/or an item

Ex. Sensitive feature = movie's popularity / user's gender

A sensitive feature is restricted to a binary type

Recommendation Independence

[Kamishima+ 12, Kamishima+ 13, Kamishima+ 16, Kamishima+18]

Recommendation Independence

statistical independence

between a recommendation outcome, R, and a sensitive feature, S

$$\Pr[R \mid S] = \Pr[R] \equiv R \perp\!\!\!\perp S$$

- No information about a sensitive feature influences the outcome
- The status of the sensitive feature is explicitly excluded from the inference of the recommendation outcome

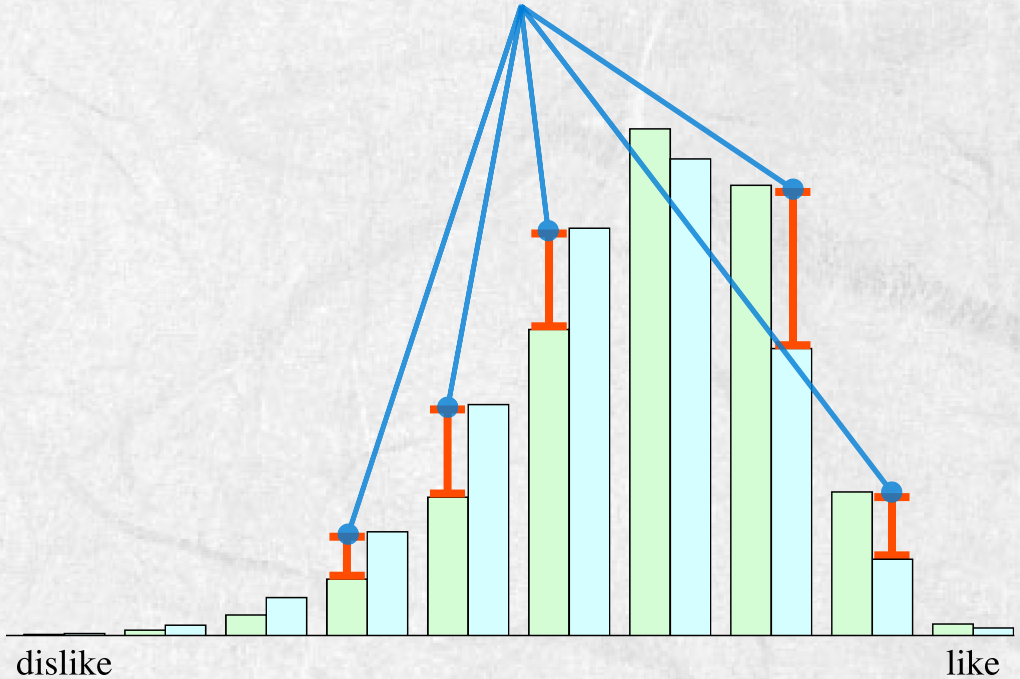
Independence-Enhanced Recommendation

Preferred items are predicted
so as to satisfy a constraint of recommendation independence

Effect of Independence Enhancement

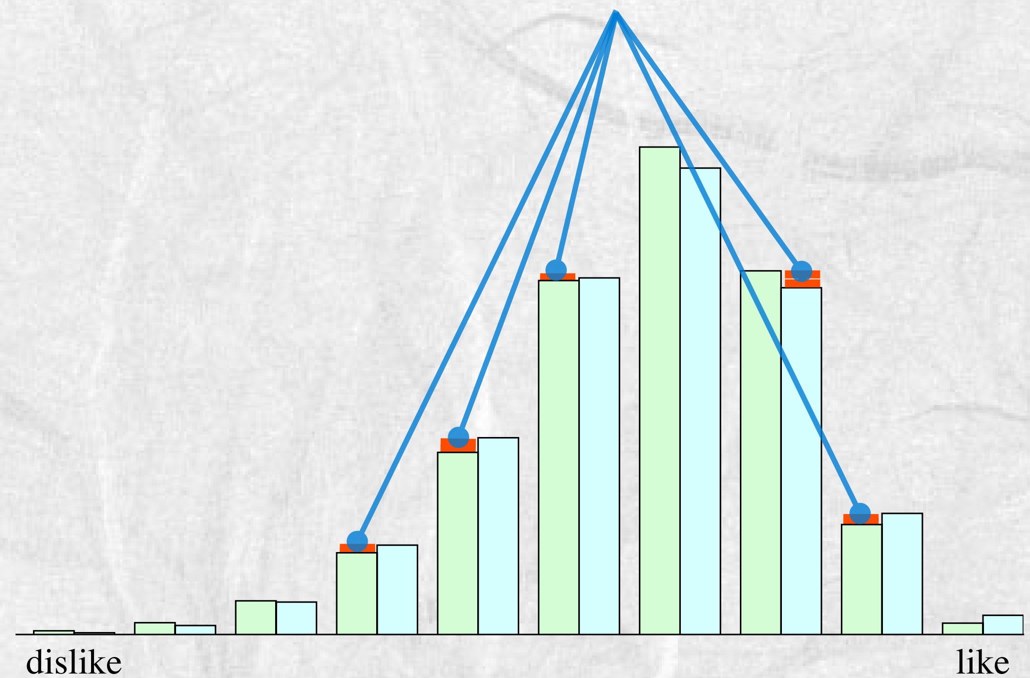
Standard

two distributions are largely diverged



Independence-enhanced

two distributions become closer



a sensitive feature = whether a movie is newer or older

✻ each bin of histograms of predicted scores for older and newer movies

The bias that older movies were rated higher could be successfully canceled by enhancing independence

Applications of Recommendation Independence

Adherence to Laws and Regulations

[Sweeney 13]

A recommendation service must be managed while adhering to laws and regulations

suspicious placement in keyword-matching advertisements

Advertisements indicating arrest records were more frequently displayed for names that are more popular among individuals of African descent than those of European descent

⬇

Socially discriminative treatments must be avoided

sensitive feature = users' demographic information

⬇

Legally or socially sensitive information can be excluded from the inference process of recommendation

Fair Treatment of Content Providers

System managers should fairly treat their content providers

Fair treatment in search engines

[Bloomberg]

The US FTC has investigated Google to determine whether the search engine ranks its own services higher than those of competitors

Fair treatment in recommendation

A hotel booking site should not abuse their position to recommend hotels of its group company

sensitive feature = a content provider of a candidate item

⬇

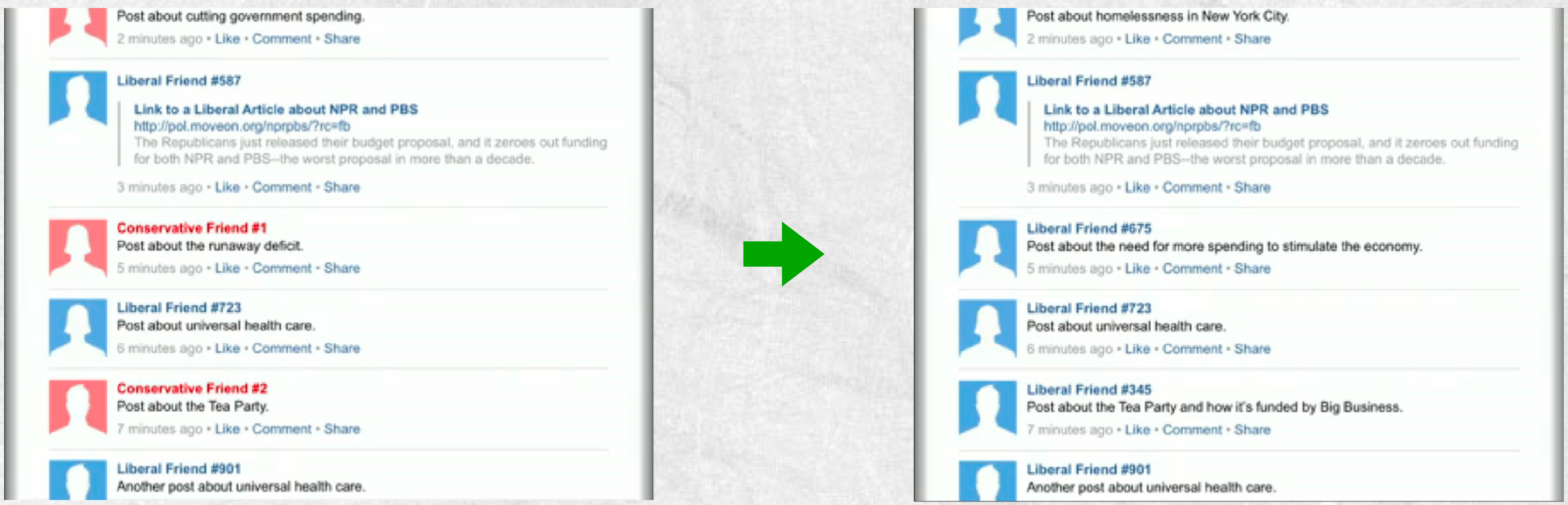
Information about who provides a candidate item can be ignored, and providers are treated fairly

Exclusion of Unwanted Information

[TED Talk by Eli Pariser, http://www.filterbubble.com/]

Information unwanted by a user is excluded from recommendation

Filter Bubble: To fit for Pariser's preference, conservative people are eliminated from his friend recommendation list in Facebook



sensitive feature = a political conviction of a friend candidate

⬇

Information about whether a candidate is conservative or progressive can be ignored in a recommendation process

Independence-Enhanced Recommendation

Independence-Enhanced Recommendation

[Kamishima+ 12, Kamishima+ 13, Kamishima+ 16, Kamishima+18]

Predicting Ratings: a task to predict a rating value that a user would provide to an item

Random variables: user X, item Y, rating R, sensitive feature S

Standard Recommendation	**Independence-Enhanced Recommendation**
Dataset $\mathcal{D} = \{(x_i, y_i, r_i)\}$	Dataset $\mathcal{D} = \{(x_i, y_i, r_i, s_i)\}$
⬇	⬇
Prediction Function $\hat{r}(x, y)$	Prediction Function $\hat{r}(x, y, s)$

Regularization Approach

[Kamishima+ 12, Kamishima+ 13, Kamishima+18]

Regularization Approach: Adopting a regularizer imposing a constraint of independence while training a recommendation model

Objective Function

independence parameter: control the balance between independence and accuracy

L2 regularizer

empirical loss

regularization parameter

$$\underbrace{\sum_D \mathrm{loss}\left(r_i, \hat{r}(x_i, y_i, s_i)\right)}_{\text{empirical loss}} - \eta\, \boxed{\mathrm{indep}(R, S)} + \frac{1}{2} \lambda \|\mathbf{\Theta}\|^2$$

independence term: a regularizer to constrain independence

The larger value indicates that recommendation outcomes and sensitive values are more independent

Independence Terms

Mutual Information with Histogram Models [Kamishima+ 12]

- computationally inefficient

Mean Matching [Kamishima+ 13]

$$-\left(\operatorname{mean}\left(\mathcal{D}^{(0)}\right) - \operatorname{mean}\left(\mathcal{D}^{(1)}\right)\right)^2$$

- matching means of predicted ratings for distinct sensitive groups
- improved computational efficiency, but considering only means

Mutual Information with Normal Distributions [Kamishima+ 18]

$$-\left(\mathrm{H}(R) - \sum_s \Pr[s]\,\mathrm{H}(R|s)\right)$$

Distribution Matching with Bhattacharyya Distance [Kamishima+ 18]

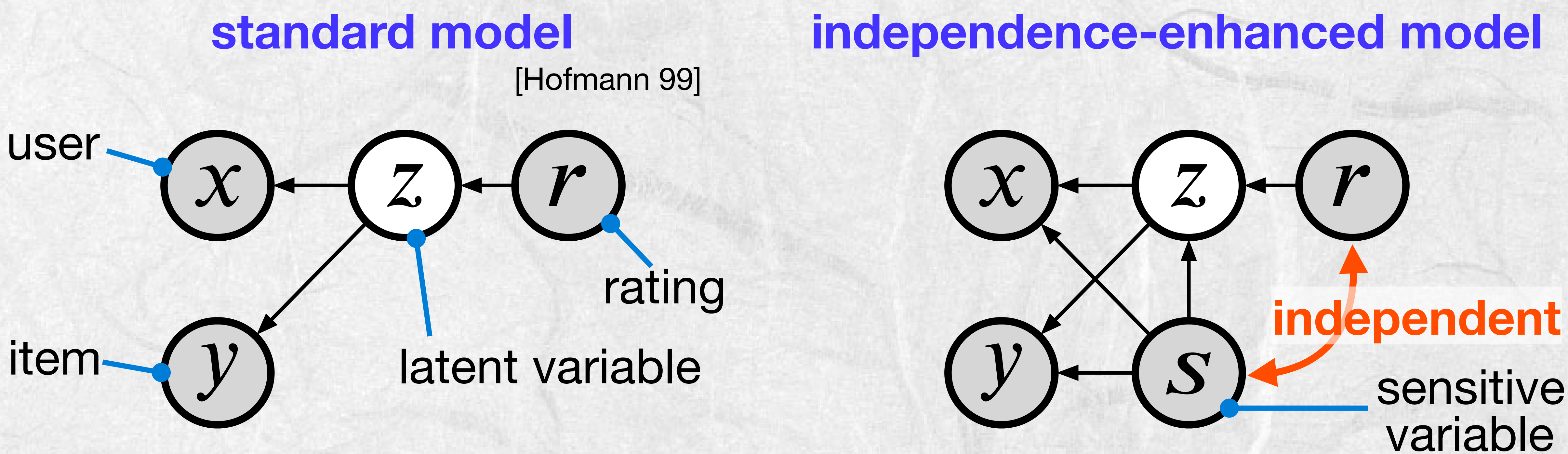
$$-\left(-\ln \int \sqrt{\Pr[r|S{=}0]\Pr[r|S{=}1]}\,dr\right)$$

- These two terms can take both means and variances into account, and are computationally efficient

Model-based Approach

[Kamishima+ 16]

Model-based Approach: a sensitive variable is added to a recommendation model so that it satisfies an independence constraint



A model-based approach is inferior to a regularization approach

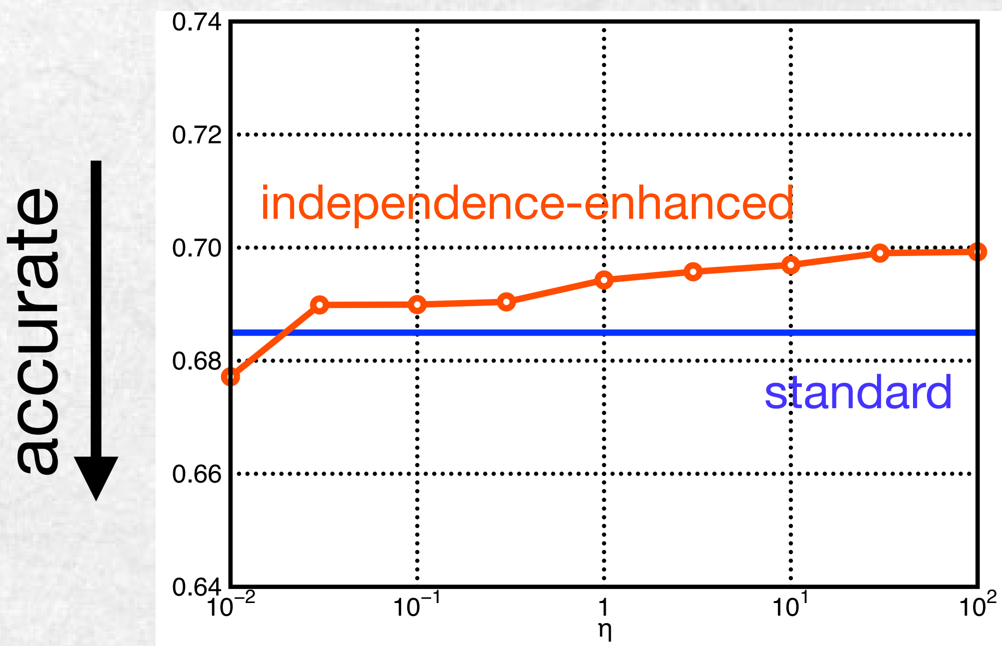
Generative process of ratings are assumed to be probabilistic in this model, but it is actually deterministic

[Kamishima+ 18b]

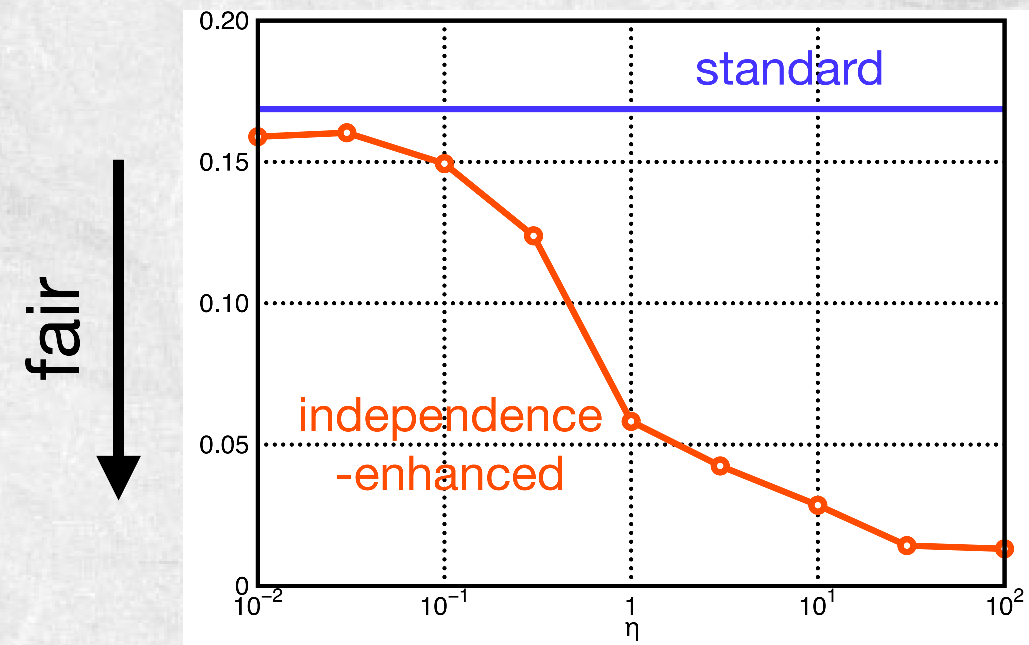
Experiments: Accuracy vs Fairness

- We apply a regularization method with mutual information with normal distributions to Movielens 1M with the Year sensitive feature
- The changes of accuracy and independence measures according as the enhancement of Independence

Accuracy
mean absolute error



Independence
Kolmogorov-Smirnov test statistics



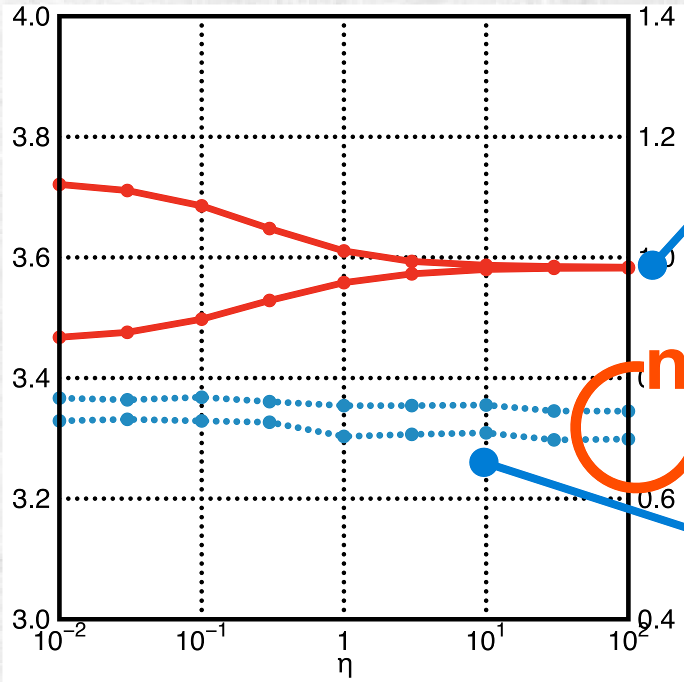
Recommendation independence could be successfully enhanced by slightly sacrificing prediction accuracy

✻ Details of experimental conditions are shown in Sec. 4.2.1 and Fig. 2

Experiments: Means & Variances

- Comparison between our previous method, **mean match**, with our latest method, **mutual information with normal distributions**
- The changes of means and variances of predicted ratings according as the enhancement of independence

mean matching

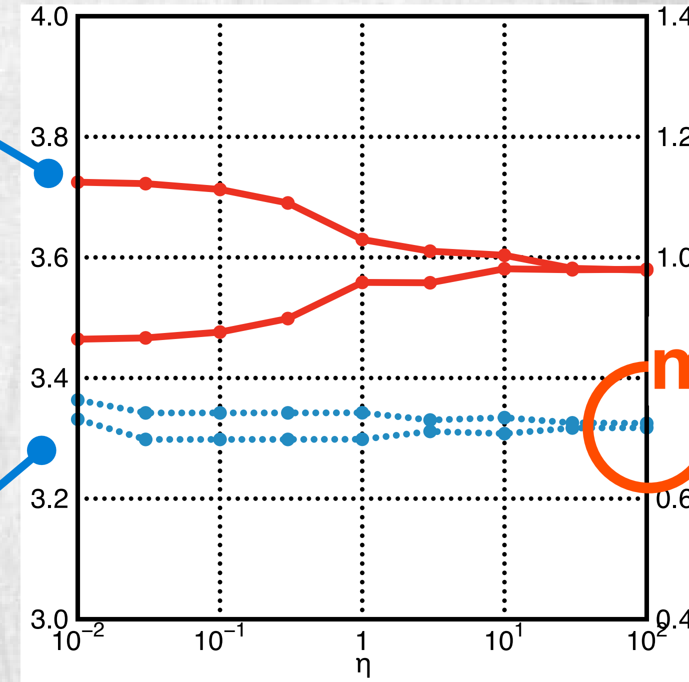


differences of means

not match

differences of standard deviations

mutual information with Gaussian models



match

Our previous method cannot control the variances of predicted ratings, but our new method can

✻ Details of experimental conditions are shown in Sec. 4.2.1 and Fig. 3

Conclusions

Contributions

- We proposed a notion of recommendation independence
- We have been developed methods for independence-enhanced recommendation
- Enhancement of recommendation independence have been empirically examined
- Our new independence terms could take variances of outcomes into account

Future work

- Recommendation independence for a find-good-items task
- Sensitive features other than a binary type, such as a continuous type
- Other types of independence, such as equalized odds [Hardt+ 17, Yao+ 18]
- In cases that are not point-estimation, such as Bayesian inference
- Introduce conditional fairness or confounding variables

Additional Information

Program Codes (plan to open in March)

http://www.kamishima.net/iers/

Our Survey Slide of Fairness-Aware Data Mining

http://www.kamishima.net/archive/fadm.pdf

Acknowledgment

- We gratefully acknowledge the valuable comments and suggestions of Dr. Paul Resnick
- We would like to thank the Grouplens research lab and Dr. Mohsen Jamali for providing datasets
- This work is supported by MEXT/JSPS KAKENHI Grant Number JP24500194, JP15K00327, and JP16H02864

Bibliography I

Ò. Celma and P. Cano.
From hits to niches?: or how popular artists can bias music recommendation and discovery.
In *Proc. of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, 2008.

S. Forden.
Google said to face ultimatum from FTC in antitrust talks.
Bloomberg, Nov. 13 2012.
⟨http://bloom.bg/PPNEaS⟩.

A. Gunawardana and G. Shani.
A survey of accuracy evaluation metrics of recommendation tasks.
Journal of Machine Learning Research, 10:2935–2962, 2009.

M. Hardt, E. Price, and N. Srebro.
Equality of opportunity in supervised learning.
In *Advances in Neural Information Processing Systems 29*, 2016.

J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl.
Evaluating collaborative filtering recommender systems.
ACM Trans. on Information Systems, 22(1):5–53, 2004.

Bibliography II

T. Hofmann and J. Puzicha.
Latent class models for collaborative filtering.
In *Proc. of the 16th Int'l Joint Conf. on Artificial Intelligence*, pages 688–693, 1999.

M. Jamali and M. Ester.
A matrix factorization technique with trust propagation for recommendation in social networks.
In *Proc. of the 4th ACM Conf. on Recommender Systems*, pages 135–142, 2010.

T. Kamishima and S. Akaho.
Considerations on recommendation independence for a find-good-items task.
In *Workshop on Responsible Recommendation*, 2017.

T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma.
Enhancement of the neutrality in recommendation.
In *The 2nd Workshop on Human Decision Making in Recommender Systems*, 2012.

T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma.
Efficiency improvement of neutrality-enhanced recommendation.
In *The 3rd Workshop on Human Decision Making in Recommender Systems*, 2013.

T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma.
Model-based and actual independence for fairness-aware classification.
Data Mining and Knowledge Discovery, 32:258–286, 2018.

Bibliography III

- T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma.
  Recommendation independence.
  In *The 1st Conf. on Fairness, Accountability and Transparency*, volume 81 of *PMLR*, pages 187–201, 2018.
- T. Kamishima, S. Akaho, H. Asoh, and I. Sato.
  Model-based approaches for independence-enhanced recommendation.
  In *Proc. of the IEEE 16th Int'l Conf. on Data Mining Workshops*, pages 860–867, 2016.
- J. A. Konstan and J. Riedl.
  Recommender systems: Collaborating in commerce and communities.
  In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, Tutorial*, 2003.
- Y. Koren.
  Factorization meets the neighborhood: A multifaceted collaborative filtering model.
  In *Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pages 426–434, 2008.
- E. Pariser.
  The Filter Bubble: What The Internet Is Hiding From You.
  Viking, 2011.
- R. Salakhutdinov and A. Mnih.
  Probabilistic matrix factorization.
  In *Advances in Neural Information Processing Systems 20*, pages 1257–1264, 2008.

Bibliography IV

L. Sweeney.
Discrimination in online ad delivery.
Communications of the ACM, 56(5):44–54, 2013.

S. Yao and B. Huang.
Beyond parity: Fairness objectives for collaborative filtering.
In *Advances in Neural Information Processing Systems 30*, 2017.

C. N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen.
Improving recommendation lists through topic diversification.
In *Proc. of the 14th Int'l Conf. on World Wide Web*, pages 22–32, 2005.

Extra Slides

Probabilistic Matrix Factorization

[Salakhutdinov 08, Koren 08]

Probabilistic Matrix Factorization Model

predict a preference rating of an item y rated by a user x

well-performed and widely used

Prediction Function

$$\hat{r}(x,y) = \underbrace{\mu}_{\text{global bias}} + \underbrace{b_x}_{\text{user-dependent bias}} + \underbrace{c_y}_{\text{item-dependent bias}} + \underbrace{\mathbf{p}_x \mathbf{q}_y^\top}_{\text{cross effect of users and items}}$$

Objective Function

$$\underbrace{\sum_D (r_i - \hat{r}(x_i, y_i))^2}_{\text{squared loss function}} + \underbrace{\lambda}_{\text{regularization parameter}} \|\underbrace{\boldsymbol{\Theta}}_{L_2 \text{ regularizer}}\|^2$$

For a given training dataset, model parameters are learned by minimizing the squared loss function with an L_2 regularizer.

Regularization Approach

[Kamishima+ 12, Kamishima+ 13, Kamishima+18]

Prediction Function

a prediction function is selected according to a sensitive value

$$\hat{r}(x, y, s) = \mu^{(s)} + b_x^{(s)} + c_y^{(s)} + \mathbf{p}_x^{(s)} \mathbf{q}_y^{(s)\top}$$

sensitive feature

Objective Function

independence parameter: control the balance between the independence and accuracy

$$\sum\nolimits_D \left(r_i - \hat{r}(x_i, y_i, s_i)\right)^2 - \eta \, \boxed{\mathrm{indep}(R, S)} + \lambda \|\Theta\|^2$$

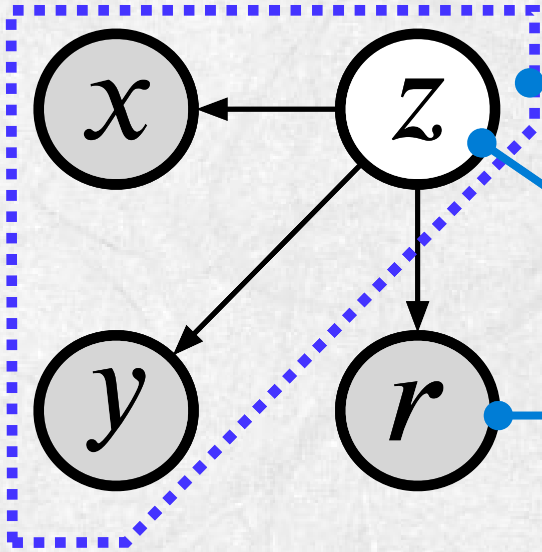
independence term: a regularizer to constrain independence

- The larger value indicates that ratings and sensitive values are more independent
- Matching means of predicted ratings for two sensitive values

Latent Class Model

[Hofmann 99]

Latent Class Model: A probabilistic model for collaborative filtering



A basic topic model, pLSA

latent topic variable

extended so as to be able to deal with ratings r given by users x to items y

Model parameters can be learned by an EM algorithm

Prediction:

$$\begin{aligned}\hat{r}(x,y) &= \mathrm{E}_{\Pr[r|x,y]}[\mathrm{level}(r)] \\ &= \textstyle\sum_r \Pr[r|x,y]\,\mathrm{level}(r)\end{aligned}$$

the r-th rating value

A rating value can be predicted by the expectation of ratings

Independence-Enhanced LCM

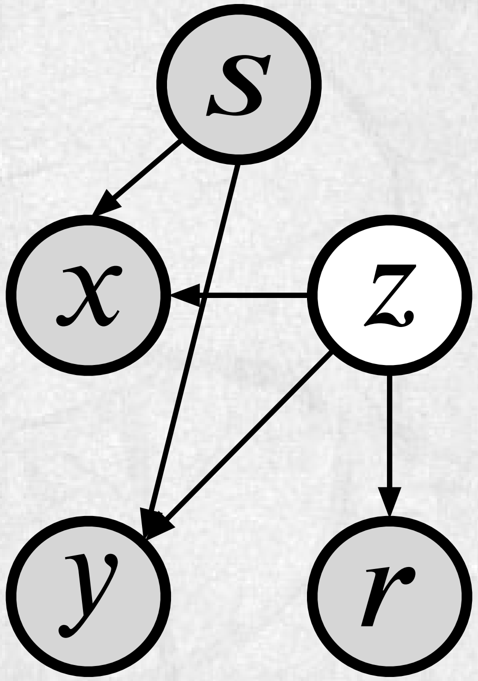
[Kamishima+ 16]

Independence-Enhancement by a Model-based Approach

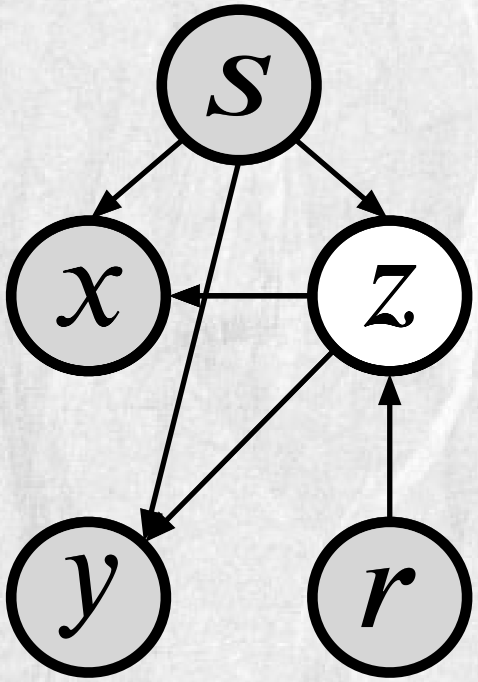
A sensitive variable is embedded into the original LCM

- A rating and a sensitive variable are mutually independent
- A user, an item, and a rating are conditionally independent given Z

Type 1 model



Type 2 model



A type 2 model can more strictly enhance recommendation independence, because in addition to X and Y, Z depends on a sensitive variable

Ugly Duckling Theorem

[Watanabe 69]

Ugly Duckling Theorem

In classification, one must emphasize some features of objects and must ignore the other features

It is impossible to make recommendation that is independent from any sensitive features

Independence-enhanced Recommendation

a sensitive feature must be specified by a user and other features are ignored

In a case of a Facebook example, A recommender system enhances independent from a political conviction, but it is allowed to make biased recommendations in terms of other features, for example, the birthplace or age of friend candidates

For a Find-Good-Items Task

[Kamishima+ 17]

Find Good Items: a task to find some items preferred by a user

⬇

making a preference score independent, instead of a predicted rating

Preference Score: How strongly a user prefers an item

sigmoid function

$$\hat{r}(x, y, s) = \mathrm{sig}\left(\mu^{(s)} + b_x^{(s)} + c_y^{(s)} + \mathbf{p}_x^{(s)} {\mathbf{q}_y^{(s)}}^\top\right)$$

$$\textstyle\sum_D \mathrm{CE}\left(r_i - \hat{r}(x_i, y_i, s_i)\right) - \eta \,\mathrm{indep}(R, S) + \lambda \|\Theta\|^2$$

cross-entropy loss

enhancing the independence between a preference score and a sensitive feature

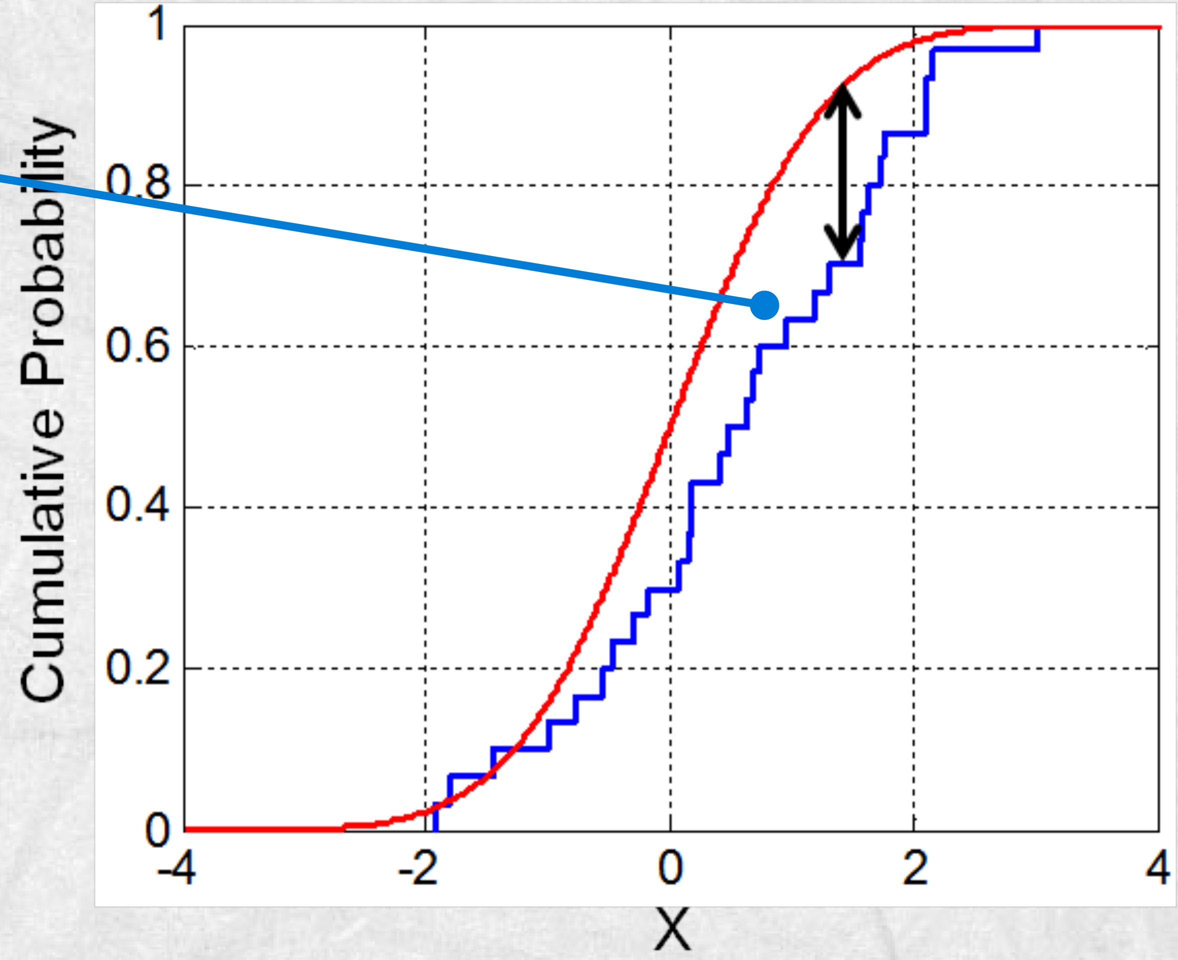
Kolmogorov-Smirnov Statistic

The statistic of the two-sample Kolmogorov-Smirnov test
a nonparametric test for the equality of two distribution

⬇

Evaluating the degree of independence
by measuring the equality between $\Pr[R \mid S{=}0]$ and $\Pr[R \mid S{=}1]$

Kolmogorov-Smirnov statistic
the area between two empirical cumulative distributions

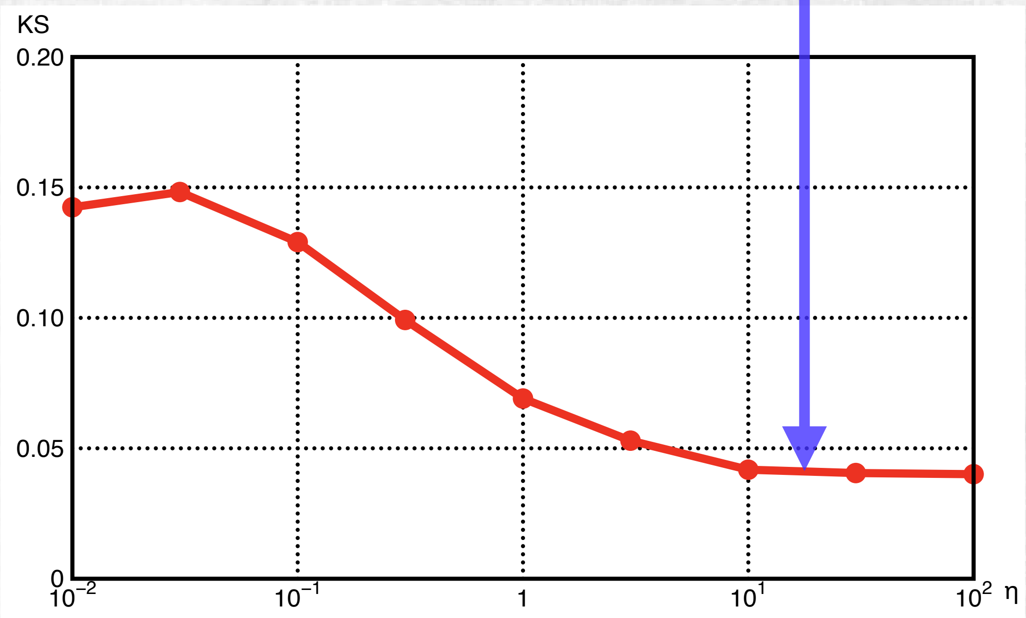


Wikipedia

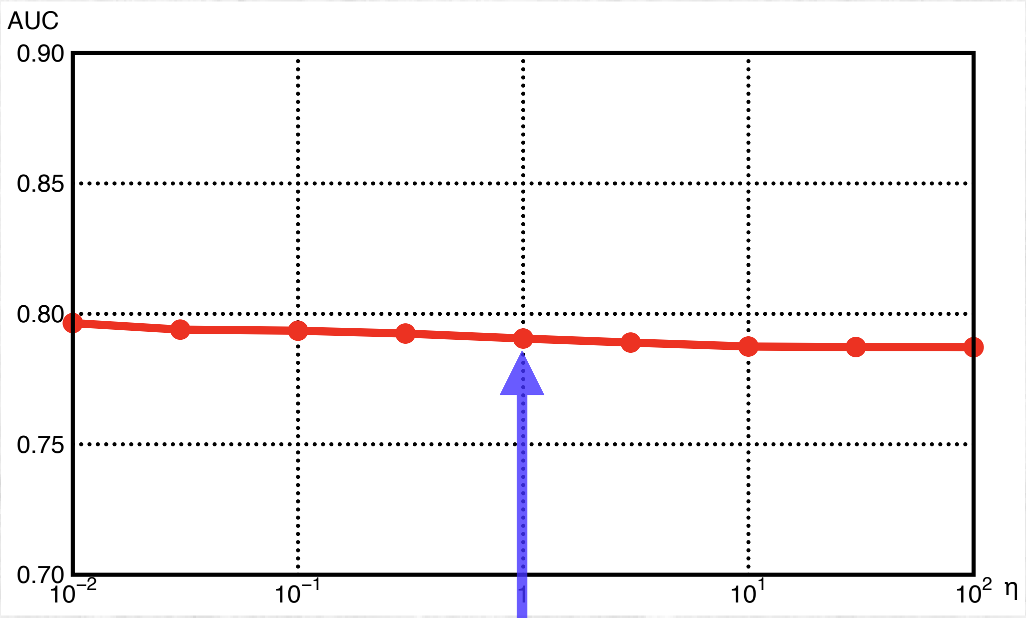
Preference Score vs Sensitive Feature

[Kamishima+ 17]

Observation 1: **A preference score could be successfully made independent from a sensitive feature**



Kolmogolov-Smirnov statistic between $\Pr[R \mid S{=}0]$ and $\Pr[R \mid S{=}1]$



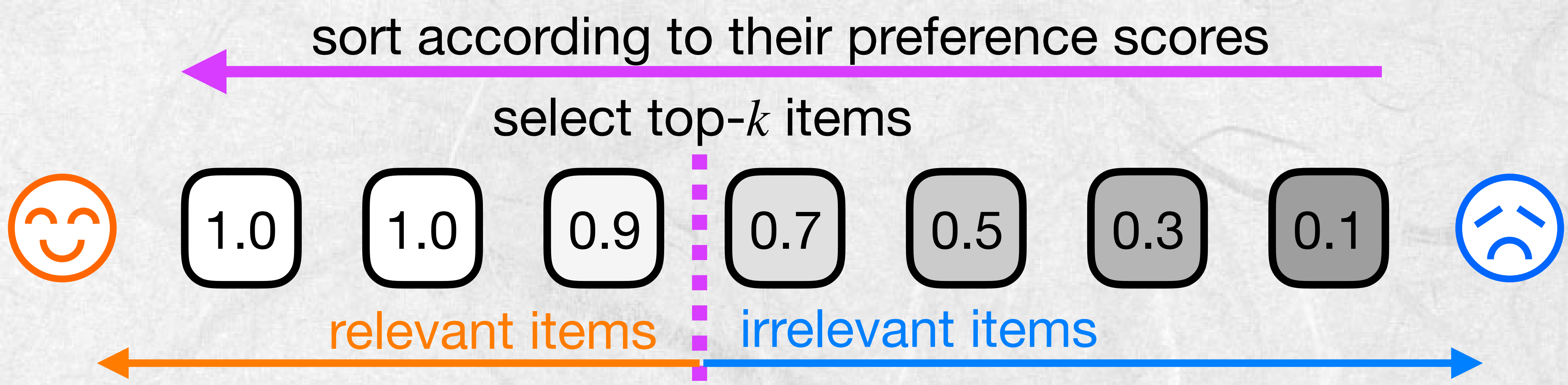
ranking accuracy (AUC)

Observation 2: **A ranking accuracy (AUC) did not worsen so much** by enhancement of the recommendation independence
This is was contrasted with the increase of a prediction error (MAE) in a predicting ratings task

Relevance and Sensitive Feature

[Kamishima+ 17]

Recommending top-k items whose preference scores are the largest



check the independence from a relevance, not from a preference score

Observation 3: **The relevance of items was not independent from a sensitive feature** for some values of k, in particular, small k

A need for a new method that fits for a ranked item list

Popularity Bias

[Celma 08]

Popularity Bias
the tendency for popular items to be recommended more frequently

[Jamali+ 10]

Flixster data
The degree popularity of an item is measured
by the number of users who rated the item

short-head (top 1%)
share in ratings: 47.2%
mean rating: **3.71**

↔

long-tail (bottom 99%)
share in ratings: 52.8%
mean rating: **3.53**

Short-head items are frequently and highly rated

sensitive feature = popularity of items

↓

Popularity bias can be corrected

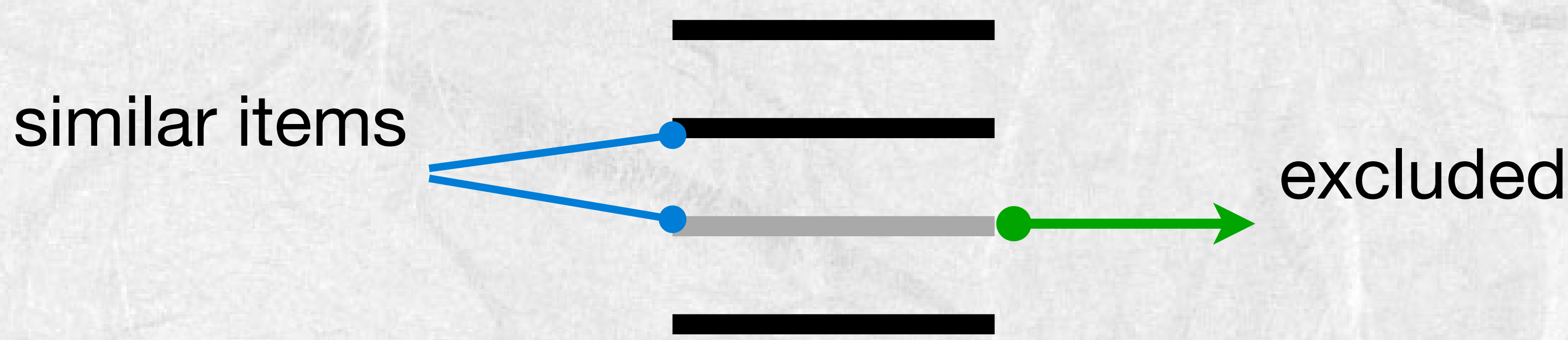
Recommendation Diversity

[Ziegler+ 05]

Recommendation Diversity

Similar items are not recommended in a single list, to a single user, to all users, or in a temporally successive lists

recommendation list



Diversity

Items that are similar in a specified metric are excluded from recommendation results

The mutual relations among results

Independence

Information about a sensitive feature is excluded from recommendation results

The relations between results and sensitive values

Diversity vs Independence

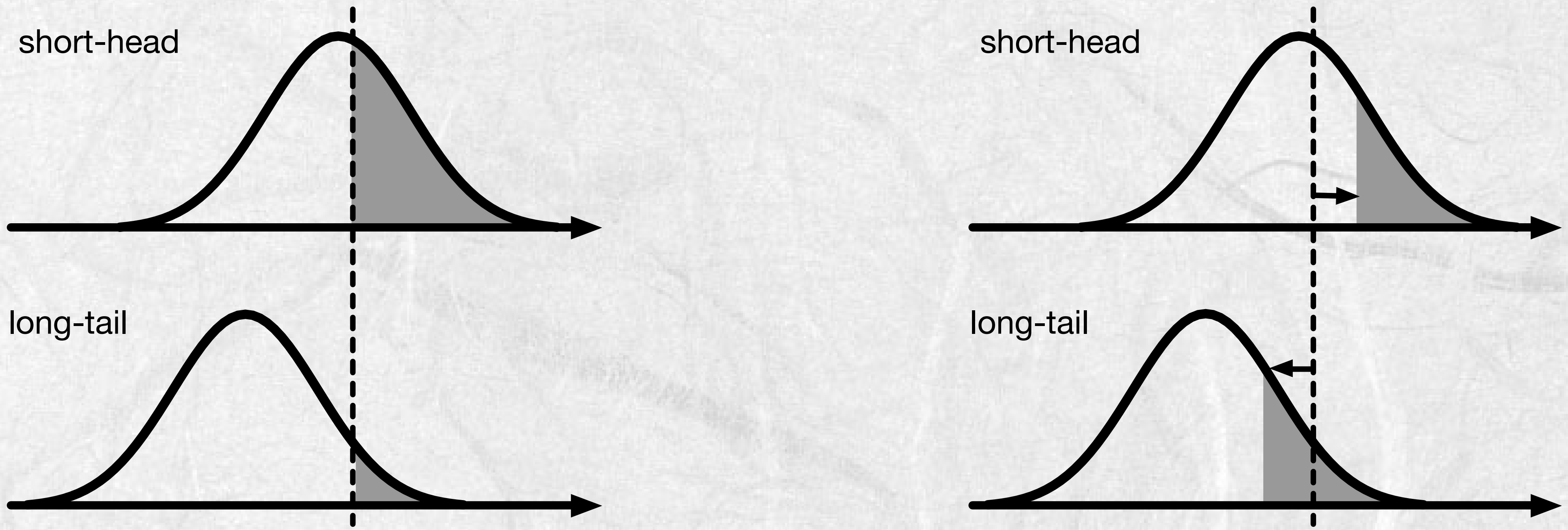
Diversity	**Independence**
Depending on the definition of similarity measures	Depending on the specification of sensitive feature
↓	↓
Similarity	**Sensitive Feature**
A function of **two items**	A function of **a user-item pair**

Because a sensitive feature depends on a user, neutrality can be applicable for coping with users' factor, such as, users' gender or age, which cannot be straightforwardly dealt by using diversity

Diversity vs Independence

standard recommendation

diversified recommendation



Because a set of recommendations are diversified by abandoning short-head items, **predicted ratings are still biased**

⇕

Prediction ratings themselves are unbiased by enhancing recommendation independence

Privacy-Preserving Data Mining

recommendation results, R, and sensitive features, S,
are statistically independent

⬇

mutual information between a recommendation result, R,
and a sensitive feature, S, is zero

$$\mathrm{I}(R; S) = 0$$

⬇

In a context of privacy-preservation

Even if the information about R is disclosed,
the information about S will not exposed

In particular, a notion of the *t*-closeness has strong connection