NDBI021, Lecture 2

User preferences, 2/1 ZK+Z,

Wed 12:20 – 13:50 S8

Wed 14:00 – 15:30 SW2 (odd weeks)

https://www.ksi.mff.cuni.cz/~peska/vyuka/ndbi021/2022/

matfyz

siret research group

https://ksi.mff.cuni.cz

How to express user preferences

Feedback variants for users

What tools are there to express preferences

- „Expressing by doing“ (implicit feedback)
- Rating/(dis)approving (explicit feedback)
- Filtering, searching
- Explicit comparison (A is better than B)
- Critiquing
- Writing a review

- Did I forgot anything?

What tools are there to express preferences

- Explicit feedback
- Implicit feednack
- Searching / filtering

Why doing user preference research feels like being a parent?

How to collect user preferences?

- **Rating, filtering, comparison, reviews… via designated GUI**
  - How to store e.g. searching / filtering may be a bit tricky…
- **Implicit feedback**
  - **Server-side (limited expressibility)**
  - **Client-side (triggered JS events)**
  - **Beyond (eye tracking, other biometrics)**
    - Limited applicability (lab studies)
    - Can provide leads on interpretation of the previous two
- **Questionaires, role playing**
  - Lab studies only (in most cases)
  - Can provide leads on interpretation of other collection methods

Explicit feedback

Let the users to tell you

Explicit feedback

- Information given consciously by the user to express his/her preference
  - Via dedicated GUI
- Rating (likert scale) of objects
  - N-ary preference (5 / 10 degrees of preference most common, sometimes pref. slider)
  - Binary preference (likes – dislikes)
  - Unary preference (likes only)
- Simple enough? Nothing to research here?
  - Well... ☺

Explicit feedback

- How rating scale influence user's rating behavior?
  https://www.tandfonline.com/doi/full/10.1080/0144929X.2017.1322145
  - Granularity of the rating scale
  - Presence/absence of neutral point
  - Labeling
- How much are the ratings inconsistent for test-retest?
  - https://xamat.github.io/pubs/xamatriain_umap09.pdf
  - 10.1177/0013164404268674
- **Are different rating scales affecting RS performance?**
  - ?? :-/ (not much research... Bachelor/diploma thesis opportunity?☺
  - http://ceur-ws.org/Vol-997/umap2013_lbr_7.pdf
    - It seems that 3-point likert scale has smaller MAE than 5-point scale
    - But what about scaling effect?

4

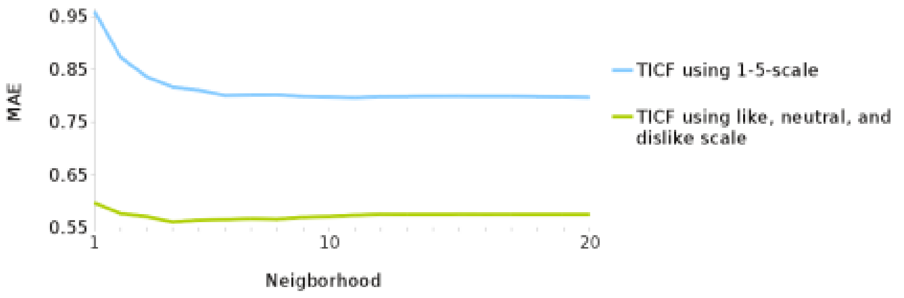


Fig. 1. MAE when using 1-5-scale ratings and a like, neutral, and dislike scale.

Explicit feedback: how scale influence behavior?

- How rating scale influence user rating behavior? https://www.tandfonline.com/doi/full/10.1080/0144929X.2017.1322145
  - Granularity of the rating scale
    - 10.1007/s11205-007-9171-x: evaluate happiness on 4, 5, 7 and 11 points likert scale; then re-scale to 11 points: 11-point scale has higher happiness than 4 and 7 (higher scale higher ratings?)
    - Other authors did not found such re-scaling issues
    - 10.1016/S0001-6918(99)00050-5: 2,3,4 point least reliable and least discriminating, wider options preferred (7-10), but 2-4 points quicker to use
      - **Less granularity imply higher willingness to use?**
      - **Binary/unary schemes less intrusive?**
    - 10.1177/0013164404268674: test-retest scenario; more pints (at least 3) imply higher reliability

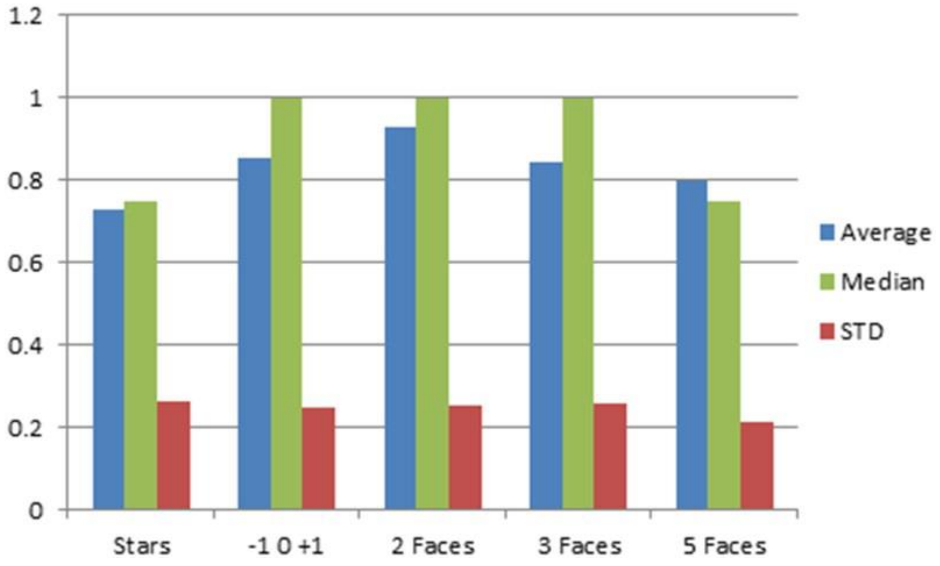
Explicit feedback: how scale influence behavior?

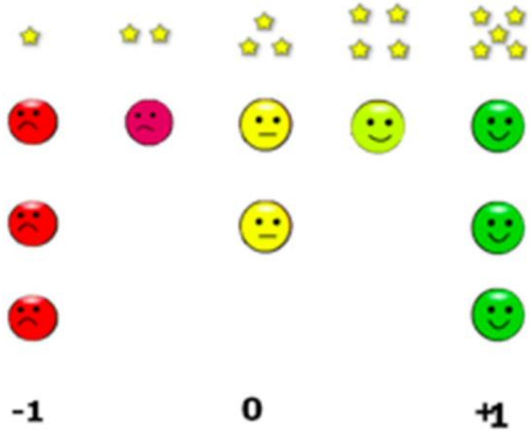
- How rating scale influence user rating behavior?
  https://www.tandfonline.com/doi/full/10.1080/0144929X.2017.1322145
  - Granularity of the rating scale: how different rating scales correlate on real-world services

		Cricketer	Filmeter	FilmCrave	IMDb	MovieLens
Criticker	Pearson Correlation	1	0.791[a]	0.943[a]	0.944[a]	0.946[a]
Filmeter	Pearson Correlation	0.791[a]	1	0.783[a]	0.767[a]	0.740[a]
FilmCrave	Pearson Correlation	0.943[a]	0.783[a]	1	0.934[a]	0.915[a]
IMDb	Pearson Correlation	0.944[a]	0.767[a]	0.934[a]	1	0.933[a]
MovieLens	Pearson Correlation	0.946[a]	0.740[a]	0.915[a]	0.933[a]	1

Explicit feedback: how scale influence behavior?

- How rating scale influence user rating behavior?
https://www.tandfonline.com/doi/full/10.1080/0144929X.2017.1322145
  - Granularity of the rating scale: same site with different rating scales implemented: how the results differ?
  - Note the average





Explicit feedback: how scale influence behavior?

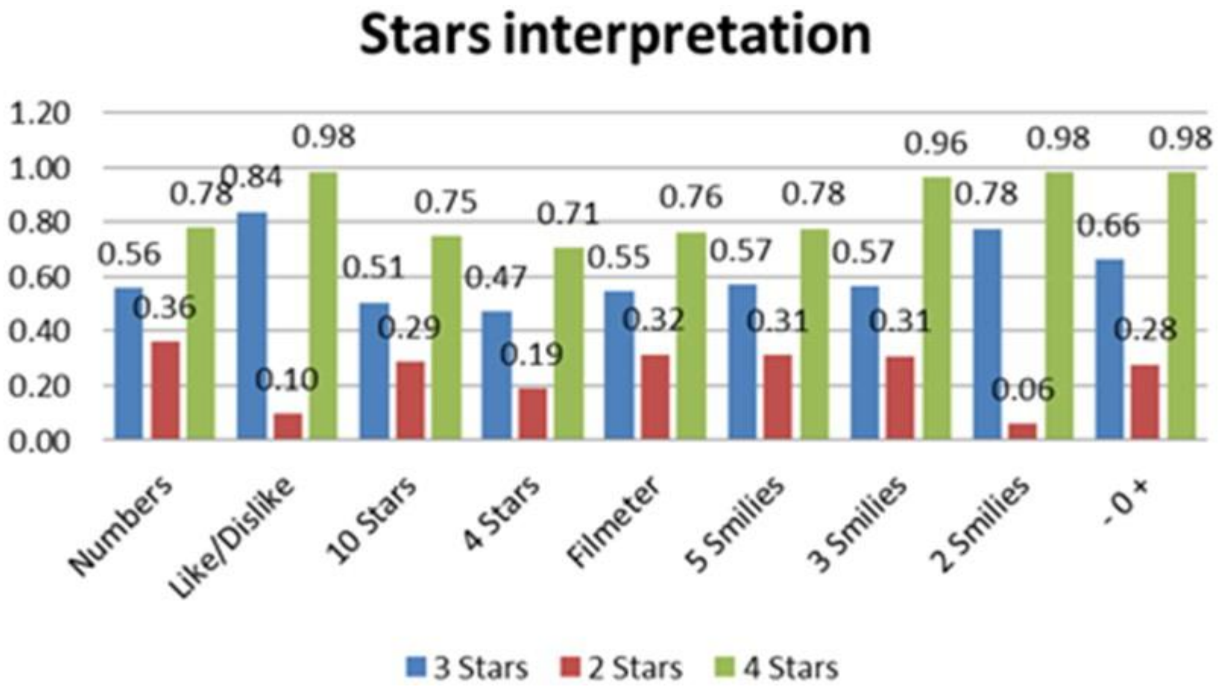
- How rating scale influence user rating behavior? https://www.tandfonline.com/doi/full/10.1080/0144929X.2017.1322145
  - Neutral point
    - https://www.rangevoting.org/MB_V2_N3_Garland.pdf :
      - some respondents may choose the midpoint in order to provide a less negative answer, because of a social desirability bias
      - rating scales with no midpoint force the real indifferent to make a choice, causing a distortion towards higher or lower answers
    - 10.1016/j.ijresmar.2010.02.004:
      - with neutral points in the rating scale, we will have less extreme responses and higher ratings

Explicit feedback: how scale influence behavior?

- How rating scale influence user rating behavior? https://www.tandfonline.com/doi/full/10.1080/0144929X.2017.1322145
  - Labels
    - http://www.websm.org/uploadi/editor/1368430817Hereshey_1993_The_Biasing_Effects_of_scale_checking.pdf : Ordering of labels could matter
      - collected students' attitudes towards their college
      - 'strongly agree', 'agree', 'undecided', 'disagree' and 'strongly disagree' and opposite order,
      - first scale resulted in a significantly greater degree of agreement.
    - https://academic.oup.com/poq/article/79/1/145/2330061?login=true :
      - Using 11-point likert scale (0 – 10 vs. 10-0), significant bias towards left side
    - https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2423064 (label's value matters)
      - If negative labels are used (e.g. -4,.., 4), it is perceived more negatively (vs. 1,...,9)
      - -4,.., 4 produces more positive evaluations than 1,…,9

Explicit feedback: how scale influence behavior?

- How rating scale influence user rating behavior? https://www.tandfonline.com/doi/full/10.1080/0144929X.2017.1322145
  - You have 2*, 3* and 4* out of five star chart, how will you translate it to other rating schemes
    - Translation differs, but there are some similar outcomes

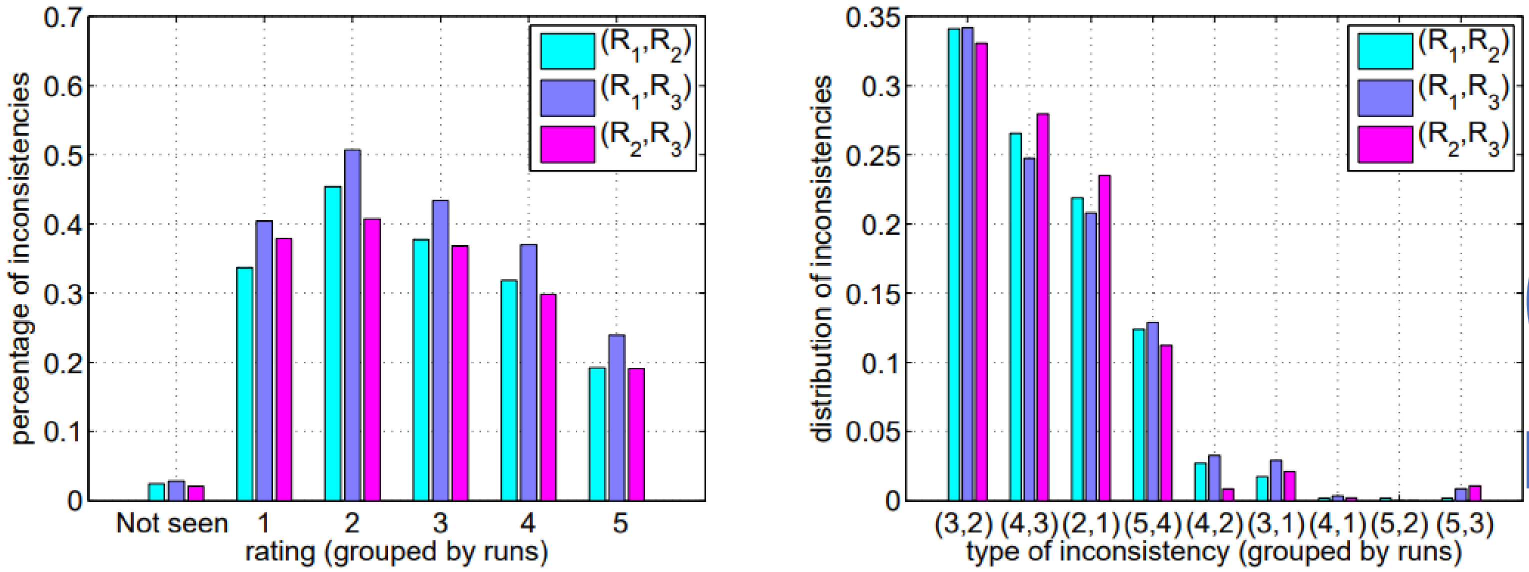


Explicit feedback: how scale influence behavior?

- Summary:
  - Be very careful while changing GUI or using external feedback data
    - Some transformation may be necessary

Explicit feedback: how much inconsistent it can be?

- https://xamat.github.io/pubs/xamatriain_umap09.pdf
- Netflix dataset, both popular & unpopular movies
- Three trials, 5-scale rating + unseen of 100 movies: 1->2 at least one day apart, 2->3 at least 14 days apart, different ordering of items (random -> popular -> random)

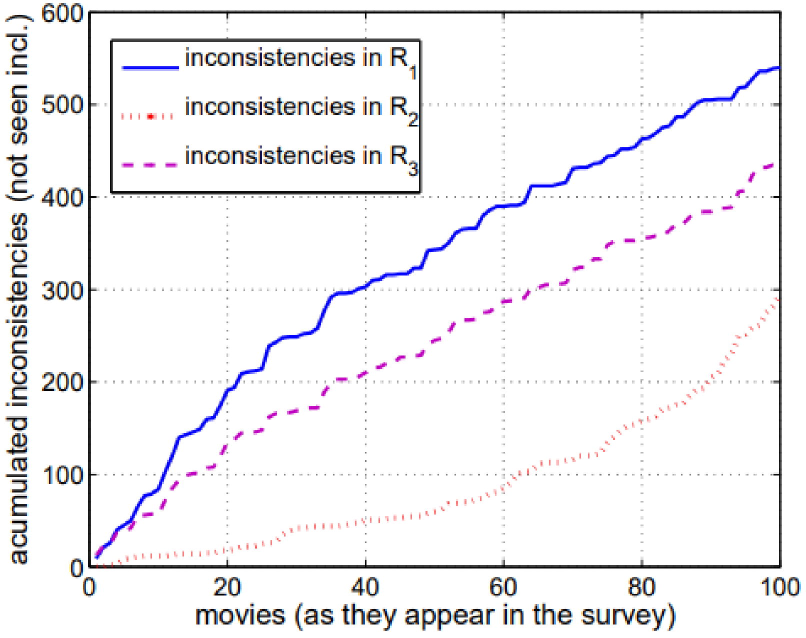


Usually, rating differs just by one point; mediocre movies more unstable

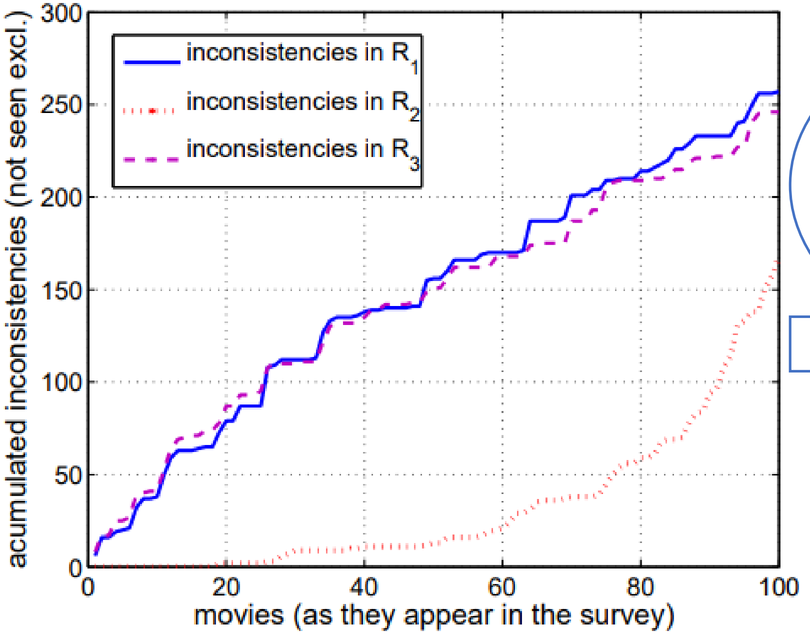
Fig. 2: Users Inconsistencies. (a) Percentage of inconsistencies by rating value and (b) Distribution of types of inconsistencies

Explicit feedback: how much inconsistent it can be?

- https://xamat.github.io/pubs/xamatriain_umap09.pdf
- Netflix dataset, both popular & unpopular movies
- Three trials, 5-scale rating + unseen of 100 movies: 1->2 at least one day apart, 2->3 at least 14 days apart, different ordering of items (random -> popular -> random)



(a) Taking into account "not seen" values.

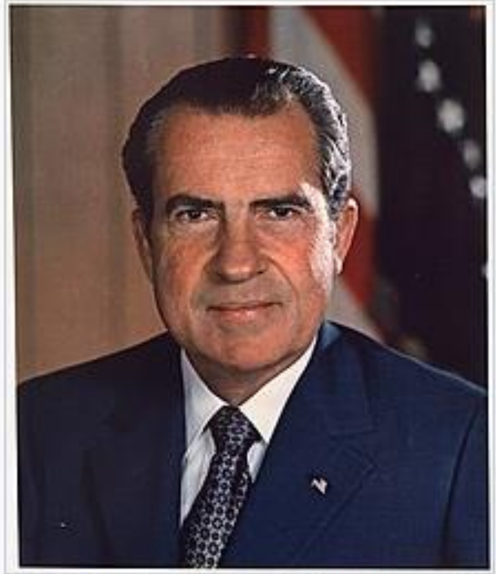


(b) "Not seen" values not taken into account.

Inconsitencies are more frequent in less popular (less known?) items

Explicit feedback: how much inconsistent it can be?

- Assimilation/Contrast effect on sequence of ratings
- https://psycnet.apa.org/record/2001-01676-002
- https://en.wikipedia.org/wiki/Assimilation_and_contrast_effects
  - Main findings:
    - a user is likely to give a lower rating to an item if the preceding one deserved a very high evaluation.
    - However, if successive items are comparable in their ratings, the user is likely to assimilate the second item to the preceding one and give the same rating to both



Thinking about Richard Nixon, a politician strongly associated with scandals, decreases the perceived trustworthiness of politicians in general (*assimilation effect*), but increases the perceived trustworthiness of every other specific politician assessed (*contrast effect*).[1]

Explicit feedback: how much inconsistent it can be?

- Summary:
  - Test robustness of your models against small rating variations (as they may be slightly unstable)

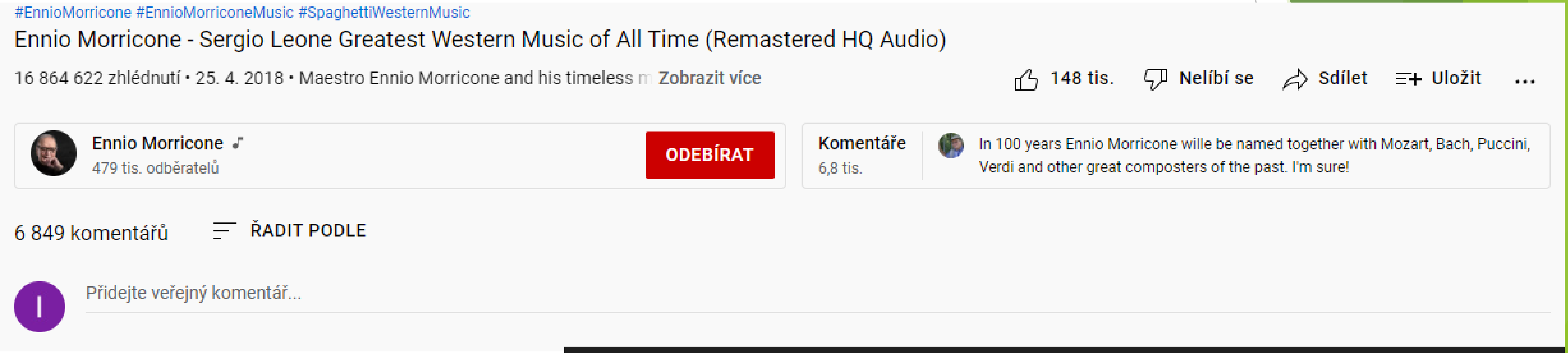
Explicit feedback: other variants

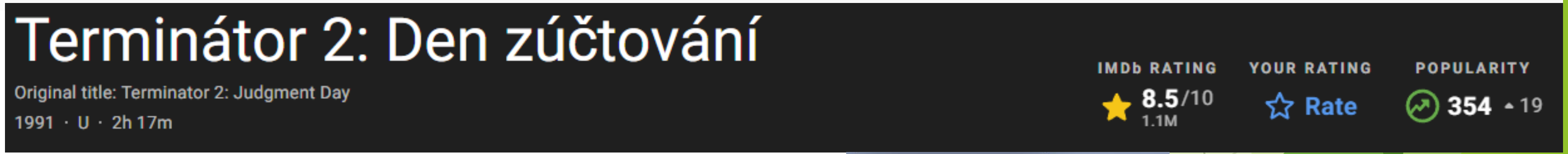
- Explicit comparison of items / groups of items
  - Never seen outside of preference elicitation models
  - And one dating app...
  - ***May be fun for users -> higher engagement***... But only for specific use-cases
- Explicit rating of item’s attributes
  - Multimodal rating
  - Not frequent, but relevant for well-defined cases
    - Booking.com example
    - Does not have to map to „attributes“ as defined for item
- Writing a review (is it really explicit feedback?)
  - Emotion/polarity detection, feature detection
  - Details maybe later – if enough time

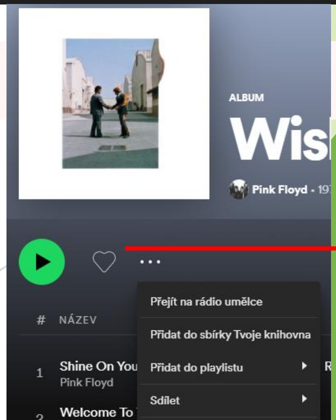
Explicit feedback

How does the industry feel about that?

Explicit feedback in industry







Artist, album & track

Explicit feedback in industry

Pánská mikina KILPI LAURER-M černá

kilpi

XS S M L XL XXL 3XL

Kilpi hodnocení

31 spokojených zákazníků

1 nespokojený zákazník

Jaká byla velikost produktu oproti očekávání?

+ Velikost sedí (27)
− Je mi to malé (4)
− Je mi to velké (1)

Jak jste spokojeni s kvalitou produktu?

+ Materiál je uspokojivý (22)
+ Kvalitní materiál (10)

Explicit feedback in industry

Explicit feedback in industry

Miroslav, Praha 8

Hodnoceno 22.10.2020, varianta 32" ViewSonic VX3276-2K-MHD

★★★★☆ Ověřený nákup

- \+ velikost
- \+ obraz
- \+ vzhled
- \+ cena

- − zvuk z reproduktoru

---

dejv

Hodnoceno 25.09.2019, varianta 32" ViewSonic VX3276-2K-MHD

★★★★★

Dlouho jsem poohlížel po levném a velkém monitoru. Tento monitor je v nízké cenové relaci a jak se říká za málo peněz hodně muziky

- \+ Cena
- \+ Velikost
- \+ Kvalita
- \+ IPS

- − zatím žádné

Explicit feedback in industry

Some users are sometimes willing to

- **Provide ratings**
  - Sometimes aspect-based ratings (mostly pre-defined, widely recognized categories)
    - Does not have to correspond to object's attributes directly
  - Write review

- *Explicit / implicit borderline:*
  - *Add object to some list / organize favorite objects / provide tags for them*
  - *Share items*

- ***Is this frequent enough so we can infer preferences of individual users?***

Implicit feedback

User’s actions will speak for themselves…

Implicit feedback

- **Server-side (limited expressibility)**
- **Client-side (triggered JS events)**
- **Beyond (eye tracking, other biometrics)**
  - Limited applicability (lab studies)
  - Can provide leads on interpretation of the previous two

Implicit feedback

Server-side

- Stream of visited pages
- Asynchronous loading of page content (e.g. more results)
- Proxy for time on page / dwell time (very coarse)
  - http://www.hongliangjie.com/publications/recsys2014.pdf (RecSys 2014 best paper)

- Not much information available
  - But non-intrusive & cannot be turned off or altered easily

Implicit feedback

Client-side

- Any JS event can be captured, processed and stored…
  - But which ones are relevant?
    - And also… what is their semantics? Does it differ from explicit feedback?
  - How to interpret implicit feedback?
  - How to establish negative preference from implicit feedback?

- *Peska, IPIget: The Component for Collecting Implicit User Preference Indicators*
  https://www.researchgate.net/publication/305495313_IPIget_The_Component_for_Collecting_Implicit_User_Preference_Indicators

What to Collect as Implicit feedback

Not very explored area

- Domain dependence (how surprising☺)
- Mostly, academic researchers work with pre-collected datasets
  - The decision on what to collect was already done
- Not many known industry papers with details on implicit feedback collection

However…

What to Collect as Implicit feedback

Not very explored area

However… common identifiers (cummulative feedback):

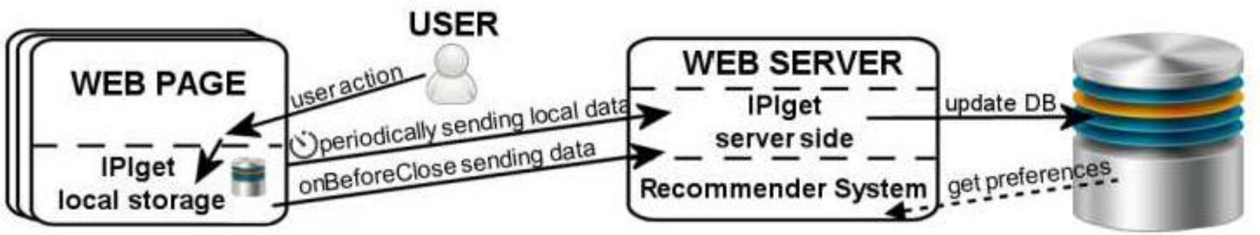
- (count of) page visits => object visits
- Time on page / dwell time
  - Beware to count only while focus is on the page (http://www.hongliangjie.com/publications/recsys2014.pdf)
- Objects consumption statistics (playcounts, viewtime, purchase, add to basket,…)
- **!!! Impressions !!!** (what was shown to the user)

What to Collect as Implicit feedback

Peska, IPIget: The Component for Collecting Implicit User Preference Indicators
https://www.researchgate.net/publication/305495313_IPIget_The_Component_for_Collecting_Implicit_User_Preference_Indicators

- Main target: small e-commerce vendors
  - Previously mentioned events
  - Other aggregated events: print, search, copy, text selection (not much usable)
    - Non-numeric data (searched text, selected text,…)
  - Context of events
    - Scrolling to coordinates
    - Mouse position sampling
    - Mouse over pre-defined elements
  - Basic page statistics
    - Vol. Of text, images, links
    - page dimensions, window dimensions
    - position of elements
    - Page params (e.g. Catalogue, menswear,…)

Identification	User ID Page ID Session ID
Page parameters	Images count: total number of images Text size: number of letters of textual content Links count: total number of links Page size: vertical and horizontal size of the page in pixels Window size: vertical and horizontal size of the browser visible window Object list: list of displayed objects within the page and their respective positions Page variables (e.g. searched text, page type etc.)
Time	Start and end of visit
Numeric preference indicators	Time on page: total time spent on the page Page view count: # the page was opened Mouse click count: # of the mouse click event Mouse moving time: total time the mouse cursor was moving Mouse distance: total distance traveled by the cursor Scrolling time: total time the page was scrolled Scrolling distance: total distance scrolled Print page count Select text count: number of times the text was selected Copy text count: number of times the copy action was initialized Forwarded to link count: number of times the user followed a link from this page Purchase process start: user started the process of purchasing an item Purchase finished: user finished the purchase successfully
Non-numeric preference indicators	Selected and copied text Description of links user followed Log file



What to Collect as Implicit feedback

Results

Dwell time: 16.8 sec

Mouse moving time: 1.8 sec

Travelled distance: 2009px

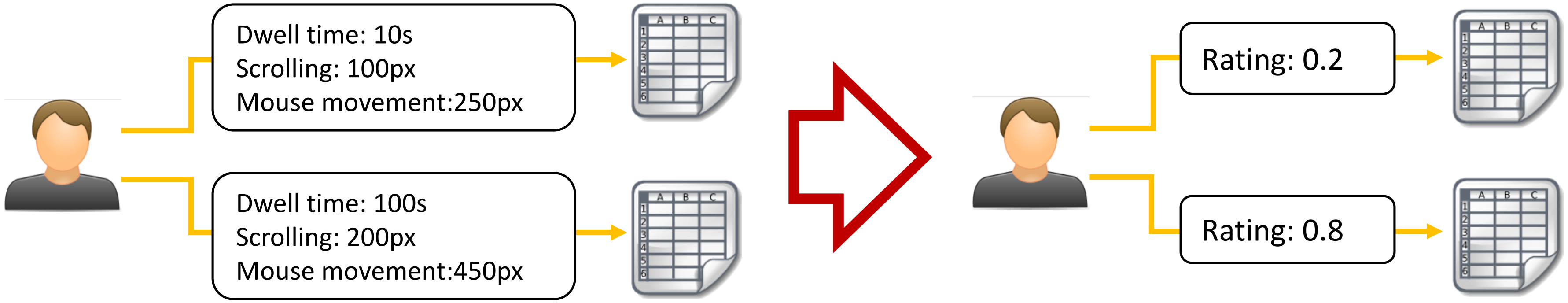
What to Collect as Implicit feedback

- Collect visible area through time (scrolling position + window dimensions)
- Store areas covered with page components
  - Items in category page
  - Areas focused on item's features
- Calculate visibility => noticeability of individual components
- If the item is clicked, it should be more preferred than not-clicked ones with high-enough noticeability

https://link.springer.com/article/10.1007/s13740-016-0061-8

How to interpret numeric implicit feedback

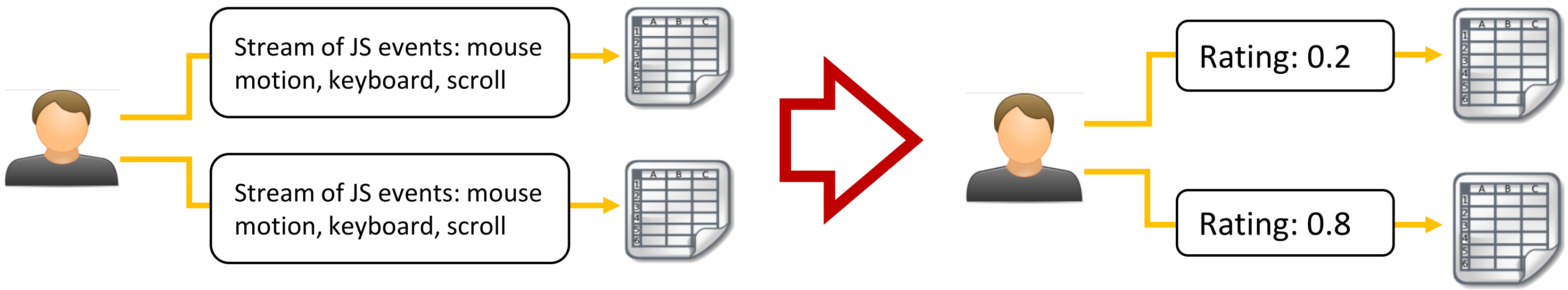
- Combine all implicit feedback features to estimated user rating
  - Standard recommender systems can be used afterwards



- The more the better hypothesis
  - Normalize data (Time on page vs. Scrolling distance vs. Vol. of visits)
  - (shifted) standardization, cummulative distribution function, log transformation
    - Make a hypothesis about what particular values mean and then confirm it via user study

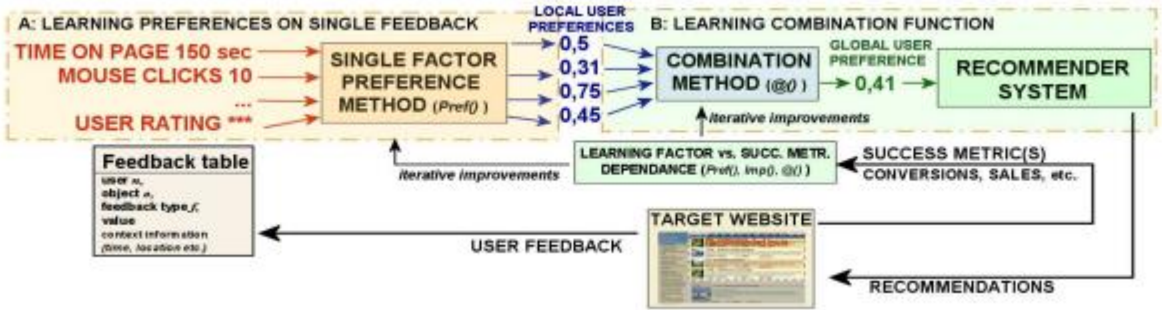
How to interpret numeric implicit feedback

- Combine all implicit feedback features to estimated user rating
  - Standard recommender systems can be used afterwards



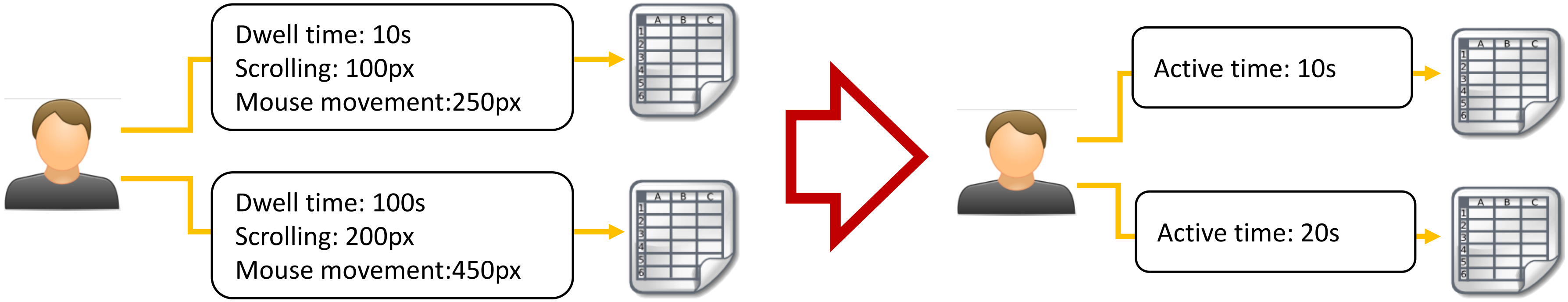
- Use feedback linked with positive/negative preference
  - Ratings, purchases
  - Train ML predictor to predict this based on other implicit feedback features
    - Note that positive preference indicators are usually very sparse => bootstrap / stratified sampling / weighting
  - Make individual preference estimators per feedback type & their aggregator (wAVG, fuzzy logic,…)
    - In case of insufficient data or specific model in mind

https://www.researchgate.net/publication/283526661_How_to_interpret_implicit_user_feedback, https://www.ksi.mff.cuni.cz/~peska/wims13.pdf



How to interpret numeric implicit feedback

- Construct single (complex) implicit feedback based proxy for user preference
  - Standard recommender systems can be used afterwards



- Active dwell time *[not confirmed by literature]*
  - Time spent on page
  - But counted only if some other events are detected in close temporal proximity => user is active

How to interpret numeric implicit feedback

- Construct single (complex) implicit feedback based proxy for user prefer...
  - Standard recommender systems can be used afterwards

...med by literature]

...page

...counted only if some other events are detected in close temporal proximity => user is active

Is that all we can do?

How to interpret numeric implicit feedback

Is that all we can do?

- Negative Implicit Feedback
  - Low values of feedback features on particular object
  - Implicit feedback on object’s categories
- Context of User Feedback

Context of user feedback

How to interpret numeric implicit feedback

- Pages may substantially vary in length, amount of content etc.
  - This could affect perceived implicit feedback features
  - Leveraging context could be important
- Consumption statistics may significantly vary for different device types
  - (http://www.hongliangjie.com/publications/recsys2014.pdf)

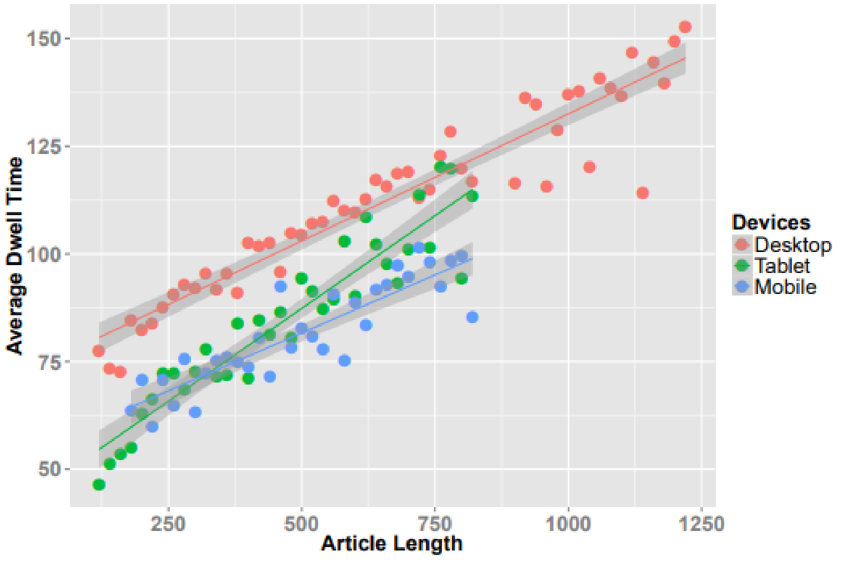


Figure 3: The relationship between the average dwell time and the article length where X-axis is the binned article length and the Y-axis is binned average dwell time.

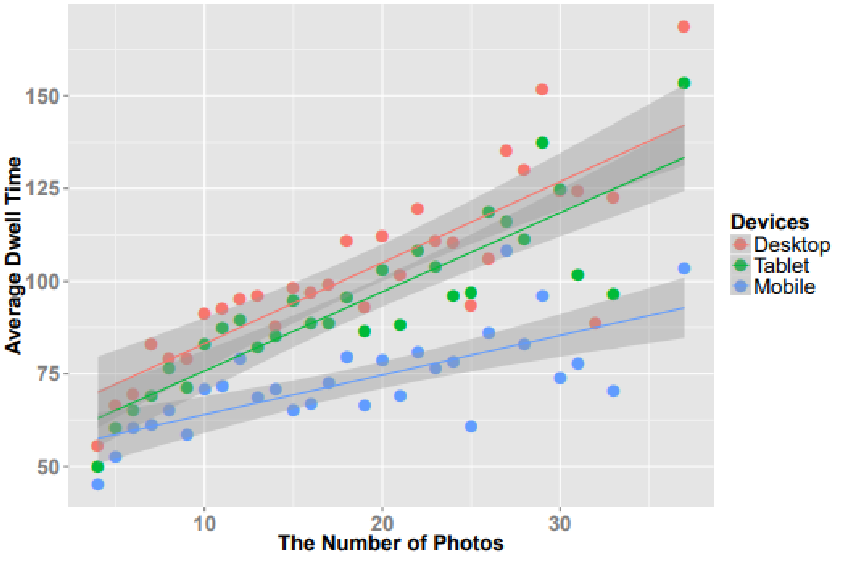
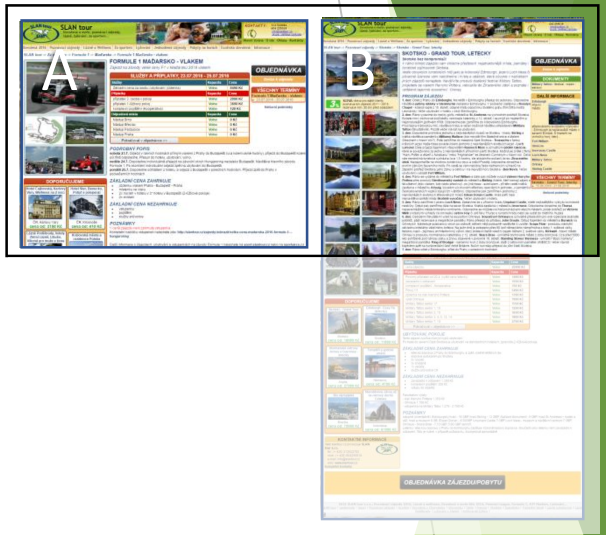
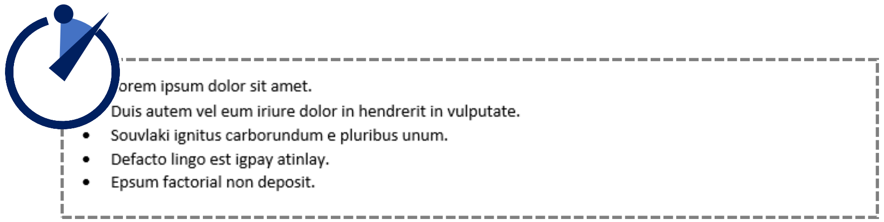


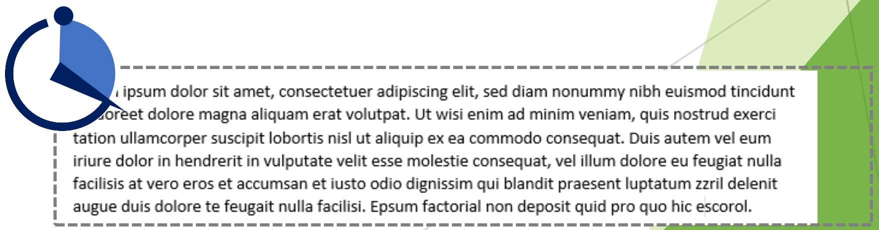
Figure 4: The relationship between the average dwell time and the number of photos on a slideshow where X-axis is the binned number of photos and the Y-axis is binned average dwell time.

How to interpret numeric implicit feedback

- Context of the user
  - Location, Mood, Seasonality...
  - *Can affect user preference*
  - *Out of scope of this paper (and this lecture☺)*
- Context of device and page
  - Page and browser dimensions
  - Page complexity (amount of text, links, images,...)
  - Device type
  - Datetime
  - *Can affect percieved values of the user feedback*







How to interpret numeric implicit feedback

- IPIget component for collecting user behavior

Implicit Feedback Features	
f_1	View Count
f_2	Dwell Time
$f_{3,4}$	Mouse Distance and Time
$f_{5,6}$	Scrolled Distance and Time
f_7	Clicks count
f_8	Hit bottom of the page
r	Purchase

Contextual features	
c_1	Number of links
c_2	Number of images
c_3	Text size
c_4	Page dimensions
c_5	Visible area ratio
c_6	Hand-held device

How to interpret numeric implicit feedback

Our approach

- Several imlicit feedback and contextual features are collected:

  $F_{u,o} = [f_1, \dots, f_i]$ $C_{u,o} = [c_1, \dots, c_j]$
- Learn estimated rating $\bar{r}_{u,o}$ for visited objects based on feedback and context
  - $F_{u,o}, C_{u,o} \rightarrow \bar{r}_{u,o} : o \in \boldsymbol{S}$
  - „The more the better" heuristics (STD, CDF)
  - Machine learning approach (dec. trees, lasso regression, ada boost)
- Incorporate context
  - As further feedback features (pass it on to the ML algorithm)
  - As baseline predictors (what is the average feedback for this context value?), re-scale actual values
- Learn rating on all objects as in traditional recommenders

  $R_u \rightarrow \hat{r}_{u,o'} : o' \in \boldsymbol{O}$

1. For each content consumption context C, collect the historical per-item time spent data and compute the mean μ_C and standard deviation σ_C, both in log space.
2. Given a new content item i's time spent t_I in its context C_i, compute the z-value in log space: $z_i = \frac{\log(t_i) - \mu_{C_i}}{\sigma_{C_i}}$.
3. Compute the normalized dwell time of item i in the article space: $t_{i,\text{article}} = \exp(\mu_{\text{article}} + \sigma_{\text{article}} \times z_i)$.

https://arxiv.org/pdf/1612.04978.pdf
https://dl.gi.de/handle/20.500.12116/916
http://www.hongliangjie.com/publications/recsys2014.pdf

Negative preference from implicit feedback

Can my consumption say I dont like it?

Negative preference from implicit feedback

Object detail level

- If more is better… „not-enough“ might mean I do not like it?
  - Where is the borderline?
- Google Analytics: bounce rate (leaving the page immediately after openning it)
- But why?
  - Did I waited too long to load page?
  - I clicked on it accidentally?
  - I found sth. better in the meantime?
  - The short description looked good, but it was missleading / did not cover important drawbacks
    - *Would this transfer into decreased feedback values?*

Negative preference from implicit feedback

List of objects (impressions needed)

- If I (repeatedly) ignore it, I probably dislike it
  - How many times do I have to ignore it?
  - Could it be that I just did not pay attention for this specific part of the page?
  - What is the chance that I changed my mind?
- We can consider uniform chance of item being unnoticed
- We can consider fixed chance of being unnoticed for certain position
  - https://link.springer.com/article/10.1007/s11257-021-09311-w
- We can consider that items are evaluated sequentially
  - If the item below was clicked, this one is probably observed as well
  - TODO: ref
- We can have detailed feedback with objects’ visibility information
  - https://link.springer.com/article/10.1007/s13740-016-0061-8

Implicit Feedback

How does the industry feel about that?

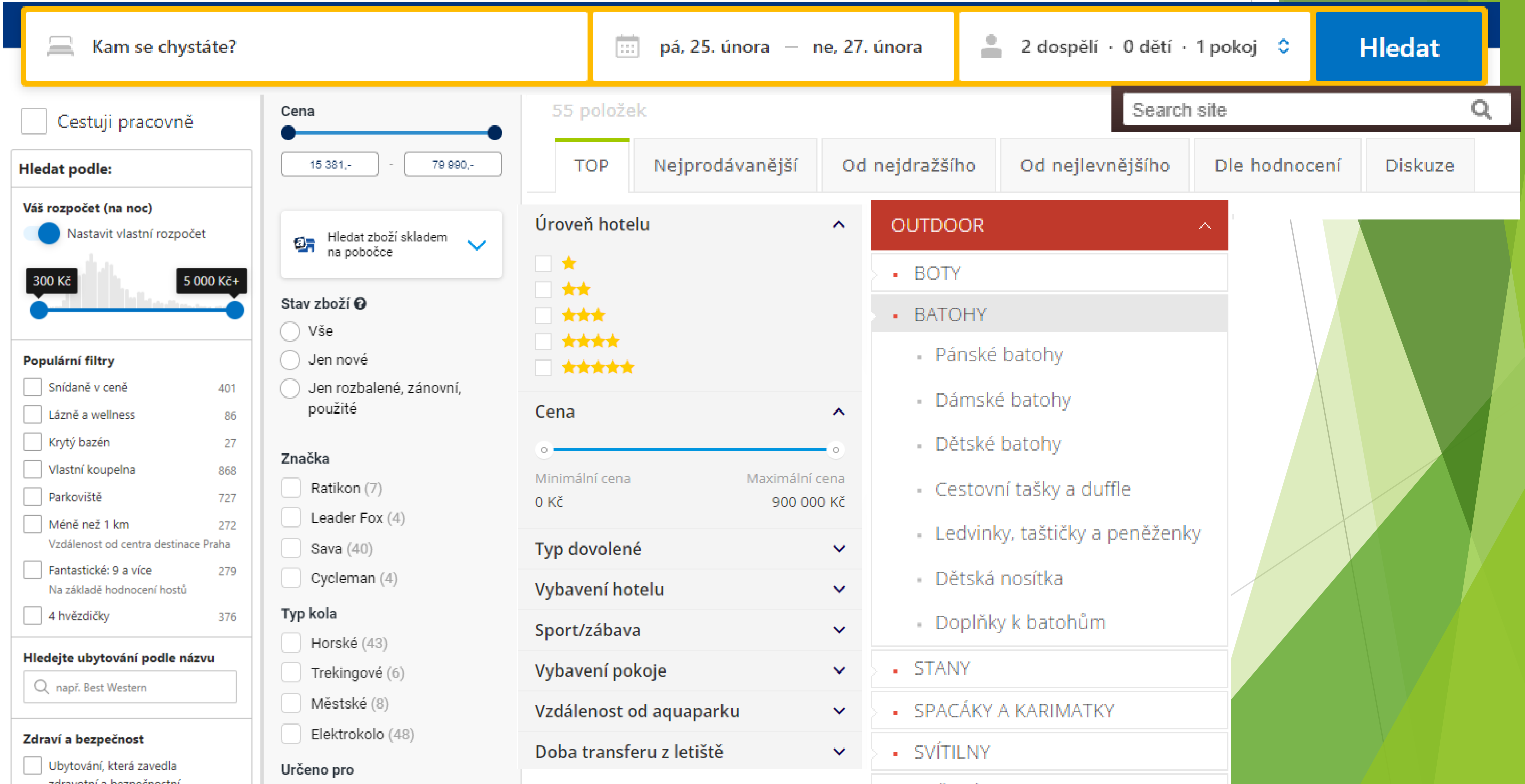
Implicit Feedback

How does the industry feel about that?

???

Searching and filtering as feedback

What would the user be willing to do?



What would the user be willing to do?

Most users do:

- **Filter content manually**
  - Browse categories
  - Apply facet search
    - Mostly direct mapping to object’s attributes
  - Use fulltext search
  - Can be utilized in the construction of attribute-level preferences
    - Beware of long-term preferences vs. short-term goals

All users do:

- Evaluate & consume content:
  - Browse items, open details, read content, play, purchase,…
  - Preferences based on implicit feedback

How to model UP

Tenative solutions for show-cases

How to model UP

Simple movies recommendation:

- **Task: discover what to watch tonight**
- How to use UP: Collaborative recommendation of movies

Basic model of UP:

- Preferences on movies (rating, watching)
  - If insufficient data: openning movie details, top search results

Enhancements:

- Learned (confirmed) preferences towards genres (multiple confirmation, enough data)
- Learned (confirmed) preferences towards other named entities (actor, director)
- List-wise preferences (Y was selected from results of XYWZ)
  - Remember impressions, not just usage

How to model UP

(Food) Recipes recommendation:

- **Task: help to decide what to cook**
  - How to use UP: **personalized searching, front-page recommendation**

Basic model of UP:

- Preferences on recipes (likes, add to list, reading sufficiently long)
- Preferences on ingredients (search count, contained in prefered recipes, confirmation?)
  - Ingredients granularity?

Enhancements:

- Learned preferences towards tags & attributes
  - Verify on a well-known subset of users (RecSys OPS)
- Best out of similar choices
  - Which goulash does the user prefer? Would that say something more generic about him/her?
- Should we allow users to further refine recommendations?
  - Faceted recommendations (https://www.researchgate.net/publication/301321425_FeRoSA_A_Faceted_Recommendation_System_for_Scientific_Articles)

How to model UP

Group music recommendation:

- **Task: create a background music playlist for an evening with friends**
- How to utilize it: fairness-aware playlist construction

- Individual preference
  - Track -> Album -> Artist (playcount, play from search, likes)
  - Maybe, preferred sequences (low-level audio analysis, but probably not for individual users)

- Group preferences
  - Playlist modifications

Preference Elicitation

Preference Elicitation

- **[WIKI] Preference elicitation** refers to the problem of developing a [decision support system]() capable of generating [recommendations]() to a user, thus assisting in decision making. It is important for such a system to model user's preferences accurately, find hidden preferences and avoid redundancy.
- Not really a definition
- The process of collecting user preferences to support decision making systems
  - Often considered w.r.t. restricted meaning of initial preference elicitation
  - Usually restricted to explicit feedback

Traditional methods (2004):

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.319.8057&rep=rep1&type=pdf

Preference Elicitation

- Additive independence of preferences
  - Preferences of items is a function of it's features preferences (wAVG)

Mutual Preferential Independence: The attributes $X = \{x_1, ..., x_n\}$ are mutually preferentially independent if every subset Y of X is preferentially independent of its complementary set.

Theorem of Additive Value Function: Given atributes $X = \{x_1, ..., x_n\}$, $n \geq 3$, an additive value function $v(x_1, ..., x_n) = \sum_{i=1}^{n} \lambda_i v_i(x_i)$ (where v and v_i are scaled from zero to one, and $\sum_{i=1}^{n} \lambda_i = 1, \lambda_i > 0$) exists if and only if the attributes are mutually preferentially independence.

Additive Independence: If the value function can be wrote as additive model, namely the condition of mutually preferentially independence is met, the attributes are said to be additive independent.

Preference Elicitation

- Additive independence of preferences

Mutual Preferential Independence: The attributes $X = \{x_1,\ldots,x_n\}$ are mutually preferentially independent if every subset Y of X is preferentially independent of its complementary set.

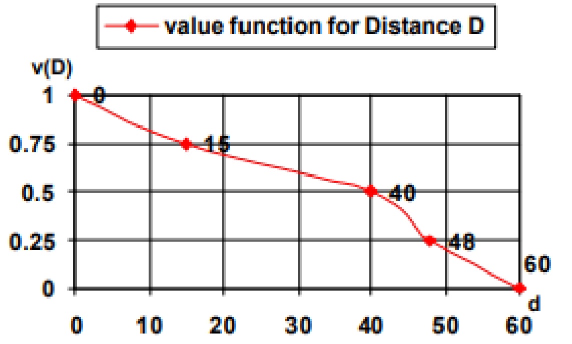
Theorem of Additive Value Function: Given atributes $X = \{x_1,\ldots,x_n\}$, $n \geq 3$, an additive value function $v(x_1,\ldots,x_n) = \sum_{i=1}^{n} \lambda_i v_i(x_i)$ (where v and v_i are scaled from zero to one, and $\sum_{i=1}^{n} \lambda_i = 1, \lambda_i > 0$) exists if and only if the attributes are mutually preferentially independence.

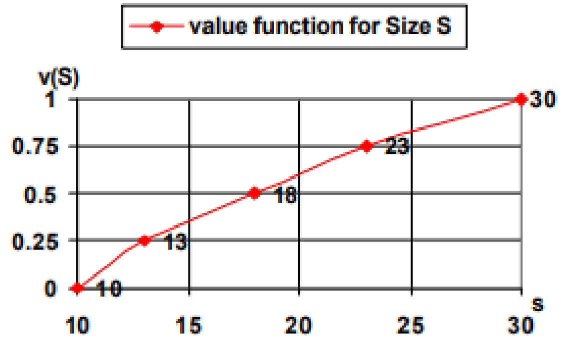
Additive Independence: If the value function can be wrote as additive model, namely the condition of mutually preferentially independence is met, the attributes are said to be additive independent.

- Similar as LMPM – only value functions does not have to be linear

Question	Hypothesized answer
1. Suppose you are at Size=20. Would you pay more of Size to change *Distance* from 60 to 30 or 30 to 0?	I would pay more to go from 60 to 30.
2. More to go from 60 to 50 or 50 to 0?	More to go from 50 to 0.
3. Give me a value, d ' say, such that you would give up the same in *Size* to go from 60 to d 'as from d 'to 0.	About x '=40
4. In our language, 40 is the midvalue point between 0 and 60. We label 40 by $d_{.5}$. What is your midvalue point between 0 and 40?	Let's say 15, I'd pay the same to go from 40 to 15 as 15 to 0.
5. In that case $d_{.75}$ =15. What is your midvalue point between 40 and 60?	Oh, about 48
6. This means that $d_{.25} = 48$. Does 40 seem like a good midvalue between 15 and 48?	Sure
7. Now let's turn to the *Size* value. What is the midvalue point between 10 and 30?	Say, 18.
8. The midvalue between 18 and 30?	Say, 23.
9. The midvalue between 10 and 18?	13.

Then we can plot these few points and fairs in the curves of v_D (distance) and v_S (size).





Preference Elicitation

- Knowledge-based RS with preference elicitation
  - Start either with known example
  - Or initial search

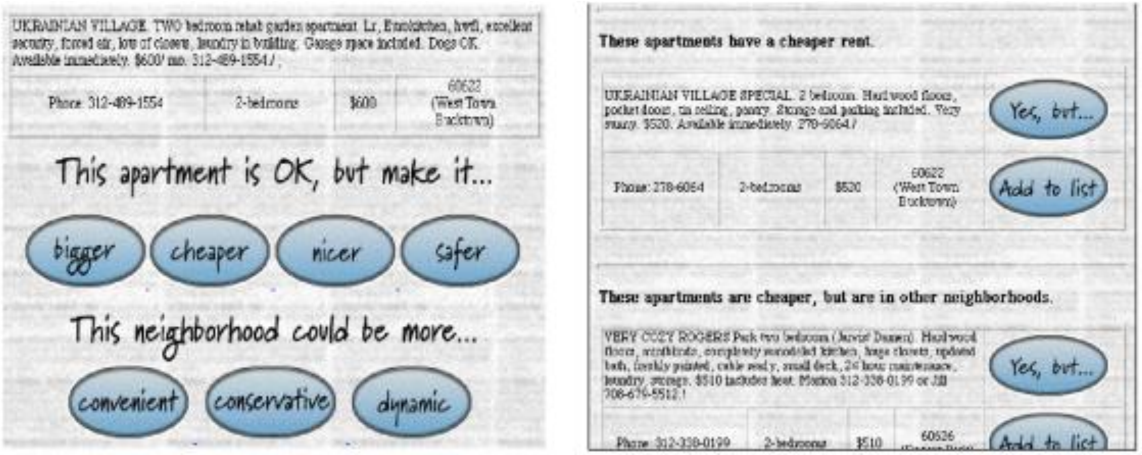


Figure 2 Tweaking an apartment in RentMe

Preference Elicitation

- Choice-based preference elicitation for collaborative filtering recommender systems https://dl.acm.org/doi/10.1145/2556288.2557069
  - Not based on meta-data, but latent factors
    - „The basic idea behind our approach is, thus, to use latent item features derived from the rating matrix and request preferences for sets of similar items instead of single items.“
    - „Since the number of interaction steps needed should be minimized, we developed a technique based on latent factors to achieve a maximum information gain with each choice.“

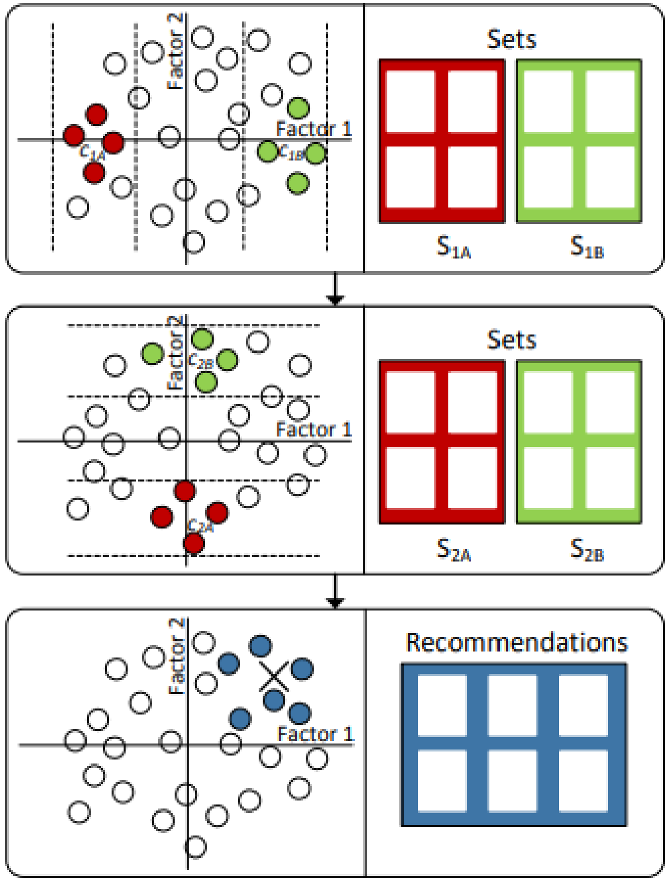
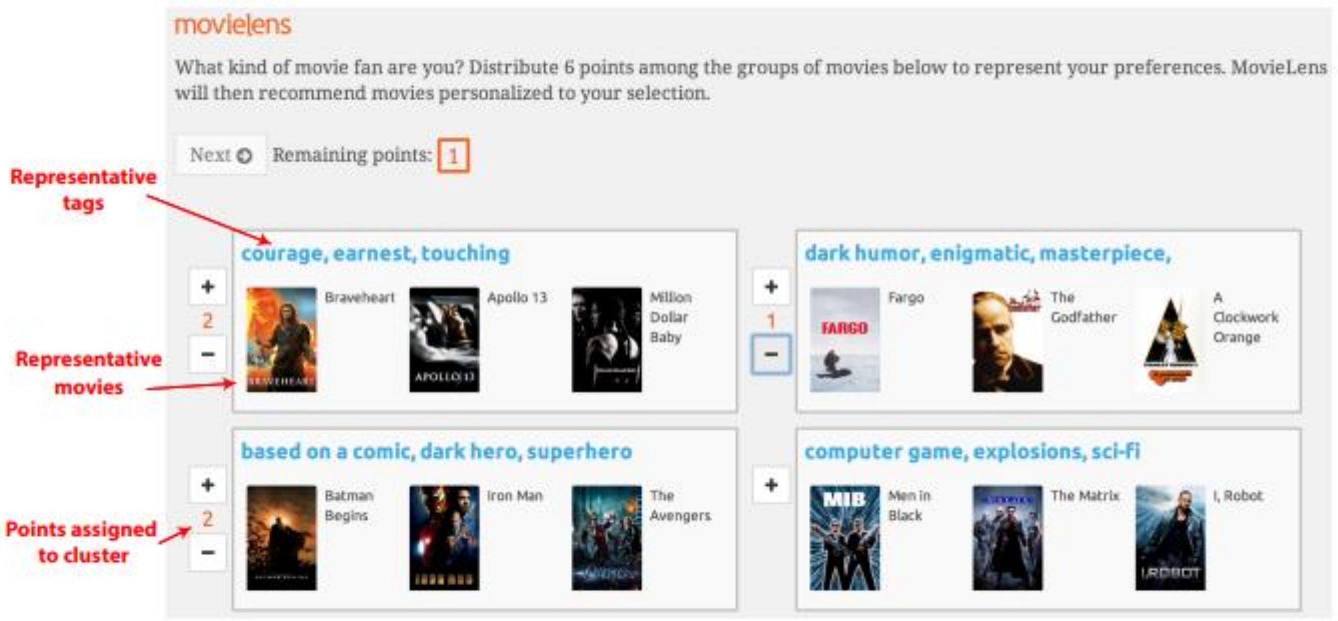


Figure 2. For each factor f taken into account, two sets of movies S_{fA} and S_{fB} are presented to the user. One set shows movies with low factor values, the other movies with high factor values. The user selects one of these sets (or indicates that he/she doesn't care). After a defined number of steps, a set of recommendations is computed.

Preference Elicitation

Using Groups of Items for Preference Elicitation in Recommender Systems
https://dl.acm.org/doi/pdf/10.1145/2675133.2675210

- New users can begin by expressing their preferences for groups of items
  - Utilize clustering to generate groups
  - Based only on movie ratings
  - For each cluster: select tags, then select best matching movies
  - Get avg. ratings of users with similar cluster prefs.

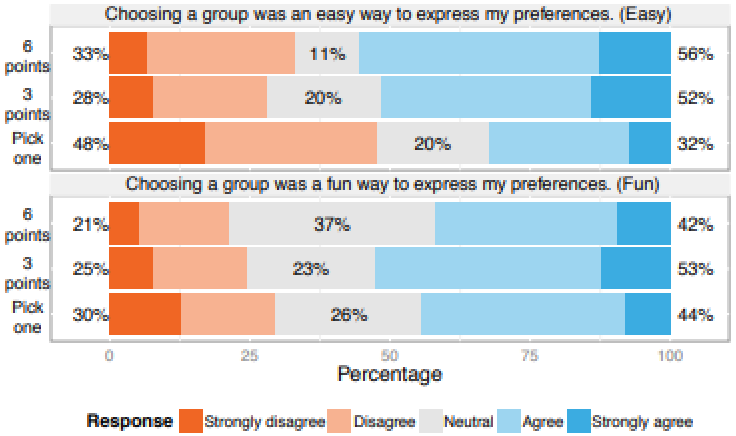


For each movie group, we first pick the top-three tags that both *uniquely describe* and are *highly relevant* to the group. Therefore, we define the measure of tag uniqueness as Equation 1 and tag relevance as Equation 2. We pick the three tags with the highest multiplication of uniqueness and relevance. (Multiplication is used to handle different scales of the two metrics.)

$$unique(t, c) = \frac{rel(t, c)}{\sum_{c_i \in C} rel(t, c_i)} \tag{1}$$

$$relevance(t, c) = \frac{rel(t, c)}{\sum_{t_i \in T_c} rel(t_i, c)} \tag{2}$$

where t denotes one of the tags T_c that appears in cluster c, and C denotes all the clusters. Note that $rel(t, c)$ is the aggregated relevance of tags t to all movies in cluster c. In our implementation, we use relevance between a tag and a movie generated from the Tag Genome [31],



Preference Elicitation

Ordered Preference Elicitation Strategies for Supporting Multi-Objective Decision Making
https://arxiv.org/pdf/1802.07606.pdf

- Utilize full ranking of items
  - User starts with two items, then iteratively place one more item at each step
  - How to select what to ask?
    - Gaussian process (model mean and variance for each datapoint) (https://ebonilla.github.io/gaussianprocesses/, https://github.com/chariff/GPro)
    - Expected improvement acquisition function (https://www.csd.uwo.ca/~dlizotte/publications/lizotte_phd_thesis.pdf)

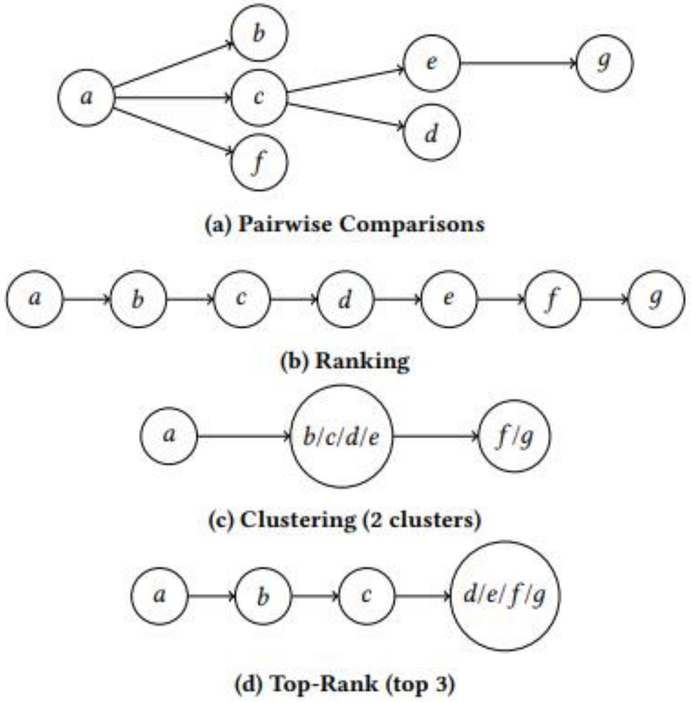
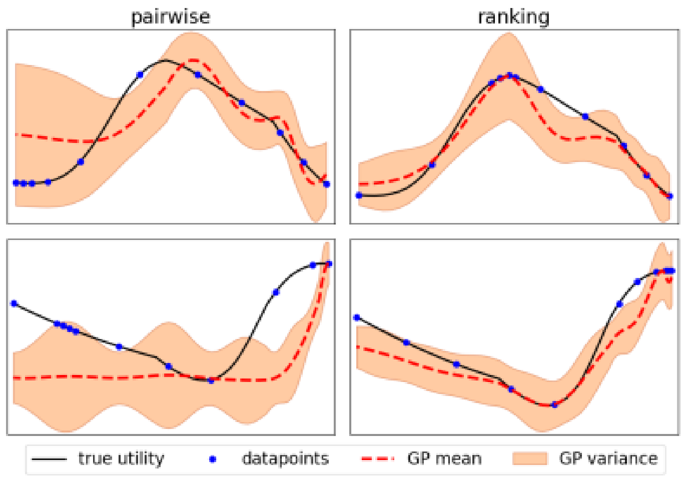


Figure 2: Possible outcomes of different query types for items a-g, with utilities $u(a) > \ldots > u(g)$. The arrows represent the preference information expressed by the user (preferred → unfavoured). Different elicitation strategies lead to different orderings: full ranking returns a total ordering (b); the other query types typically lead to partial orderings.



Preference Elicitation

- https://dl.acm.org/doi/pdf/10.1145/2792838.2796554 (Healthy recipes recommendation)
  - What was the main cause of your decision?
    - Video: https://onedrive.live.com/?authkey=%21ALYePnW0fOCHOUQ&cid=60DC0855E37985A6&id=60DC0855E37985A6%2149418&parId=60DC0855E37985A6%2149101&o=OneUp
    - Relatively simple tag-based approach

Preference Elicitation

- https://www.frontiersin.org/articles/10.3389/frobt.2017.00071/full (Constructive pref. Elicitation)
  - There exist many types of queries, like lotteries, pairwise or setwise rankings, improvements, which all share the goal of being easy to answer to and as informative as possible.
    - Choice set feedback
    - Coactive feedback (how to slightly improve a solution? – can be done from implicit feedback) https://www.jair.org/index.php/jair/article/view/10939
    - Example critiquing
  - Queries involving comparisons and rankings have come to be predominant in the literature with respect to quantitative evaluations.
  - Indeed, users are typically more confident in providing qualitative judgments like “I prefer configuration *y* over *y*′” than in specifying how much they prefer *y* over *y*′ (Conitzer, 2009; Carson and Louviere, 2011).