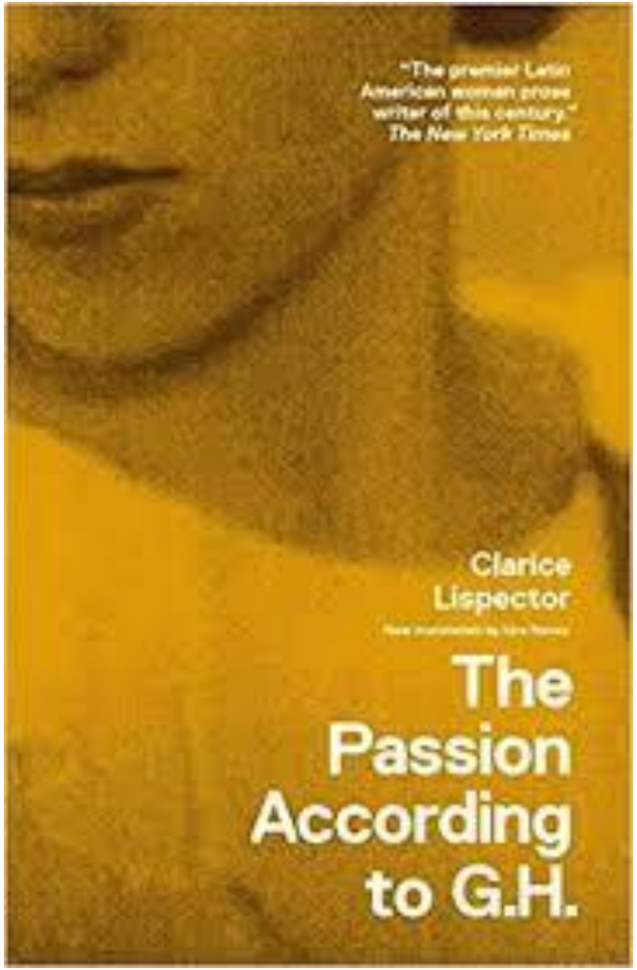
Journey to Personalisation

Dr Mahbub Gani
Lead Data Scientist

@MNonmad @Bibblio_org #RecSysLDN

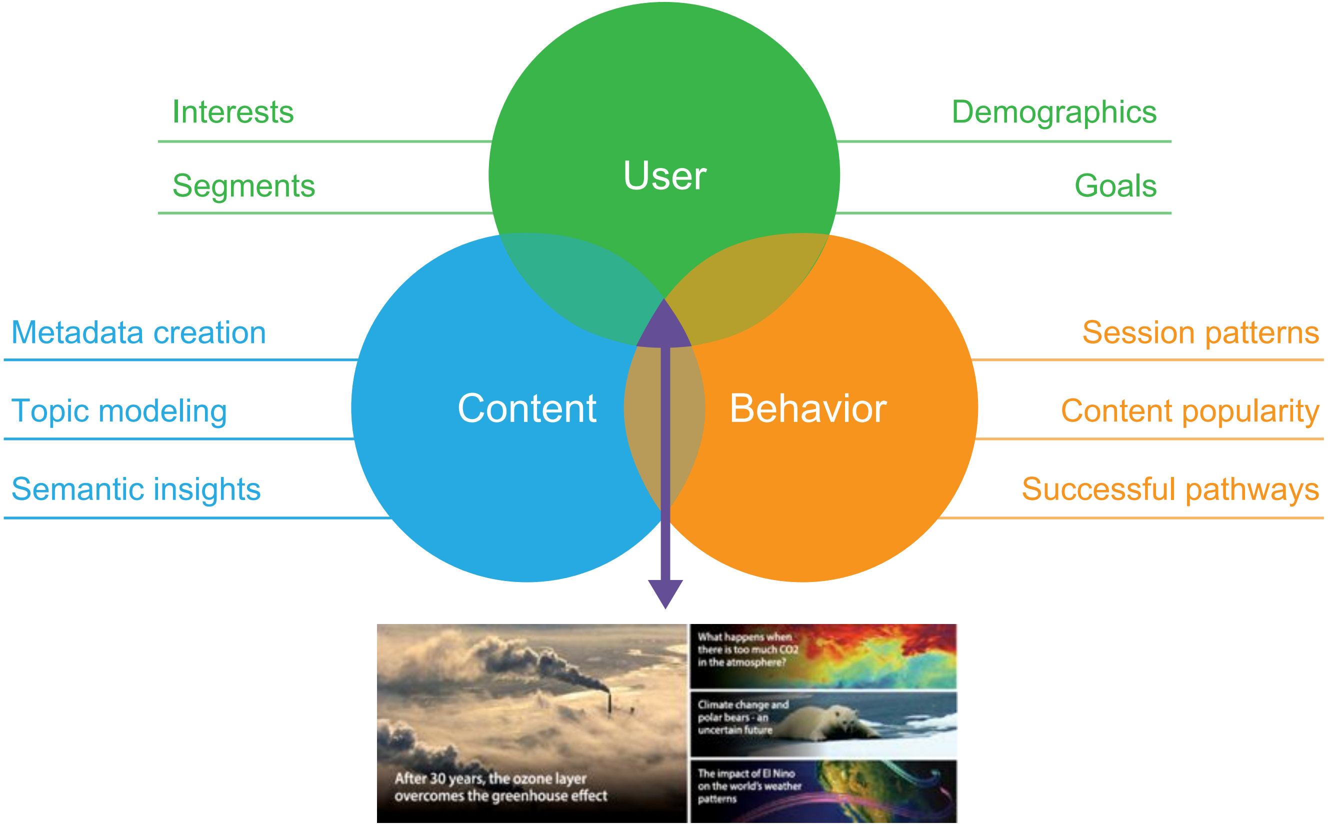


“Depersonalization like the deposing of useless individuality — the loss of everything that can be lost, while still being. To take away from yourself little by little, with an effort so attentive that no pain is felt, to take away from yourself like one who gets free of her own skim, her own characteristics. Everything that characterizes me is just the way I am most easily viewed by others and end up being superficially recognizable to myself.”

Content

- What Personalisation means for Bibblio
- Data availability
- Experiments in session identification
- Three Kings of Collaborative Filtering
- Model construction and training
- Evaluation
- Final Reflections

The ingredients of engaging recommendations



What to do when session and user ids aren't available

Missing user identifiers

- Generate proxy user ids using other forms of fingerprinting.
- E.g. device IP and agent information.
- (Need to explore constraints introduced by GDPR).

Missing session identifiers

- Approximate detection of contiguous chain of events by constructing event graph.


THREE
KINGS

Three Kings of Collaborative Filtering

Alternative Least Squares with Implicit Feedback

"Collaborative Filtering for Implicit Feedback Datasets," Hu, Koren, Volinksy (2008)

Python library implicit by Ben Fredrickson

Factorisation Machine with Bayesian Probability Ranking

"Factorization Machines," S. Rendle (2010)

C++ library libfm with BPR extension by Fabio Petroni

Matrix Factorisation with Metadata Embeddings

"Metadata Embeddings for User and Item Cold-Start Recommendations," M. Kula (2015)

Python library lightfm by Lyst

Implicit ALS

Key Concept

- Distinguish between user preferences for an item (to be estimated), and our (given) confidence (e.g. affine function of ratings, or number of clicks) over those preferences.
- Initially these preferences are set at 1 or 0 (based on whether item is clicked or not clicked respectively).
- An initial assignment of “no” preference will receive a low confidence (item may not have been clicked simply because it wasn’t seen).

Cost function

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

Solution strategy

Alternating least squares for user and item factors with some pre-computation to improve efficiency.

Factorisation Machines

Key Concept

- Include higher-order interaction terms.
- Instead of assigning independent parameter to each interaction, “factorise” interaction parameters so that parameters can be shared thereby improving learning under sparsity.

Cost function

$$\textsc{OptReg}(S,\lambda) := \underset{\Theta}{\operatorname{argmin}} \left(\sum_{(\mathbf{x},y)\in S} l(\hat{y}(\mathbf{x}|\Theta), y) + \sum_{\theta\in\Theta} \lambda_\theta \theta^2 \right)$$

$$l^{C}(y_1, y_2) := -\ln \sigma(y_1 y_2)$$

Solution strategy

- Stochastic gradient descent.
- Rendle showed how model can be “computed in linear time” by “completing the square”.
- “Learning to rank” using BPR.

Matrix Factorisation with Metadata

Key Concept

- Each user and item is characterised by (sum of) lower level features.
- Feature vectors are dense embeddings in the same factor space.
- The scores are derived by taking dot-product of the representation of the users and items in latent factor space.

Cost function

$$L\left(\mathbf{e}^U, \mathbf{e}^I, \mathbf{b}^U, \mathbf{b}^I\right) = \prod_{(u,i)\in S+} \widehat{r}_{ui} \times \prod_{(u,i)\in S-} (1 - \widehat{r}_{ui})$$

$$\widehat{r}_{ui} = f\left(\boldsymbol{q}_u \cdot \boldsymbol{p}_i + b_u + b_i\right)$$

Solution strategy

- Stochastic gradient descent (asynchronous).
- “Learning to rank” via BPR or WARP.

Model Construction and Training

- Dataset consisting of ~41k clicked events, ~26k users and ~4k items
- (Biased) train/test split:
  - Randomly select 20% of users as candidates for test set
  - Hold back 50% of the data instances belonging to the test users for the test set
- Optimise hyper-parameters using sequential model-based optimisation (scikit optimisation Python API)
- Use Precision@3 for objective
- “Optimal” performances:

	implicit	LibFM	LightFM
Prec@3	0.10	0.087	0.082
Relative Speed	Medium	High	Low

Evaluations

Recommendations

For each persona, generate *N* <= *C* recommendations for all the prototypes under evaluation. Include two control recommenders: global popularity and random.

Analysis

Rank recommenders according to accuracy, diversity and overall business score. Check inter-annotator divergence.

Personas

Sample 10 personas with history of at least *C* clicks.

Evaluations

Each evaluator performs a blind evaluation for each recommendation set against accuracy and diversity criteria. Also assign overall “business” score to each recommender.

And the Winner is ...


AND THE WINNER IS...

Subjective Evaluation Results

Algorithm Label	Algorithm Type	Mean accuracy rank (lower=better)	Mean diversity rank (lower=better)	Mean overall score/10 (higher=better)	Relative Implementation Complexity	Computational Performance
A	Random	2.81	1.47	3.66	Trivial	Very Fast
B	LibFM	0.94	1.84	6.09	High	Slow
C	Implicit ALS	0.97	2.06	5.91	Low	Moderate
D	LightFM	1.41	1.75	3.66	Moderate	Fast
E	Global Popularity	3.88	2.88	2.28	Trivial	Fast

Parting Thoughts

- “Machine Learning that Matters [and works for us]” (K. Wagstaff, 2012)
- Hyper-parameter optimisation matters too
- But subjective evaluation is what *really* matters!
- Always consider marginal opportunity cost of algorithmic “complexity”

Machine Learning that Matters

"Machine learning offers a cornucopia of useful ways to approach problems that otherwise defy manual solution. However, much current ML research ***suffers from a growing detachment from those real problems***. Many investigators withdraw into their private studies with a copy of the data set and work in isolation to perfect algorithmic performance."

bibblio LABS

“Because I’d looked at the living roach and was discovering inside it the identity of my deepest life.”

@MNonmad @Bibblio_org #RecSysLDN