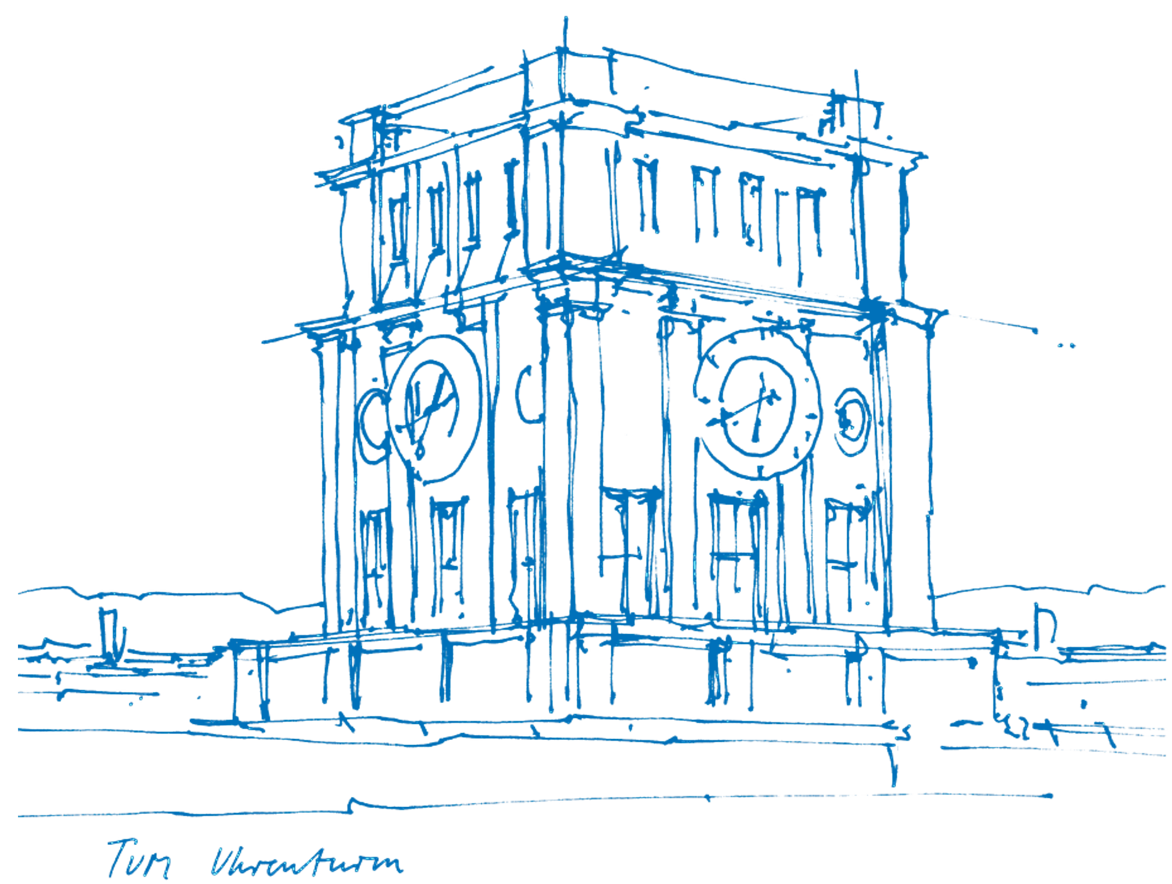
Recommending the Duration of Stay in Personalized Travel Recommender Systems

Abhishek Agarwal and **Linus W. Dietz**
Technical University of Munich



Recommend ~~where to go~~ how long to stay

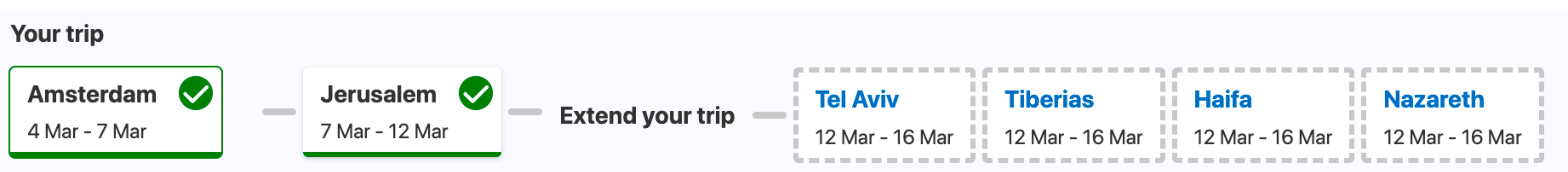


Image from https://booking.ai/booking-com-wsdm-webtour-2021-data-challenge-d814e9c1dd96

Previous work

How Long to Stay Where? On the Amount of Item Consumption in Travel Recommendation

Linus W. Dietz
Department of Informatics
Technical University of Munich
Garching, Germany
linus.dietz@tum.de

Wolfgang Wörndl
Department of Informatics
Technical University of Munich
Garching, Germany
woerndl@in.tum.de

[RecSys Late-breaking Results 2019]

Our approach

- **Two data sets:** Twitter trips & Booking.com
- **Four baselines:** User mean/mode, [Dietz & Wörndl 2019]
- **Two ML algorithms:** Gradient Boosting & Scikit DT
- **Three embeddings:** One-hot, Global, Personalized
- **Two metrics:** RMSE & MAE

Data sets

	Booking.com	Twitter
#Users	96,643	24,146
#Trips	734,102	852,131
#Origin Countries	5	97
#Destination Countries	107	105
Domestic trips	4.5%	91.3%
Date Range	Jan 2016 – Feb 2017	Oct 2010 – Jul 2021

[Goldenberg & Levin 2021] [Self collected]

Embeddings

	M-OHE	M-GE	M-PE
Trip type	Y	Y	Y
Traveler type clustering	Y	Y	Y
User home country	Y	N	N
Destination country	Y	N	N
City embeddings	N	Y	Y
Traveler embeddings	N	N	Y

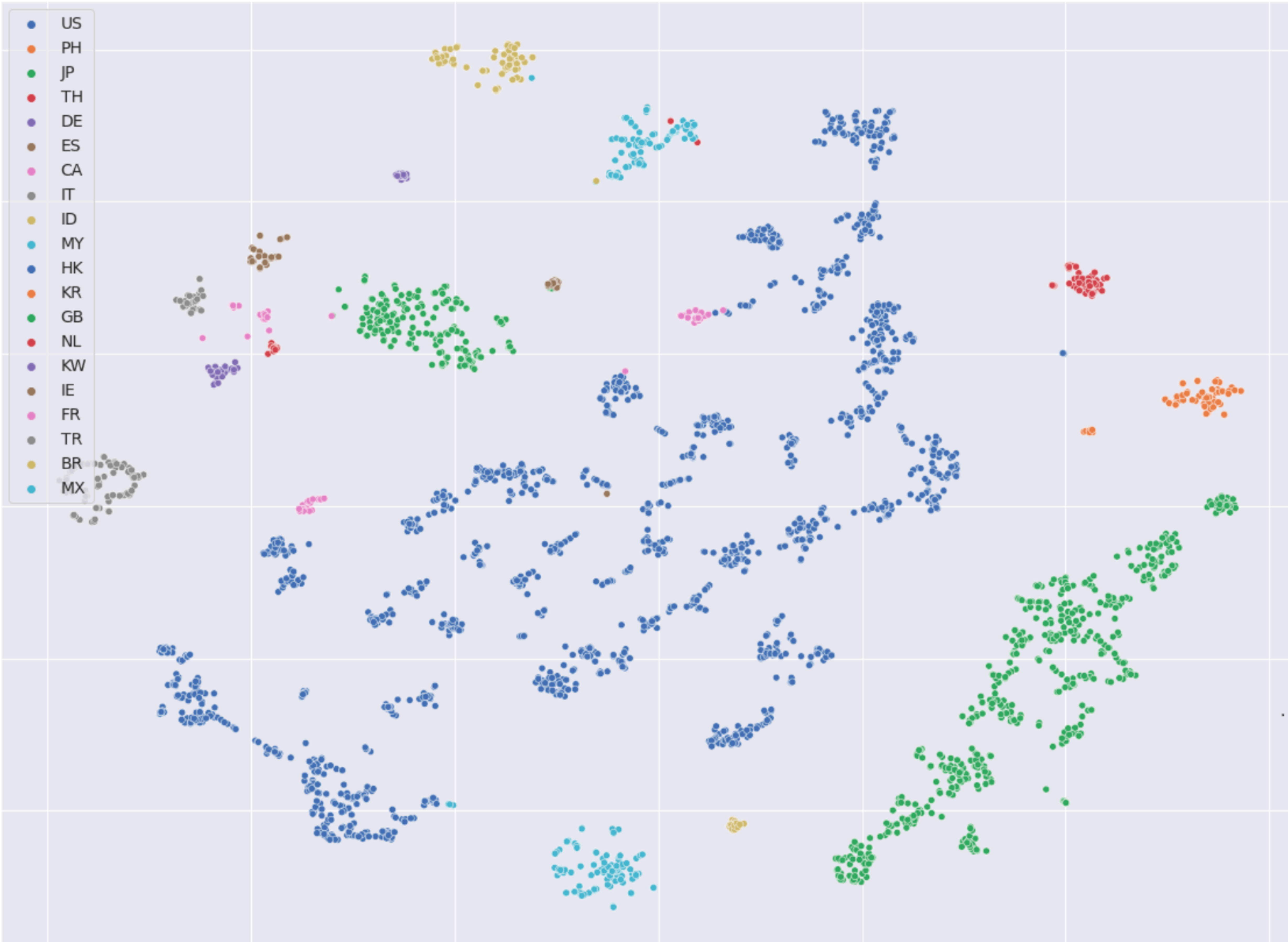
Embeddings: Traveler Clustering

- Identify different traveler types [Dietz et. al 2020]
- Features:
  - # domestic / international trips
  - # unique domestic / international cities visited
  - mean duration of domestic / international trips
- Determine clusters number: Silhouette width & SSE
- **Result:** 6 clusters in both data sets

Embeddings: Cities

- Construct mobility graph
- Use Node2Vec to determine city embeddings

	Booking.com	Twitter
#Nodes	5,046	3,523
#Edges	88,623	62,260
Density	0.0069	0.01


US
PH
JP
TH
DE
ES
CA
IT
ID
MY
HK
KR
GB
NL
KW
IE
FR
TR
BR
MX

Embeddings: Travelers

- Similar to city embeddings
- Per user: average embedding of all the past cities the user has visited

Embeddings

	M-OHE	M-GE	M-PE
Trip type	Y	Y	Y
Traveler type clustering	Y	Y	Y
User home country	Y	N	N
Destination country	Y	N	N
City embeddings	N	Y	Y
Traveler embeddings	N	N	Y

Results: Booking.com

Approach		MAE	MAE Rounded	RMSE	RMSE Rounded
User Mean		0.835	0.821	1.168	1.214
User Mode		0.742	0.742	1.233	1.233
User Percentile – Country		0.726	0.726	1.185	1.185
User Percentile – City		0.769	0.769	1.209	1.21
M-OHE	Scikit-DT	0.678	0.6	0.955	1.01
	CatBoost	0.678	0.601	0.954	1.01
M-GE	Scikit-DT	0.545	0.483	0.787	0.837
	CatBoost	0.541	0.475	0.777	0.827
M-PE	Scikit-DT	0.563	0.491	0.804	0.853
	CatBoost	**†0.534**	**†0.466**	**†0.767**	**†0.815**

Results: Twitter

Approach		MAE	MAE Rounded	RMSE	RMSE Rounded
User Mean		0.962	0.958	1.307	1.336
User Mode		**†0.806**	**†0.806**	1.49	1.49
User Percentile – Country		0.823	0.823	1.471	1.472
User Percentile – City		0.856	0.856	1.442	1.443
M-OHE	Scikit-DT	0.932	0.974	1.261	1.286
	CatBoost	0.931	0.975	1.259	1.283
M-GE	Scikit-DT	0.884	0.872	1.229	1.268
	CatBoost	0.882	0.876	1.222	1.262
M-PE	Scikit-DT	0.892	0.898	1.236	1.278
	CatBoost	0.844	0.820	**†1.183**	**†1.228**

Limitations

- Need richer data sets
- Accuracy with Deep/Neural learning?
- Not a sequential solution!

Conclusions

- Hard problem
- Mobility enables substantial feature engineering
- Recommending the user mode is not a huge mistake!