A novel approach for venue recommendation using cross-domain techniques

Pablo Sánchez, Alejandro Bellogín

pablo.sanchezp@uam.es, alejandro.bellogin@uam.es

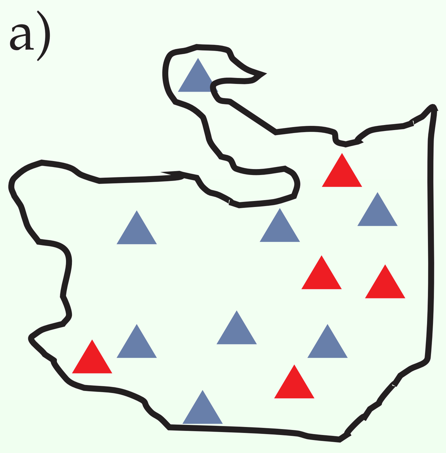
Information Retrieval Group, Department of Computer Science, Universidad Autónoma de Madrid

IRG
IR Group @UAM
UAM
Universidad Autónoma de Madrid

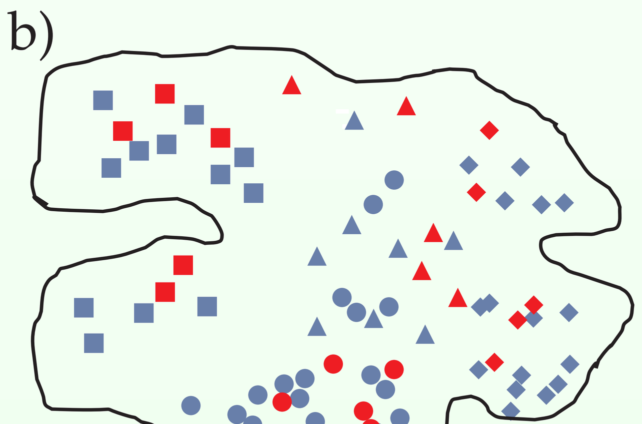
Venue Recommendation: Traditional Evaluation

- Two common approaches: consider each city as an independent dataset (a) or every check-in of many cities as one dataset (b) [1, 2].

a)



b)

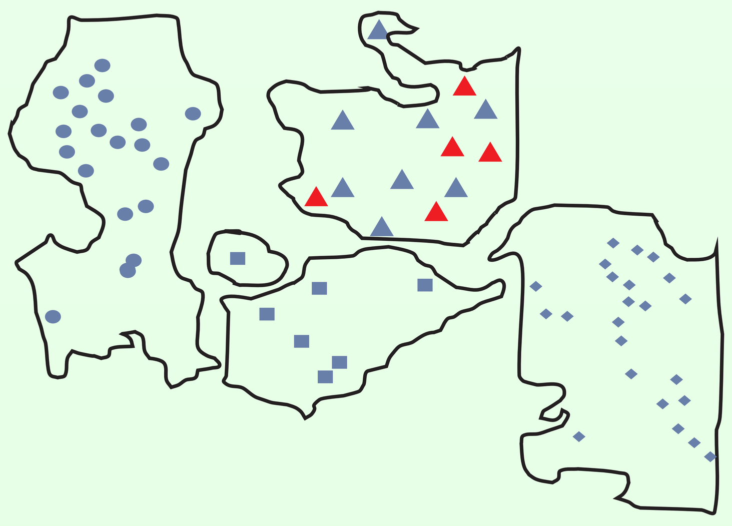


- Option a: it allows to isolate behavior on one city, but no external information can be exploited.
- Option b: by training once, many different cities can be evaluated, but no control about dominant cities is possible.

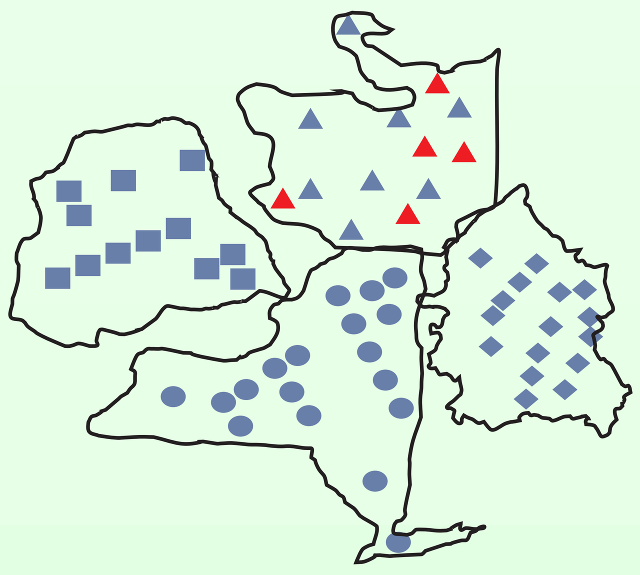
Venue Recommendation as Cross-Domain

- We propose to consider each city as an independent domain, using one target domain (test) and many source domains (training).
- Best options to learn and transfer knowledge? Our proposals: use most popular cities (more data) or closest cities (more overlap).

Nearest Cross-Domain (N-CD)



Most-Popular Cross-Domain (P-CD)



Experiments and Results

- Dataset: 33M Foursquare check-ins. Temporal split: 6 months for training, 1 month for test.
- Recommenders: closest venues (AvgDis), hybrid (PGN), UB, IB, HKV, MF with geographical information (IRenMF).
- Results using NDCG@5.

City		AvgDis	PGN	UB	IB	HKV	IRenMF
IST	N-CD	0.001	0.068	**0.073**	0.057	0.071	0.059
	Δ(%)	−9.7	1.6	0.3	−3.2	▲2.0	▼−14.8
	P-CD	0.001	0.068	**0.073**	0.059	0.068	0.052
	Δ(%)	−0.1	▲0.9	0.4	0.0	−3.4	▼−24.7
MEX	N-CD	0.001	0.044	0.045	0.013	**0.045**	0.040
	Δ(%)	▲13.3	2.2	1.6	−6.5	−5.0	▼−6.8
	P-CD	0.001	0.044	**0.045**	0.013	0.037	0.037
	Δ(%)	−0.2	▲1.3	1.2	−0.1	▼−22.1	−13.6
MOS	N-CD	0.002	0.033	0.038	0.017	**0.040**	0.034
	Δ(%)	▼−6.9	0.8	2.5	−0.7	▲3.3	−1.1
	P-CD	0.002	0.032	**0.037**	0.018	0.036	0.029
	Δ(%)	−0.6	0.1	0.3	▲1.1	−7.7	▼−17.4
SAO	N-CD	0.001	**0.057**	0.056	0.016	0.056	0.046
	Δ(%)	▼−7.1	0.4	▲15.4	5.5	15.2	7.3
	P-CD	0.001	**0.057**	0.049	0.015	0.047	0.034
	Δ(%)	−9.2	▲0.5	−0.2	−0.2	−2.1	▼−20.2
TOK	N-CD	0.000	**0.073**	0.073	0.048	0.064	0.071
	Δ(%)	▼−15.6	4.9	5.4	−0.2	▲8.7	4.2
	P-CD	0.001	**0.070**	0.069	0.048	0.064	0.064
	Δ(%)	−0.3	−0.2	−0.2	−0.1	▲8.6	▼−6.1

- Performance improvement for P-CD usually negligible.
- N-CD usually produces larger improvements with less data involved.
- UB and HKV exploit more successfully the information coming from source domains.
- Cross-domain techniques tend to deteriorate performance of techniques based on geographical distances.

Conclusions and Future Work

- Using Cross-Domain techniques in venue recommendation improves the performance of many recommenders.
- Selecting the cities by proximity is a good strategy to improve the results, confirming that **better data is more useful than more data**. "Everything is related to everything else, but near things are more related than distant things" [3].
- Future: explore different ways to select cities and exploit categorical information.

https://xkcd.com

References

[1] Yiding Liu, Tuan-Anh Pham, Gao Cong, Quan Yuan An Experimental Evaluation of Point-of-interest Recommendation in Location-based Social Networks. In PVLDB (2017), pp. 1010–1021.

[2] Yong Liu, Wei Wei, Aixin Sun, Chunyan Miao Exploiting Geographical Neighborhood Characteristics for Location Recommendation. In CIKM (2014), pp. 739–748.

[3] Harvey J. Miller Tobler's First Law and Spatial Analysis. In Annals of the Association of American Geographers (2004), pp. 284–289.

Source code available at:
`https://bitbucket.org/PabloSanchezP/TempCDSeqEval`

RecSys 2018, 12th ACM Conference on Recommender Systems. October 2-7, 2018. Vancouver, Canada.