Spatial Summarization of Image Collections

Diego A. Ballesteros Villamizar

ETH Zürich

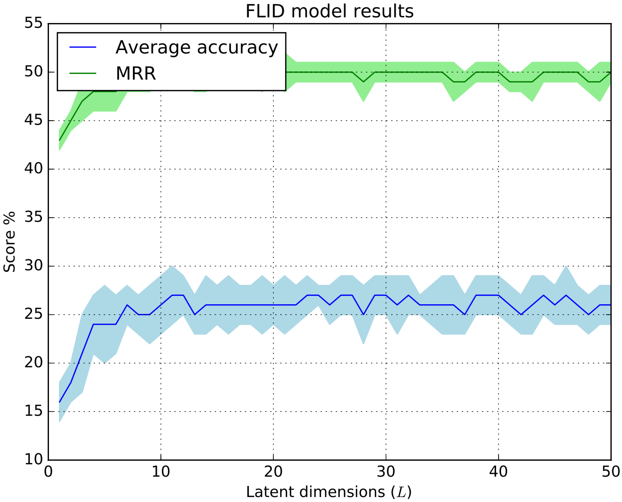
December 2nd, 2015

Outline

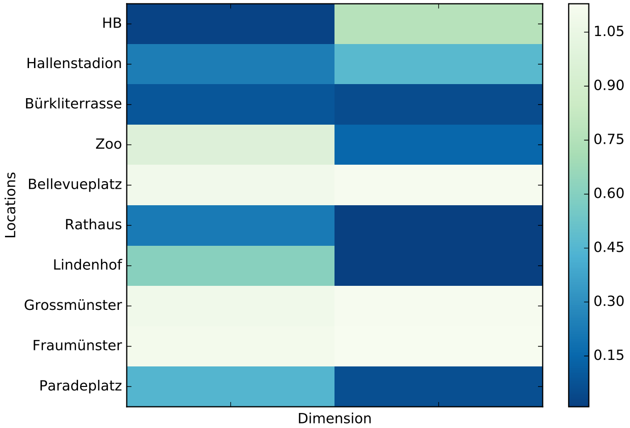
Recap

Model	Acc	σ_{Acc}	MRR	σ_{MRR}
Modular	16.42	2.42	44.47	1.40
FLID $(d = 2)$	18.95	2.60	45.35	1.64
FLID $(d = 5)$	24.50	4.14	48.73	2.85
FLID $(d = 10)$	26.32	3.35	50.00	1.92
FLID $(d = 20)$	26.75	2.18	50.20	1.53

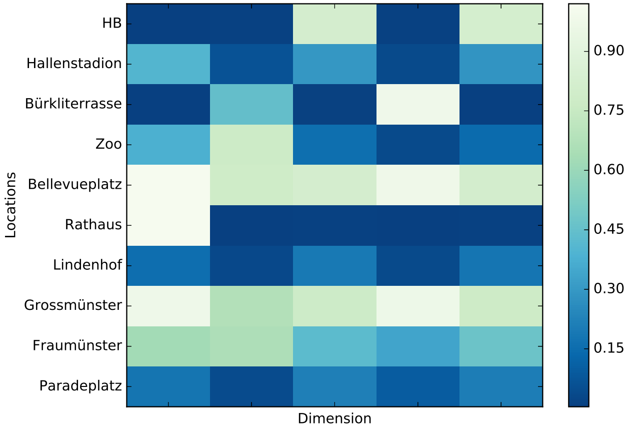
Choosing d


FLID model results
Average accuracy
MRR
Score %
Latent dimensions (L)
55
50
45
40
35
30
25
20
15
10
0
10
20
30
40
50

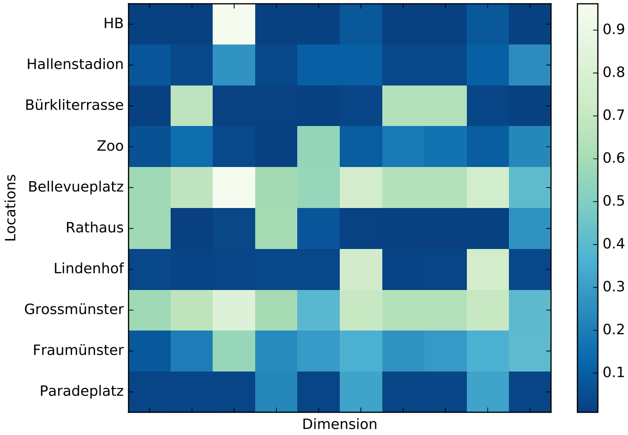
Diversity Encoding ($d = 2$)



Diversity Encoding ($d = 5$)



Diversity Encoding ($d = 10$)



Outline

Proximity Model

Assuming a partial ordered sequence S of length k for geo-located items, the model defines a score $Q(i \mid S)$.

$$\forall i \notin S : Q(i \mid S) = \frac{1}{d(i, s_k)} \tag{1}$$

Where $d(a, b)$ is the great-circle distance between items a and b, and s_k is the k_{th} element of the sequence.
The set of items to suggest for S is ordered descendingly according to $Q(i|S)$.

Markov Model

As proposed by [Kurashima et al., 2010]. The probability of visiting a location given a location history $S = \langle s_1 \dots s_k \rangle$ can be modeled as:

$$P(s_{k+1} = i \mid s_1 \dots s_k) = P(s_{k+1} = i \mid s_k) \tag{2}$$

Which can be estimated with maximum likelihood from the data as:

$$P(s_{k+1} = i \mid s_k) = \frac{N(l_{t+1} = i, l_t = s_k)}{N(l_t = s_k)} \tag{3}$$

Where $N(l_{t+1} = i, l_t = s_k)$ is the number of times that i was visited immediately after s_k, and $N(l_t = s_k)$ is the number of times that s_k was visited.

The set of items to suggest for S is ordered descendingly according to $P(i \mid S)$.

Results

The new baseline models were trained on the data and the same 10-fold evaluation was performed, the results are:

Model	Acc	σ_{Acc}	MRR	σ_{MRR}
Modular	16.42	2.42	44.47	1.40
FLID ($d = 10$)	26.32	3.35	50.00	1.92
Markov	30.18	2.59	53.41	1.76
Proximity	11.61	0.99	33.99	0.99

Outline

Definitions

- $\boldsymbol{V}$ The set of locations, with cardinality $N = |\boldsymbol{V}|$.
- M The number of features that define each location.
- L The number of latent concepts.
- $\boldsymbol{X}$ The feature matrix for the locations, $\boldsymbol{X} \in \mathbb{R}^{N \times M}$.
- $\boldsymbol{a}$ The vector that defines the utility weight for each feature, $\boldsymbol{a} \in \mathbb{R}^{M}$.
- $\boldsymbol{B}$ The matrix that defines the contribution of each feature to a latent diversity concept, $\boldsymbol{B} \in \mathbb{R}^{M \times L}$.

FLID with Features

The FLID model with features is defined by:

$$P(S \mid \boldsymbol{a}, \boldsymbol{B}) = \frac{1}{Z} \exp\left(\sum_{i \in S} \boldsymbol{x}_i \boldsymbol{a} + \sum_{l=1}^{L} \left(\max_{i \in S} \boldsymbol{x}_i \boldsymbol{b}_l - \sum_{i \in S} \boldsymbol{x}_i \boldsymbol{b}_l\right)\right) \quad (4)$$

Where $\boldsymbol{x}_i$ is the i-th row of $\boldsymbol{X}$, and $\boldsymbol{b}_l$ is the l-th column of $\boldsymbol{B}$. This is analog to the previous model if we define $\boldsymbol{u}$ and $\boldsymbol{W}$ as:

$$\boldsymbol{u} = \boldsymbol{X}\boldsymbol{a}$$
$$\boldsymbol{W} = \boldsymbol{X}\boldsymbol{B}$$

NCE learning

To modify the NCE learning algorithm, it is just necessary to change the definition of the sub-gradient as follows:

$$\left(\nabla_{\boldsymbol{a}} \log \frac{1}{\hat{Z}} \tilde{P}\left(S \mid \boldsymbol{a}, \boldsymbol{B}\right)\right)_m = \sum_{i \in S} x_{i,m} \tag{5}$$

$$\left(\nabla_{\boldsymbol{B}} \log \frac{1}{\hat{Z}} \tilde{P}\left(S \mid \boldsymbol{a}, \boldsymbol{B}\right)\right)_{m,l} = x_{i^*,m} - \sum_{i \in S} x_{i,m} \tag{6}$$

Where $x_{i^*,m}$ in equation 6 is:

$$i^* = \operatorname*{argmax}_{i \in S} \boldsymbol{x}_i \boldsymbol{b}_l$$

The subgradient for $\hat{Z}$ is unchanged.

References

Kurashima, T., Iwata, T., Irie, G., and Fujimura, K. (2010).
Travel route recommendation using geotags in photo sharing sites.
In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 579–588.