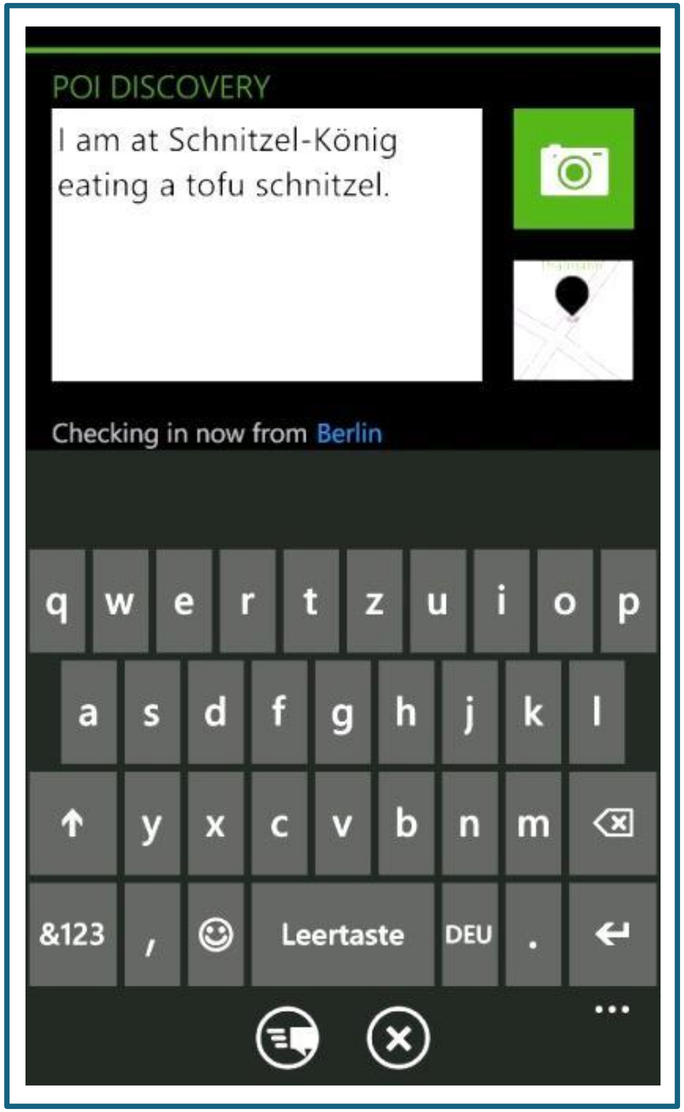
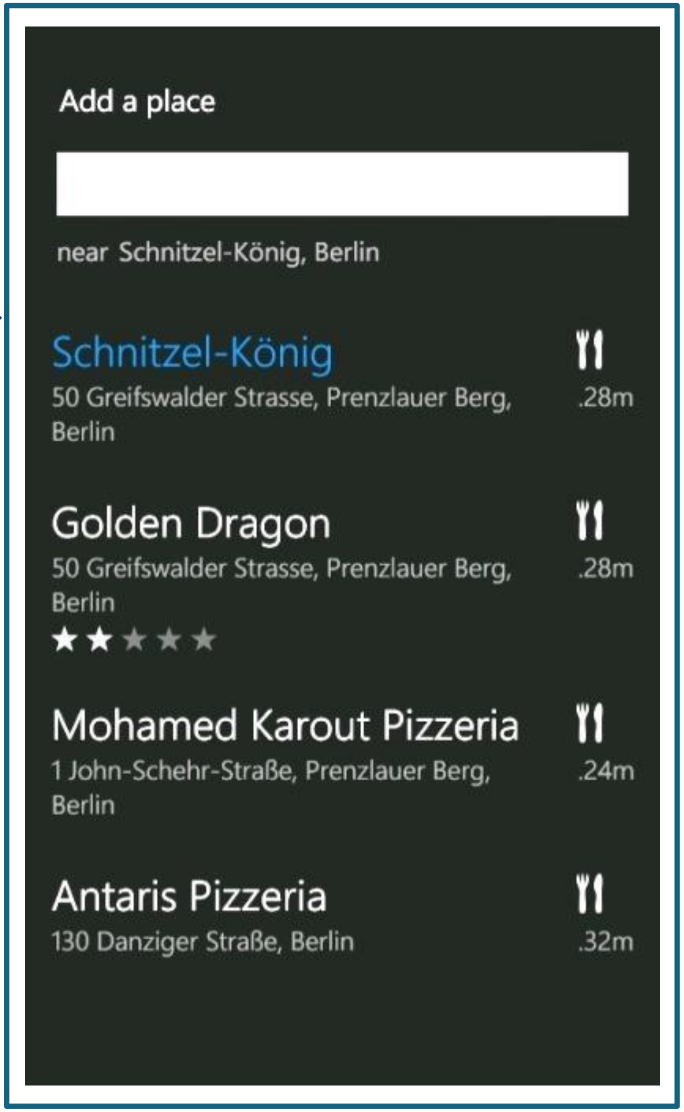
Relevance Optimization of Check-In Candidates

Steffen Bickel

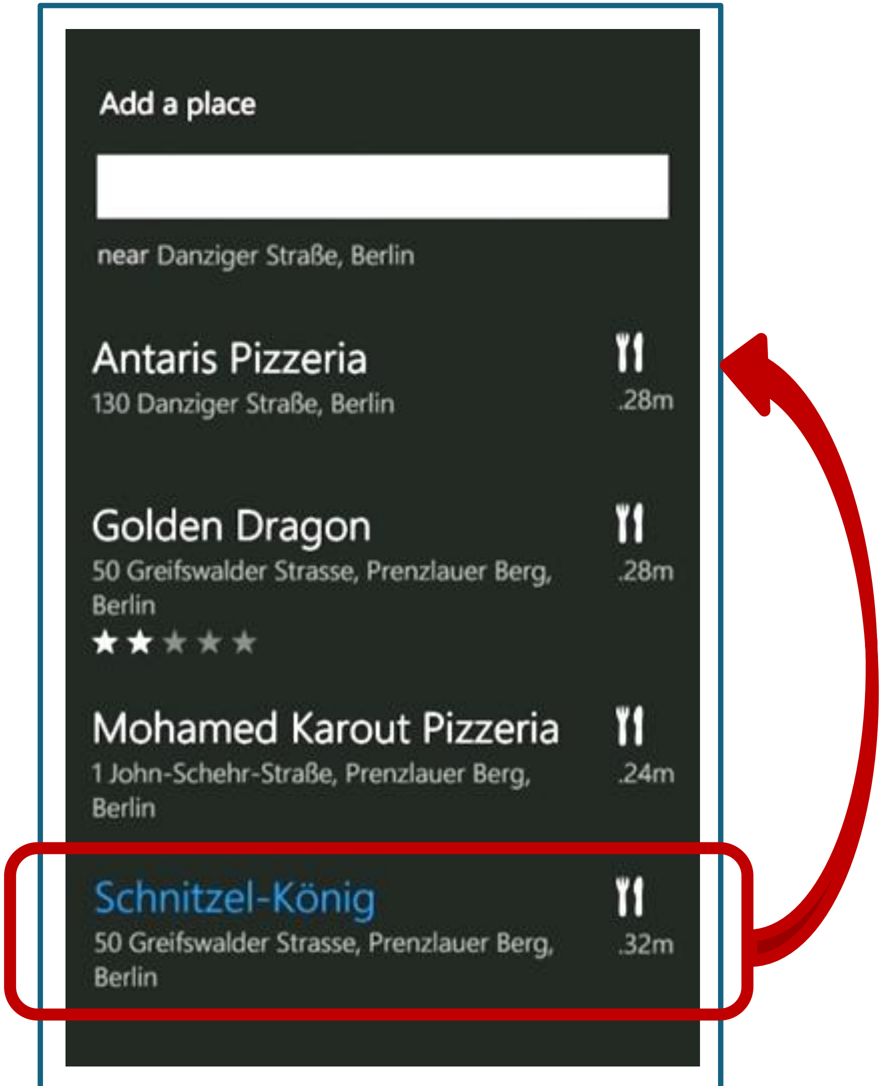
NOKIA

Check-In from Nokia Pulse App





Goal: Show Check-In Place at Rank 1



NOKIA

Goal: Show Check-In Place at Rank 1

Main Performance Measure: Precision@1

Precision@1 is percentage of cases where
correct check-in place is shown at rank 1 of the candidate list.

Goal: Show Check-In Place at Rank 1 or Within Top-5 Results

Also important for a good user experience:

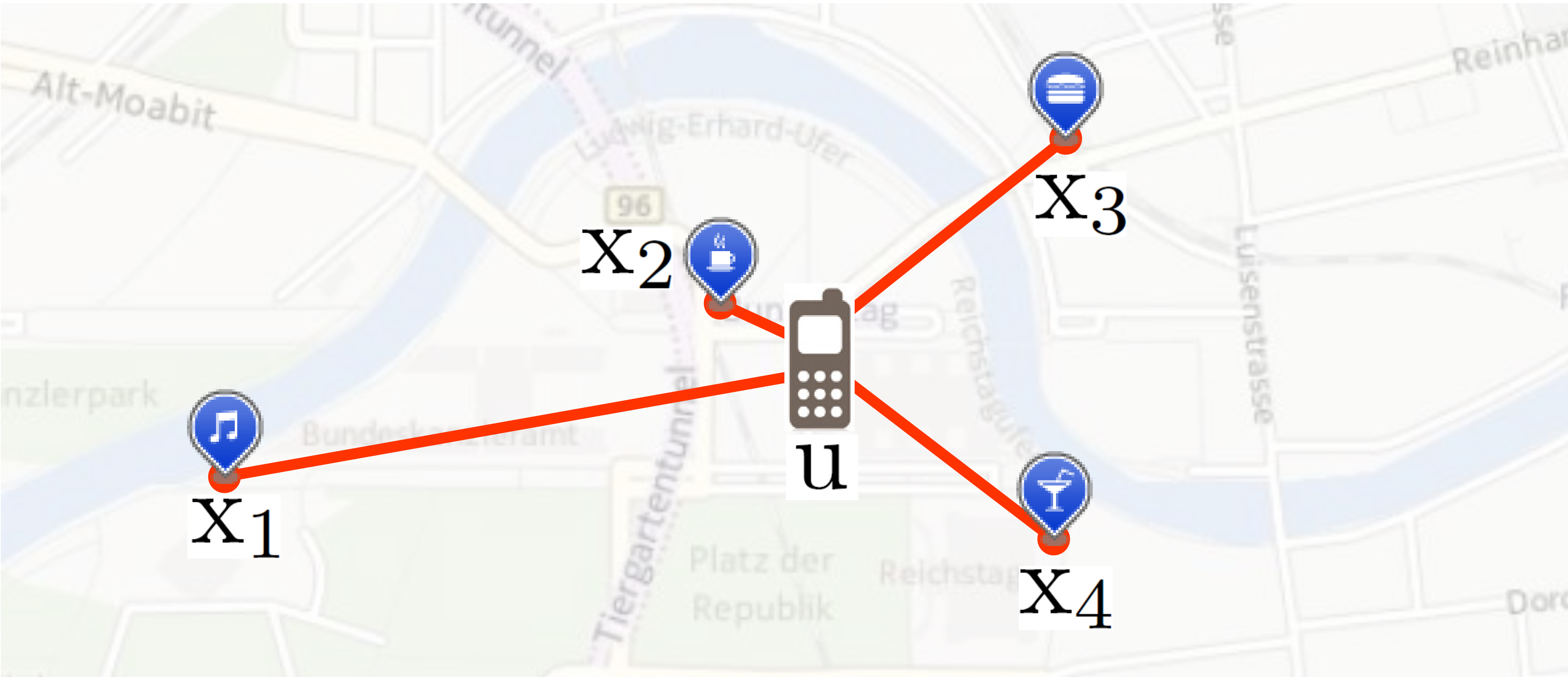
Precision@5

Precision@5 is percentage of cases where
correct check-in place is shown within top-5 results of candidate list.

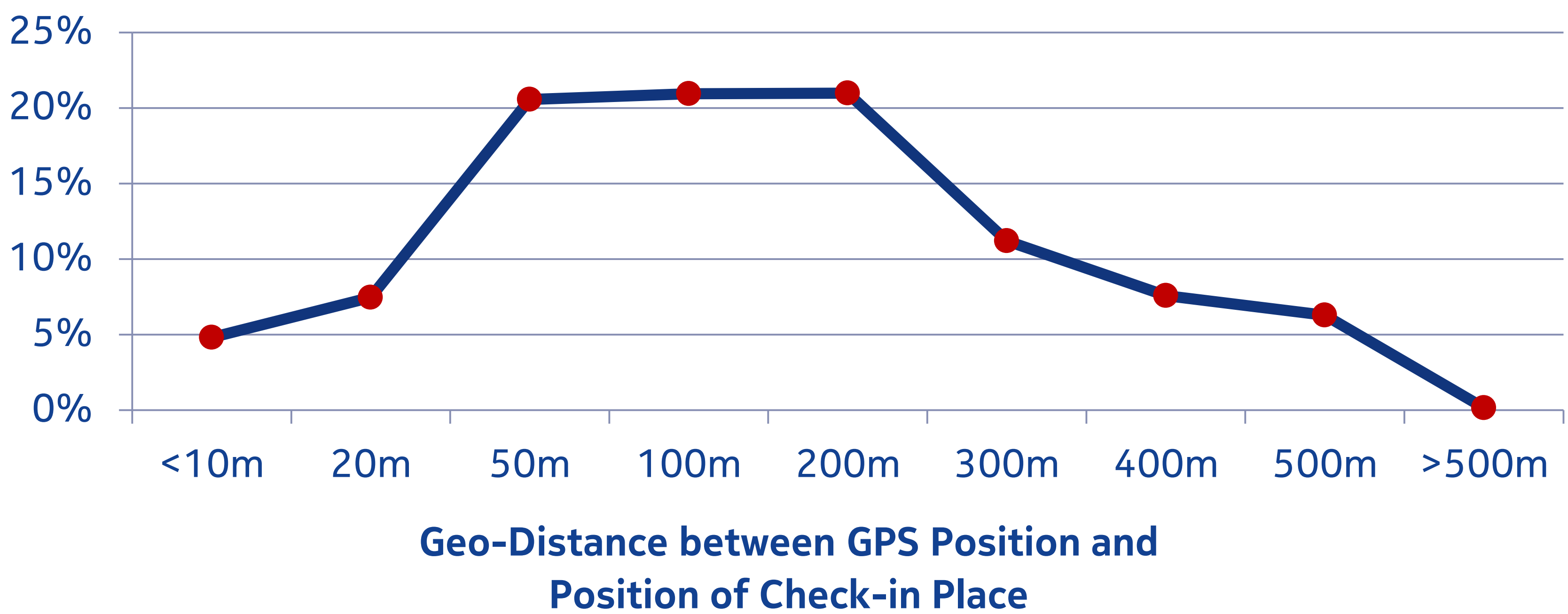
Relevance Modeling

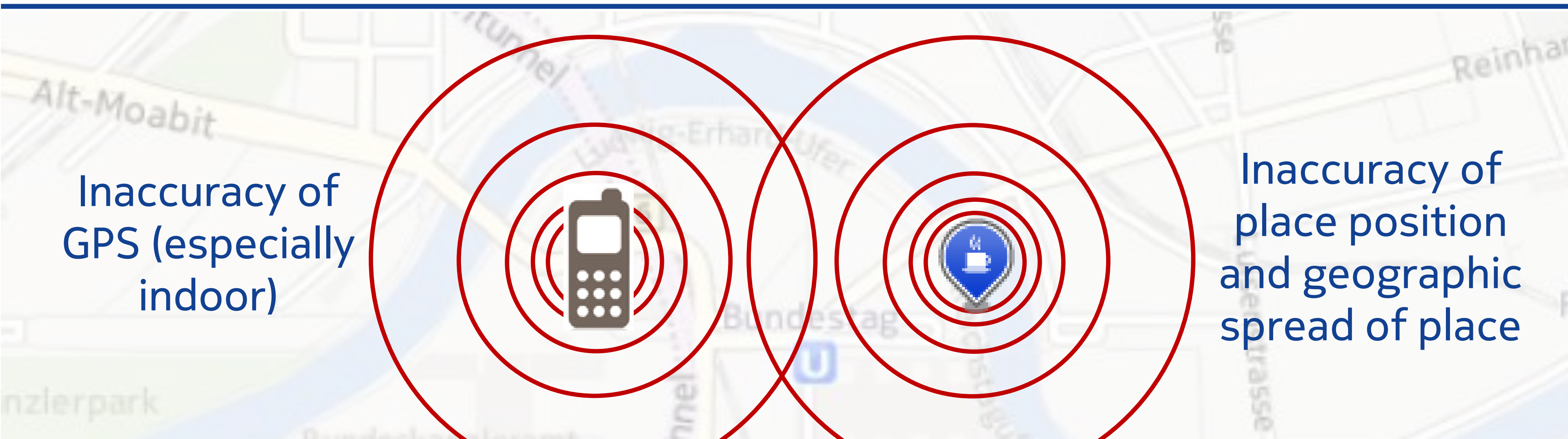
Relevance Factor: Distance = Baseline

$$\mathrm{score}(\mathrm{x}, \mathrm{u}) = -\mathrm{dist}(\mathrm{x}, \mathrm{u})$$



Inaccuracy of Geo-Positions





Relevance Factor: Popularity

www.flickr.com/photos/tomklaver/1499723565

Relevance Factor: Popularity

Places with many historic check-ins are more relevant

Closer places are more relevant

Relevance Factor: Popularity

$$\mathrm{score}(\mathrm{x}, \mathrm{u}) = (\alpha + \mathrm{n}_{\mathrm{x}})\ \mathrm{e}^{-\mathrm{dist}(\mathrm{x},\mathrm{u})/\sigma}$$

constant base count

number of historic check-ins on place x

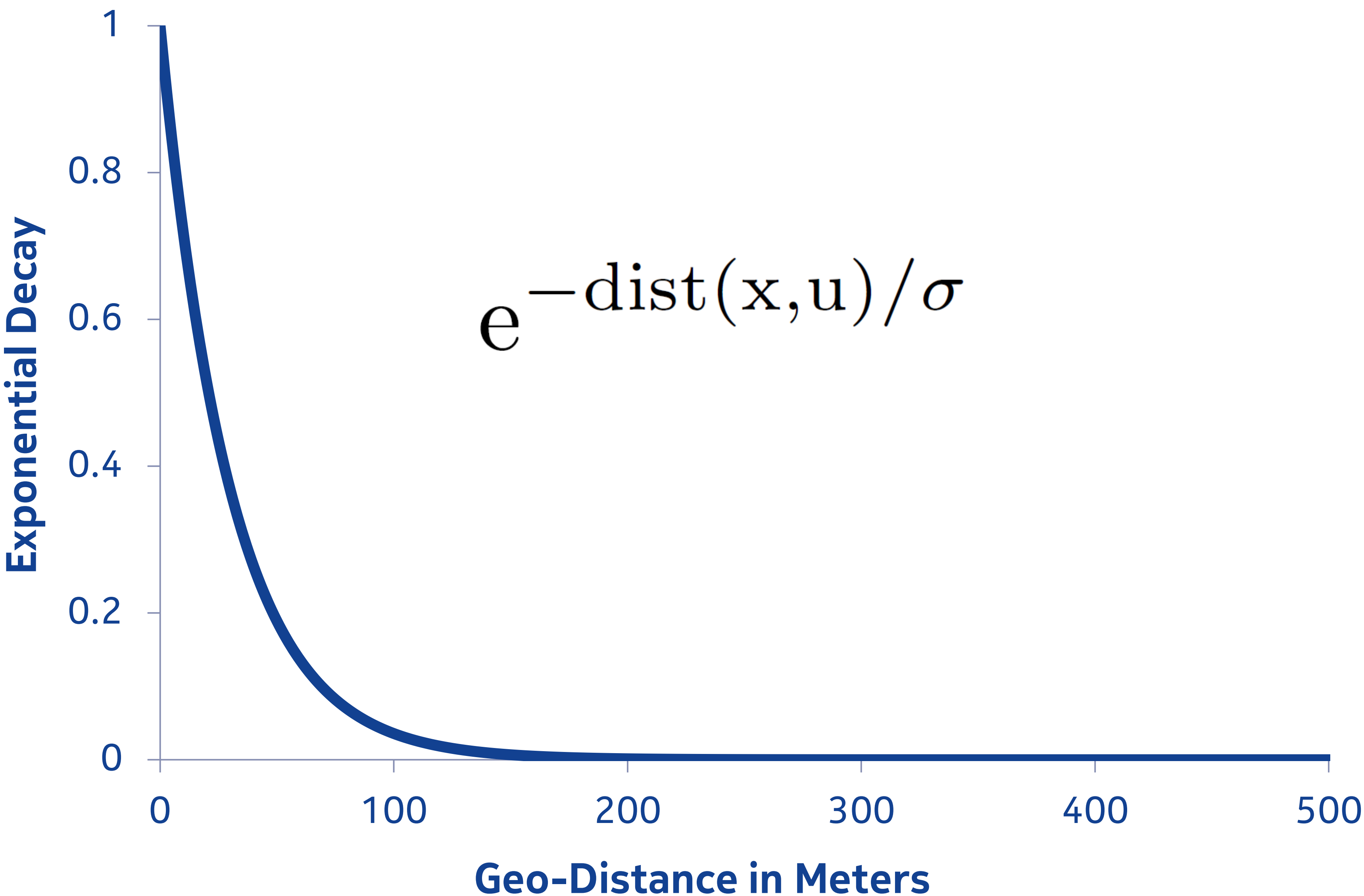
exponential decay on geo-distance to x

Places with many historic check-ins are more relevant

Closer places are more relevant

NOKIA

Exponential Decay on Geo-Distance



NOKIA

Relevance Factor: Popularity

$$\mathrm{score}(\mathrm{x},\mathrm{u}) = (\alpha + \mathrm{n_x})\ \mathrm{e}^{-\mathrm{dist}(\mathrm{x},\mathrm{u})/\sigma}$$

constant base count

number of historic check-ins on place x

exponential decay on geo-distance to x

Places with many historic check-ins are more relevant

Closer places are more relevant

Personalization

Personalization

$$\mathrm{score}(\mathrm{x}, \mathrm{u}) = (\alpha + \mathrm{n}_{\mathrm{x}} + \beta \mathrm{n}_{\mathrm{xu}})\ \mathrm{e}^{-\mathrm{dist}(\mathrm{x},\mathrm{u})/\sigma}$$

number of historic check-ins (all users)	number of personal historic check-ins	exponential decay on geo-distance to x
Places with many historic check-ins are more relevant	Revisits by current user are more relevant	Closer places are more relevant

Bootstrapping with Search Clicks

$$\mathrm{score}(\mathrm{x},\mathrm{u}) = (\alpha + \mathrm{n}_{\mathrm{x}} + \beta \mathrm{n}_{\mathrm{xu}} + \gamma \mathrm{c}_{\mathrm{x}})\ \mathrm{e}^{-\mathrm{dist}(\mathrm{x},\mathrm{u})/\sigma}$$

number of historic check-ins (all users)	number of personal historic check-ins	number of clicks on place x in search results	exponential decay on geo-distance to x
Places with many historic check-ins are more relevant	Revisits by current user are more relevant	Searched places are more relevant	Closer places are more relevant

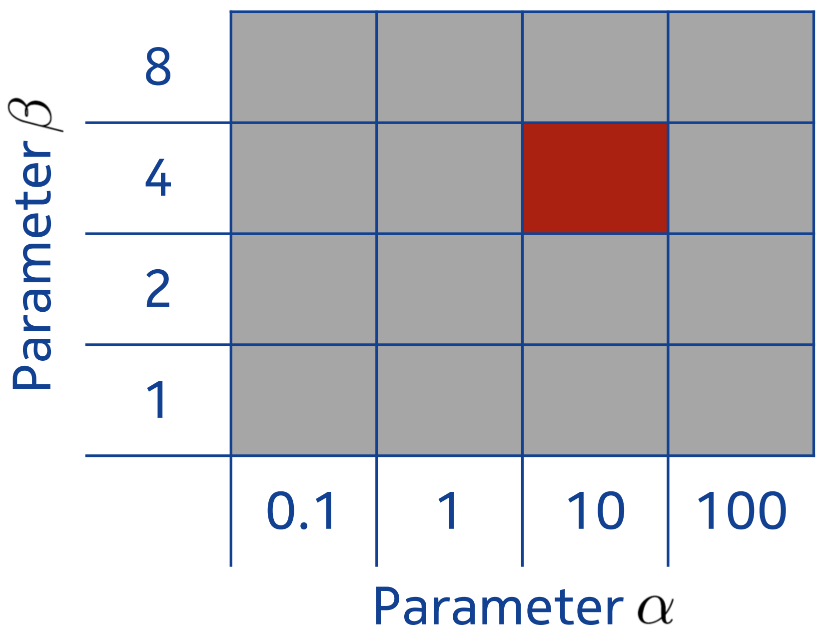
Parameter Learning
and
Evaluation Setup

4 Parameters Need to be Learned

$$\mathrm{score}(\mathrm{x}, \mathrm{u}) = (\alpha + \mathrm{n_x} + \beta \mathrm{n_{xu}} + \gamma \mathrm{c_x})\ \mathrm{e}^{-\mathrm{dist}(\mathrm{x},\mathrm{u})/\sigma}$$

Parameter Learning with Grid Search – Machine Learning for Slackers

All parameter combinations in grid are evaluated on tuning data.
Combination with best Precision@1 is chosen.



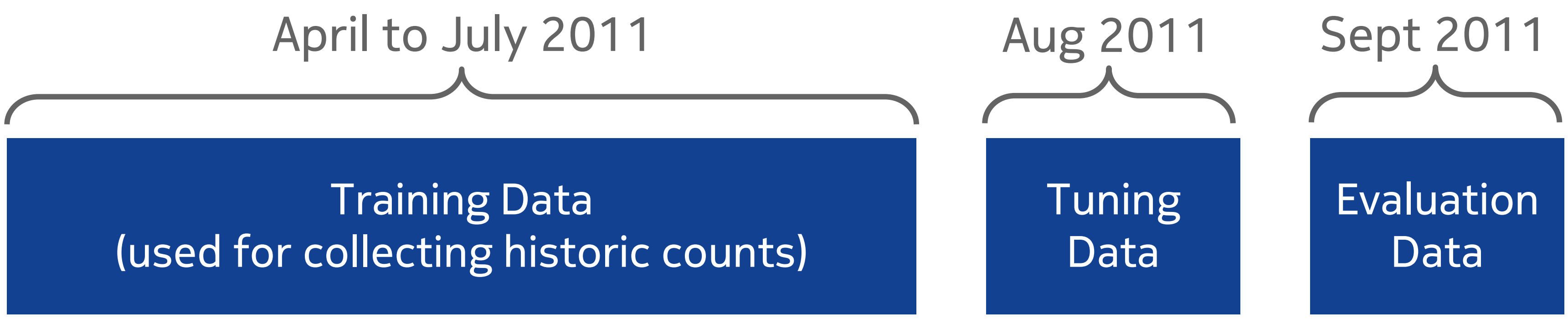
4 Parameters Need to be Learned

$$\text{score}(\text{x}, \text{u}) = (\alpha + \text{n}_\text{x} + \beta \text{n}_{\text{xu}} + \gamma \text{c}_\text{x})\ \text{e}^{-\text{dist}(\text{x},\text{u})/\sigma}$$

4 parameters → 4 dimensional search grid

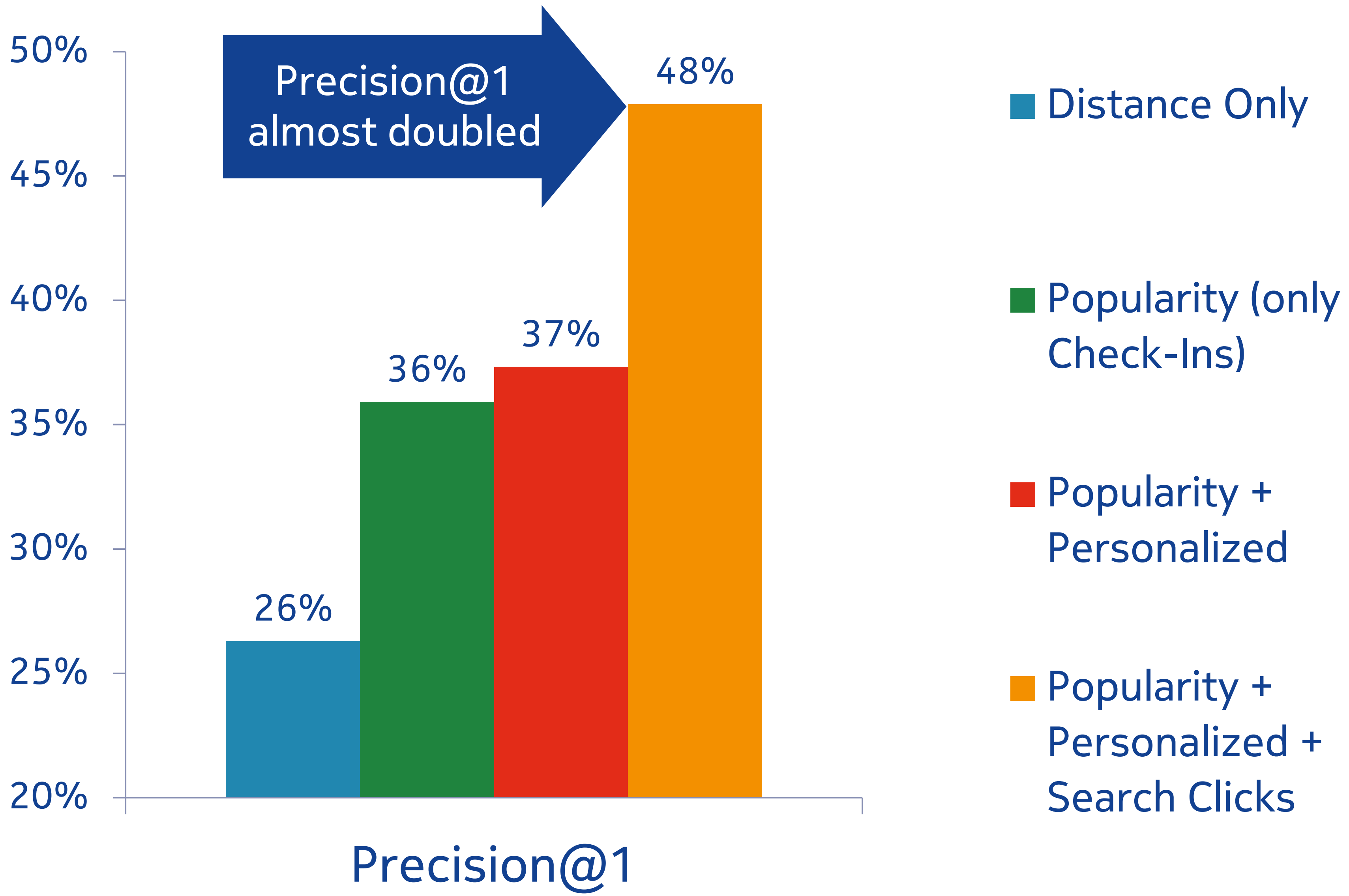
Evaluation Setup

Split of historic check-in data into 3 chunks.

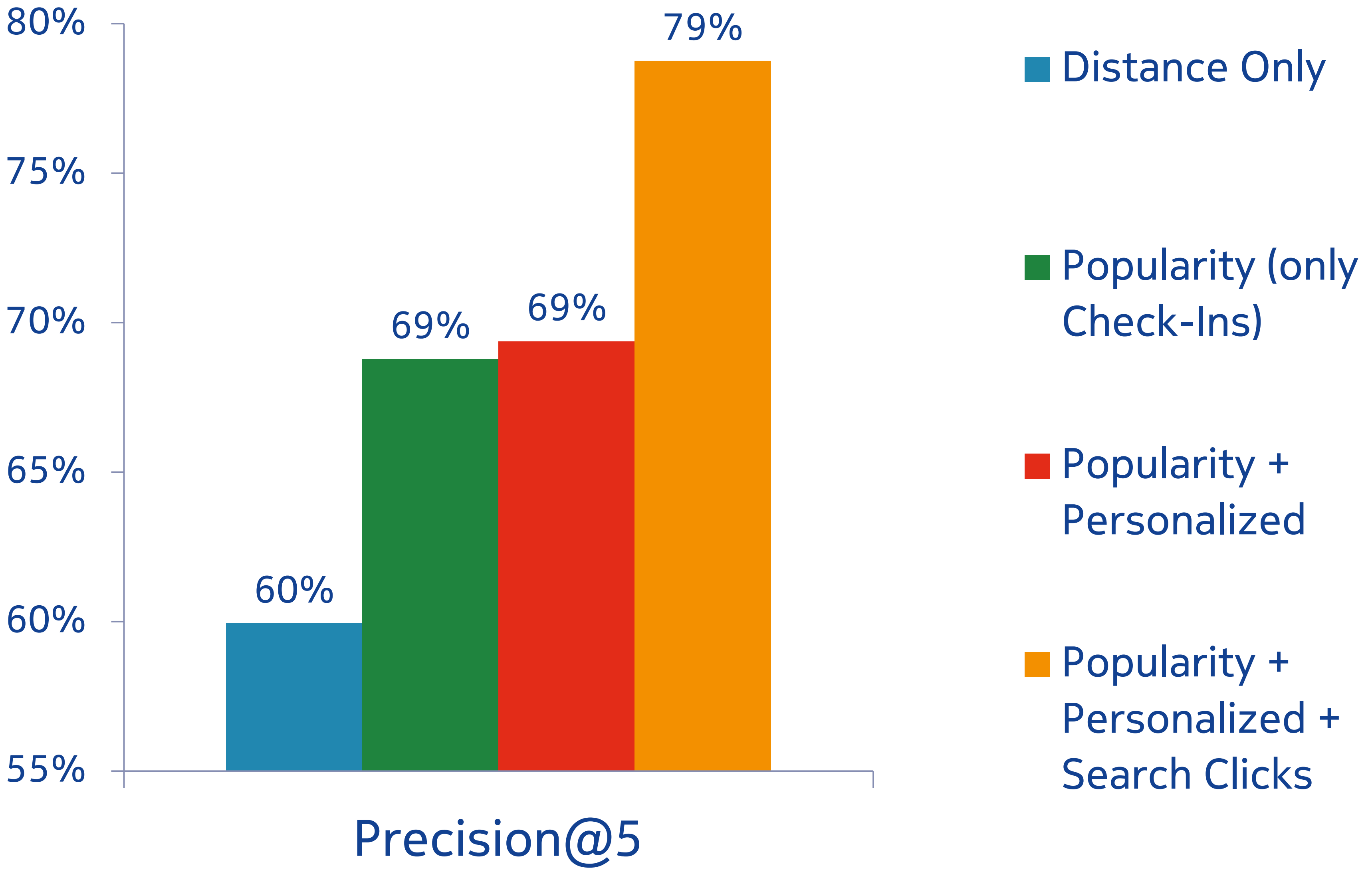


Evaluation Results

Evaluation of Relevance Model

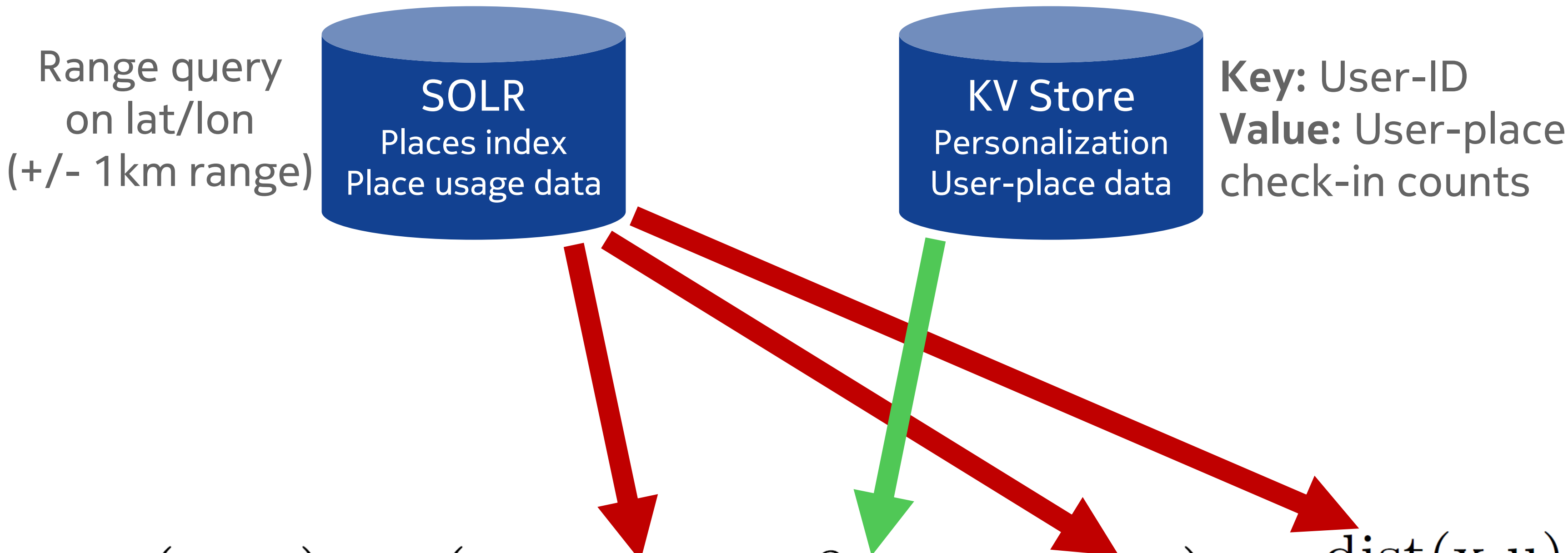


Evaluation of Relevance Model (Prec@5)



NOKIA

Implementation in Production: Sketch



$$\mathrm{score}(\mathrm{x},\mathrm{u}) = (\alpha + \mathrm{n_x} + \beta \mathrm{n_{xu}} + \gamma \mathrm{c_x})\ \mathrm{e}^{-\mathrm{dist}(\mathrm{x},\mathrm{u})/\sigma}$$

Sorting by score,
show top-k results to user

NOKIA

Summary

- Relevance model improves Precision@1 from 26% to 48%
- Usage data from different use-case (search clicks) helps.
- Grid search on model parameters is very simple but powerful way of doing machine learning.
- Statistical relevance model easy to implement in production on top of SOLR.

Thank You!

steffen.bickel@nokia.com

NOKIA