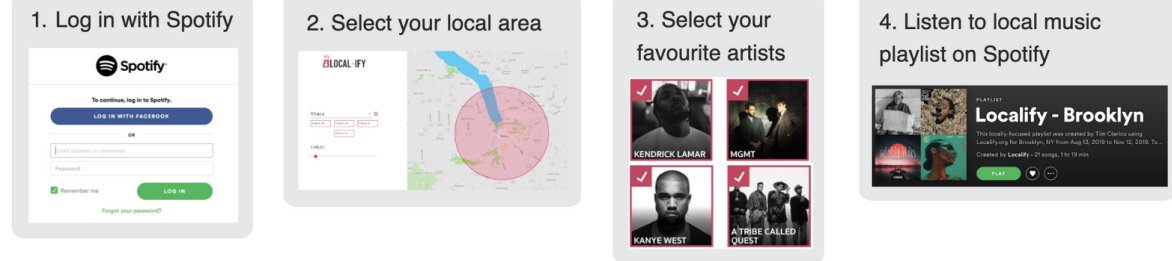# Can We Determine Artist Origin from Past Live Events?

Michael Zhou, Thorsten Joachims, Doug Turnbull

Cornell University · Ithaca College

Motivation: Localify

Localify is a music recommender system that promotes local music by artists from a city of a user's interest, to help users connect with artists and support their local communities.

1. Log in with Spotify
2. Select your local area
3. Select your favourite artists
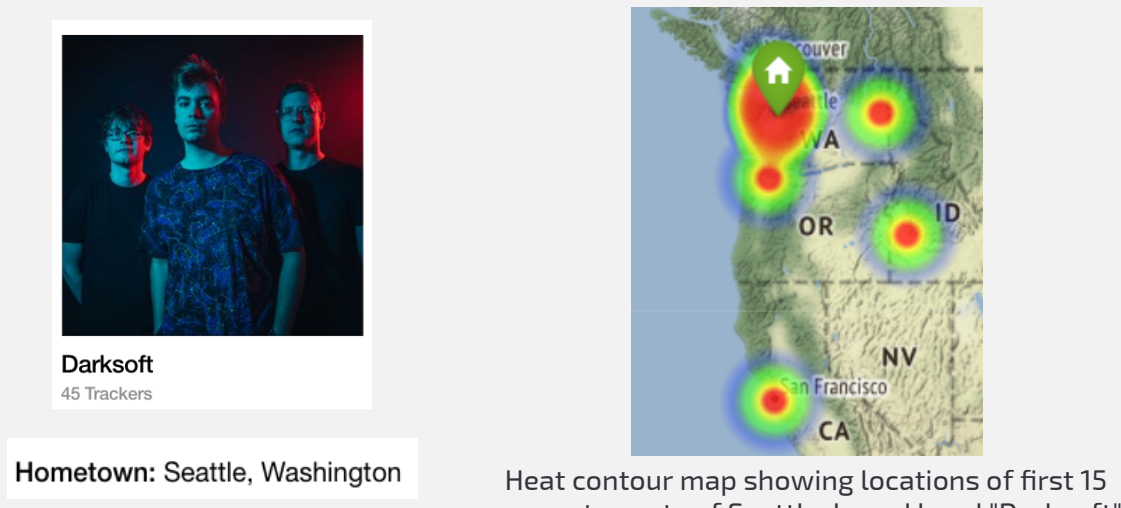4. Listen to local music playlist on Spotify

While artist origin info for popular artists may be available on biographical sources (e.g. Wikipedia), such information may not be as easy to find for long-tail, relatively obscure artists.

With that in mind, we explore a **novel** approach of heuristically approximating the artist origin via **past live events**.

Our Hypothesis

We hypothesize that when an artist first starts out, the artist will perform close to his/her origin place, before branching out of his/her origin place to tour as they gain more recognition.

Darksoft
45 Trackers

Hometown: Seattle, Washington

Heat contour map showing locations of first 15 concert events of Seattle-based band "Darksoft"

Experimental Setup

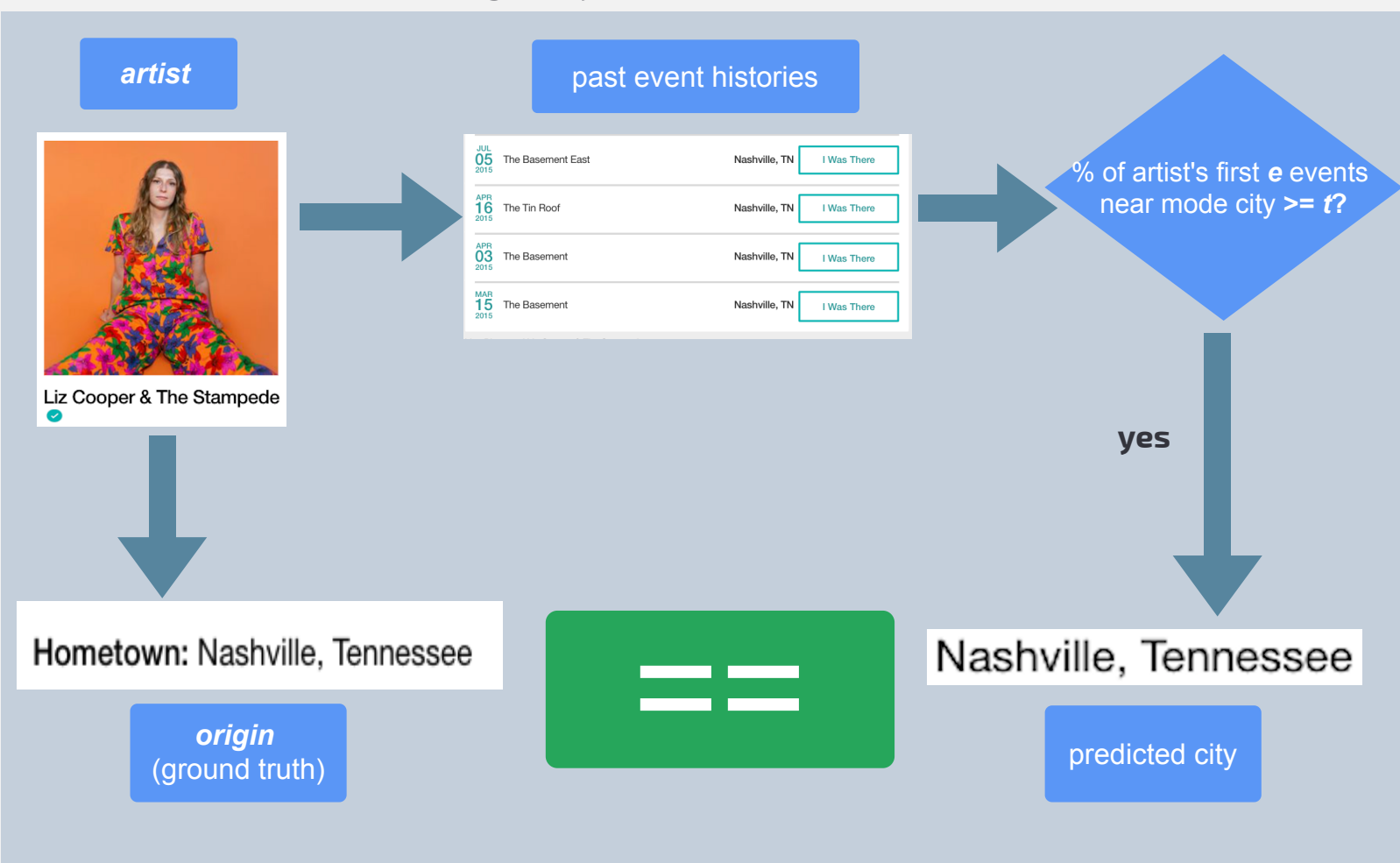Past event data from BandsInTown API (collected since 2013)

Sample S of 4766 artists spanning 249 cities sampled from local database:
- First show is recorded in 2014 or later
- From city with 10+ Wikipedia or BandsInTown-confirmed artists

Evaluation

For **e = 10, 20, 30, all**, and threshold **t = 0.8, 0.85, 0.9, 0.95**:
    For each (**artist, known origin city**) in **S**:

artist

past event histories

% of artist's first e events near mode city >= t?

yes

Liz Cooper & The Stampede

Hometown: Nashville, Tennessee

origin (ground truth)

==

Nashville, Tennessee

predicted city

gray block for each iteration

Results

Precision = # correctly predicted cities / # committed labels

Coverage = 100 x # of predicted cities / # of artists

Our "best" model consists of artists with the maximum coverage with at least 85% precision:
At least 80% of their **first 20 events** within 10 miles of their predicted origin city.

Set of Events e / Threshold t	0.8	0.85	0.9	0.95
all	0.840	0.842	0.837	0.810
first 30	0.867	0.864	0.86	0.833
first 20	**0.873**	0.884	0.881	0.867
first 10	0.793	0.810	0.810	0.831

Set of Events e / Threshold t	0.8	0.85	0.9	0.95
all	1.700	1.196	0.902	0.441
first 30	2.056	1.385	1.049	0.504
first 20	**2.476**	1.804	1.238	0.629
first 10	4.050	2.434	2.434	1.238

Precision table (above) and Coverage % table (below) for artists from cities with 10 or more confirmed artists. Each artist has 10 or more events and their first event recorded in BIT came after January 1st, 2014. The precision and coverage for the"best" model are **bolded**.

The errors are mainly due to **relocated artists** from predicted city, and/or **earliest events not being recorded in BandsInTown**.

Conclusion

1. Many artists with correctly predicted origin cities stay close to their origin.
2. Initial attempts with supervised learning did not produce better results
3. BandsInTown may contain **incomplete** historical event data - **early shows** may not be recorded

Future Work

1. Using Localify to **Crowdsource** Artist Origin
2. Further Exploration of Supervised Machine Learning Methods
3. **Combining** biographical (text) sources with past event data
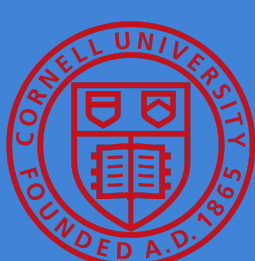
CONTACT US

Michael Zhou
mgz27@cornell.edu

Doug Turnbull
dturnbull@ithaca.edu

Thorsten Joachims
tj@cs.cornell.edu