

Where should I travel next? Modeling multi-destination trips with Recurrent Neural Networks.

Pavel Levin
Sr. Data Scientist - **Booking.com**

- Who am I?
 - **Data Scientist at Booking.com for about 4 years**
 - Studied Applied Math (McGill University, University of Ottawa)
 - Prior to Booking: medical research (psychiatry, pain research, drug resistance modeling, cancer trials)
- About Booking.com
 - Founded in 1996; part of Booking Holdings (NASDAQ:BKNG) from 2005
 - Over 29 million listings including everything from apartments and holiday homes to hotels and resorts, and even igloos and tree houses with 164,000,000+ guest reviews
 - **140,000+ destinations** in 230 countries and territories
 - **1,550,000+ rooms booked every 24 hours**
 - 17,000+ employees in over 200 offices in 70 countries

Recommending destinations

Recommending destinations

Find Deals for Any Season
From cozy country homes to funky city apartments

Where are you going? Check-in — Check-out 2 adults · 0 children Search
 I'm traveling for work

Going on vacation soon? [Save 15% or more with Mid-Year Deals](#)



New York City  1,608 properties
Average price US\$293



Tel Aviv  1,438 properties
Average price US\$201



London  11,090 properties
Average price US\$229



Jerusalem  549 properties
Average price US\$224



Rome  12,453 properties
Average price US\$248

Recommending destinations

Find Deals for Any Season
From cozy country homes to funky city apartments

Where are you going? Check-in — Check-out 2 adults · 0 children Search
 I'm traveling for work

Going on vacation soon? [Save 15% or more with Mid-Year Deals](#)

New York City 1,608 properties Average price US\$293

Tel Aviv 1,438 properties Average price US\$201

London 11,090 properties Average price US\$248

Jerusalem 549 properties

Rome 12,453 properties Average price US\$248

Booking.com

Search Hotels, guest houses, motels, and more...
e.g. city, landmark, address

Utrecht 71 properties available

Amsterdam 1297 properties available

Rotterdam 115 properties available

The Hague 120 properties available

Brussels 629 properties available

Paris 4298 properties available

Booking.com

Recommending destinations

Find Deals for Any Season
From cozy country homes to funky city apartments

Where are you going? Check-in — Check-out 2 adults · 0 children Search
 I'm traveling for work

Going on vacation soon? [Save 15% or more with Mid-Year Deals](#)



New York City 1,608 properties
Average price US\$293



Tel Aviv 1,438 properties
Average price US\$201



London 11,090 properties



Jerusalem 549 properties



Rome 12,453 properties

Booking.com

Gozo 611 properties	Tenerife 7,283 properties	Ibiza 1,301 properties	Isle of Wight 826 properties	Lake District 3,208 properties
New Forest 216 properties	Al Madinah Al Munnawarah ... 376 properties	Cotswolds 1,070 properties	Makkah Al Mukarramah Pro... 1,448 properties	Scotland 12,779 properties

Discover

You'll love relaxation, shop... United States of America

You'll love food, scenery a... Italy

Put tranquillity, scenery an... France

If tranquillity, food and be... Spain

Russia

1 2 3 4 5 6 7 8 9 10 More countries

List Your Property

Brussels 629 properties available

Paris 4,290 properties available

Booking.com

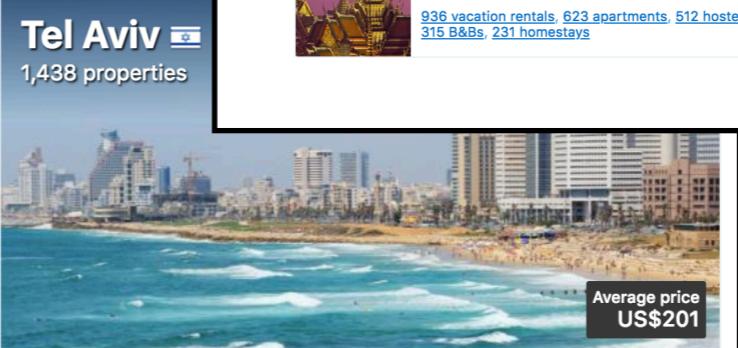
Recommending destinations

Find Deals for Any Season
From cozy country homes to funky city apartments

Where are you going? Check-in — Check-out
 I'm traveling for work

Going on vacation soon? [Save 15% or more with Mid-Year Deals](#)

New York City  1,608 properties
Average price US\$293

Tel Aviv  1,438 properties
Average price US\$201

London  11,090 properties

Jerusalem  549 properties

Rome  12,453 properties

Gozo 611 properties
New Forest 216 properties

Tenerife 7,283 properties
Al Madinah Al Munnawarah ... 376 properties

Ibiza 1,301 properties
Cotswolds 1,070 properties

Isle of Wight 826 properties
Makkah Al Mukarramah Pro... 1,448 properties

Lake District 3,208 properties
Scotland 12,779 properties

Discover

You'll love relaxation, shop... 
United States of America

You'll love food, scenery a... 
Italy

Put tranquillity, scenery an... 
France

If tranquillity, food and be... 
Spain

Russia is highly rated by tr... 
Russia

[List Your Property](#)

Apartments 664,987 apartments
Resorts 21,308 resorts
Villas 336,384 villas
Cabins 10,891 cabins
Cottages 104,093 cottages
Glamping 6,983 Glamping

Berlin  Germany
631 vacation rentals, 484 apartments, 107 homestays, 101 B&Bs, 91 guest houses

Amsterdam  Netherlands
1059 vacation rentals, 496 B&Bs, 451 apartments, 118 homestays, 53 guest houses

Bangkok  Thailand
936 vacation rentals, 623 apartments, 512 hostels, 315 B&Bs, 231 homestays

Las Vegas  United States of America
258 vacation rentals, 143 apartments, 95 vacation homes, 95 villas, 78 serviced apartments

Barcelona  Spain
2287 vacation rentals, 1922 apartments, 280 B&Bs, 270 homestays, 229 guest houses

[More destinations](#)

Booking.com

Booking.com

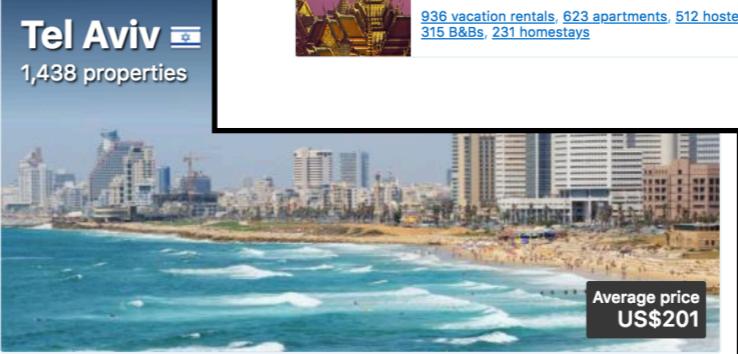
Recommending destinations

Find Deals for Any Season
From cozy country homes to funky city apartments

Where are you going? Check-in — Check-out
 I'm traveling for work

Going on vacation soon? [Save 15% or more with Mid-Year Deals](#)

New York City  1,608 properties
Average price US\$293

Tel Aviv  1,438 properties
Average price US\$201

Toilet With Grab Rails 9 [Show all 11](#)

Have you thought about staying further away?

Bat Yam
7.3 km from Tel Aviv
17 properties available *18% cheaper 

Herzlia
11.9 km from Tel Aviv
20 properties available *43% more 

Netanya
29.1 km from Tel Aviv
15 properties available *7% cheaper 

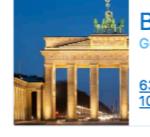
Ramat Gan
4.7 km from Tel Aviv
3 properties available

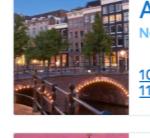
Petah Tiqwa
9.3 km from Tel Aviv
7 properties available *16% cheaper 

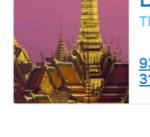
Ashdod
33.2 km from Tel Aviv
2 properties available *132% more 

Ben Gurion Airport
12.5 km from Tel Aviv
33 properties available *19% cheaper 

Apartments 664,987 apartments **Resorts** 21,308 resorts **Villas** 336,384 villas **Cabins** 10,891 cabins **Cottages** 104,093 cottages **Glamping** 6,983 Glamping

Berlin  Germany 631 vacation rentals, 484 apartments, 107 homestays, 101 B&Bs, 91 guest houses

Amsterdam  Netherlands 1059 vacation rentals, 496 B&Bs, 451 apartments, 118 homestays, 53 guest houses

Bangkok  Thailand 936 vacation rentals, 623 apartments, 512 hostels, 315 B&Bs, 231 homestays

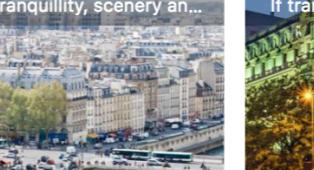
Las Vegas  United States of America 258 vacation rentals, 143 apartments, 95 vacation homes, 95 villas, 78 serviced apartments

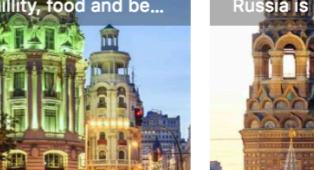
Barcelona  Spain 2287 vacation rentals, 1922 apartments, 280 B&Bs, 270 homestays, 229 guest houses

Discover

You'll love relaxation, shop...  United States of America

You'll love food, scenery a...  Italy

Put tranquillity, scenery an...  France

If tranquillity, food and be...  Spain

Russia is highly rated by tr...  Russia

[List Your Property](#)

Brussels
629 properties available

Paris
4298 properties available

Booking.com

Recommending destinations

Fitness Center 33

Not what you're looking for? Try your search again.

< 1 2 3 4 5 6 7 ... 19 20 >

Showing 1 – 50

Recommended Destinations for You:

Jerusalem 89 properties available
Compared to the past 2 days, prices in Jerusalem for your dates are now lower!

Haifa 41 properties available
Prices in Haifa have gone up for your dates since yesterday.

Eilat 9 properties available
Prices in Eilat have gone up for your dates since yesterday.

Tiberias 23 properties available
Prices in Tiberias for your dates are the lowest we've seen in 40 days!

Herzlia 20 properties available
Prices in Herzlia have gone up for your dates since yesterday.

Netanya 15 properties available
Prices in Netanya have gone up for your dates since yesterday.

Toilet With Grab Rails 9

Show all 11

Have you thought about staying further away?

Bat Yam

7.3 km from Tel Aviv
17 properties available

*18% cheaper

Herzlia

11.9 km from Tel Aviv
20 properties available

*43% more

Netanya

29.1 km from Tel Aviv
15 properties available

*7% cheaper

Ramat Gan

4.7 km from Tel Aviv
3 properties available

Petah Tiqwa

9.3 km from Tel Aviv
7 properties available

*16% cheaper

Ashdod

33.2 km from Tel Aviv
2 properties available

*132% more

Ben Gurion Airport

12.5 km from Tel Aviv
33 properties available

*19% cheaper

Check-in — Check-out

5% or more with Mid-Year Deals

Average price US\$293

Tel Aviv 1,438 properties

Average price US\$201

Rome 12,453 properties

1 2 3 4 5 6 7 8 9 10 More countries

Discover

You'll love relaxation, shop...

United States of America

You'll love food, scenery a...

Italy

Put tranquillity, scenery an...

France

If tranquillity, food and be...

Spain

Russia is highly rated by tr...

Russia

List Your Property

Brussels 629 properties available

Paris 4,230 properties available

Booking.com

Recommending destinations

Fitness Center 33 Not what you're looking for? Try your search again.

Showing 1 – 50

Recommended Destinations for You:

- Jerusalem** 89 properties available

Compared to the past 2 days, prices in Jerusalem for your dates are now lower!
- Haifa** 41 properties available

Prices in Haifa have gone up for your dates since yesterday.
- Eilat** 9 properties available

Prices in Eilat have gone up for your dates since yesterday.
- Tiberias** 23 properties available

Prices in Tiberias for your dates are the lowest we've seen in 40 days!
- Herzlia** 20 properties available

Prices in Herzlia have gone up for your dates since yesterday.
- Netanya** 15 properties available

Prices in Netanya have gone up for your dates since yesterday.

Toilet With Grab Rails 9 Show all 11

Have you thought about staying further away?

- Bat Yam** 7.3 km from Tel Aviv 17 properties available *18% cheaper
- Herzlia** 11.9 km from Tel Aviv 20 properties available *43% more
- Netanya** 29.1 km from Tel Aviv 15 properties available *7% cheaper
- Ramat Gan** 4.7 km from Tel Aviv 3 properties available
- Petah Tiqwa** 9.3 km from Tel Aviv 7 properties available *16% cheaper
- Ashdod** 33.2 km from Tel Aviv 2 properties available *132% more
- Ben Gurion Airport** 12.5 km from Tel Aviv 33 properties available *19% cheaper

Check-in — Check-out

5% or more with Mid-Year Deals

Tel Aviv 1,438 p

Average price US\$293

London 11,090 properties

Jerusalem 549 properties

Booking.com

Gozo 611 properties
New Forest 216 properties

Tenerife 7,283 prop
Al Madinal 376 proper

Discover

You'll love relaxation, shop... United States of America

You'll lov Italy

List Your Property

Brussels 629 properties available

Paris 4,298 properties available

Apartments 664,987 apartments **Resorts** 21,308 resorts **Villas** 336,384 villas **Cabins** 10,891 cabins **Cottages** 104,093 cottages **Glamping** 6,983 Glamping

Berlin Germany 631 vacation rentals, 484 apartments, 107 homestays, 101 B&Bs, 91 guest houses

Amsterdam Netherlands 1059 vacation rentals, 496 B&Bs, 451 apartments, 118 homestays, 53 guest houses

Eilat Israel 331 vacation rentals, 216 apartments, 65 villas, 65 vacation homes, 41 homestays

Las Vegas United States of America 258 vacation rentals, 143 apartments, 95 vacation ho

Pave, turn your trip to Jerusalem into an adventure!

Tel Aviv, Eilat, Ein Bokek and Tiberias are all popular choices nearby

Tel Aviv

Double from ₪ 153

Add Tel Aviv

Eilat

Double from ₪ 76

Add Eilat

Ein Bokek

Double from ₪ 220

Add Ein Bokek

Tiberias

Double from ₪ 214

Add Tiberias

Booking.com

Recommending destinations

Not what you're looking for? Try your search again.

Showing 1–50

Recommended Destinations for You:

- Jerusalem** 89 properties available

Compared to the past 2 days, prices in Jerusalem for your dates are now lower!
- Haifa** 41 properties available

Prices in Haifa have gone up for your dates since yesterday.
- Eilat** 9 properties available

Prices in Eilat have gone up for your dates since yesterday.
- Tiberias** 23 properties available

Prices in Tiberias for your dates are the lowest we've seen in 40 days!
- Herzlia** 20 properties available

Prices in Herzlia have gone up for your dates since yesterday.
- Netanya** 15 properties available

Prices in Netanya have gone up for your dates since yesterday.

Check-in — Check-out

5% or more with Mid-Year Deals

Tel Aviv 1,438 p

Your trip

Finisterre ✓ 12 Dec - 13 Dec

Santiago de Compostela ✓ 13 Dec - 14 Dec

Extend your trip

Porto 14 Dec - 15 Dec

Madrid 14 Dec - 15 Dec

Vigo 14 Dec - 15 Dec

A Coruña 14 Dec - 15 Dec

Thanks Pavel!

Your booking in Santiago de Compostela is confirmed

*43% more

*7% cheaper

*16% cheaper

*132% more

*19% cheaper

Gozo 611 properties

Tenerife 7,283 prop

New Forest 216 properties

Al Madinal 376 proper

Discover

You'll love relaxation, shop... United States of America

You'll lov Italy

Tel Aviv Double from ₪ 153 Add Tel Aviv

Eilat Double from ₪ 76 Add Eilat

Ein Bokek Double from ₪ 220 Add Ein Bokek

Tiberias Double from ₪ 214 Add Tiberias

Brussels 629 properties available

Paris 4,298 properties available

List Your Property

Booking.com

What we'll cover

- Travel as a sequence modeling problem
 - Sequence modelling
 - The multi-destination trip problem
 - RNNs
- Implementation details
 - data collection, batching, training
 - practical considerations
 - putting it in production
- Some use cases
- Beyond single-step predictions

Sequence modeling

Sequence modeling

Typical example: language model

Sequence modeling

Typical example: language model

- Sentence scoring:

$$\Pr\{She \text{ likes reading books}\} > \Pr\{She \text{ likes reading bookings}\}$$

Sequence modeling

Typical example: language model

- Sentence scoring:

$$\Pr\{\text{She likes reading books}\} > \Pr\{\text{She likes reading bookings}\}$$

- Machine Translation:

$$\Pr\{\text{The apartment offers a terrace.} \mid \text{Het appartement beschikt over een terras.}\}$$

$$> \Pr\{\text{Free WiFi access throughout is available.} \mid \text{Het appartement beschikt over een terras.}\}$$

Sequence modeling

Typical example: language model

- Sentence scoring:

$$\Pr\{She \text{ likes reading books}\} > \Pr\{She \text{ likes reading bookings}\}$$

- Machine Translation:

$$\Pr\{\text{The apartment offers a terrace.} | \text{Het appartement beschikt over een terras.}\}$$

$$> \Pr\{\text{Free WiFi access throughout is available.} | \text{Het appartement beschikt over een terras.}\}$$

- Image captioning:

$$\Pr\{\text{Room with a fireplace} |$$



$$| \text{ } \} > \Pr\{\text{Room with a private pool} |$$



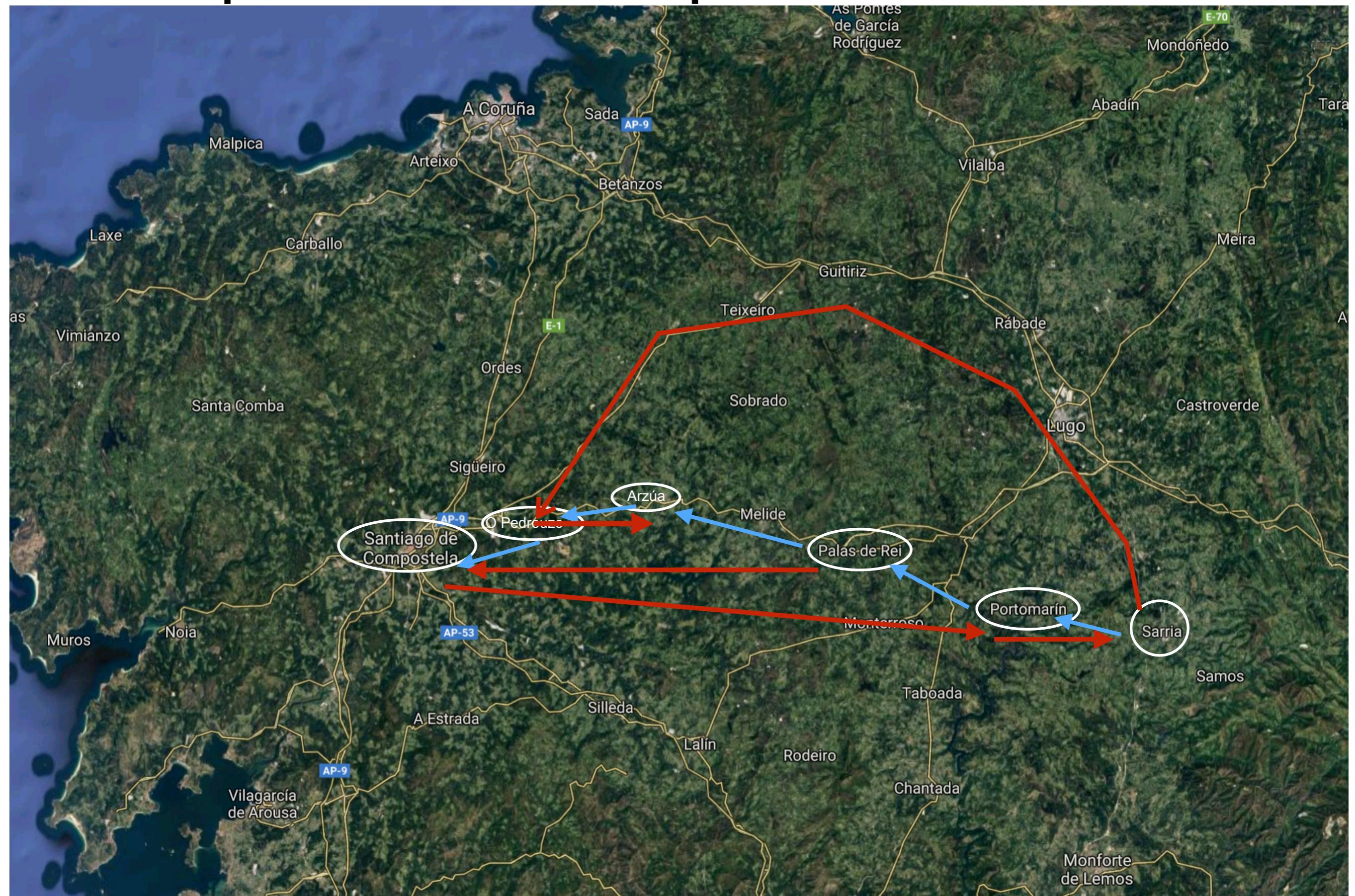
Trips are sequences too!



Trips are sequences too!



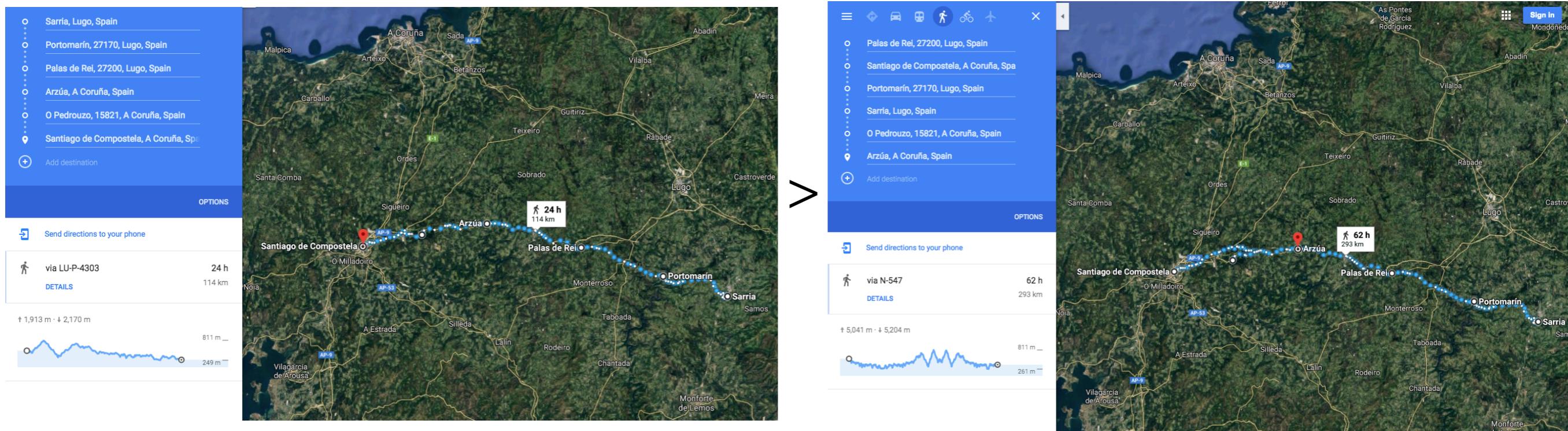
Trips are sequences too!



Trips are sequences too!

$\Pr\{(Sarria \rightarrow Portomarin \rightarrow Palas de Rei \rightarrow Arzua \rightarrow Pedrouzo \rightarrow Santiago de Compostela)\}$

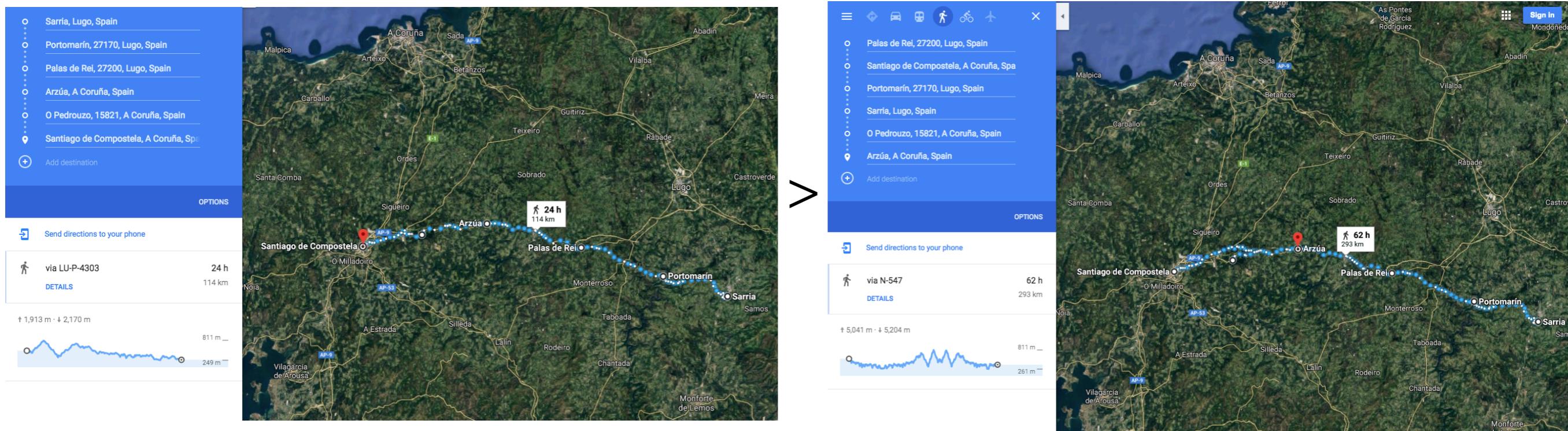
> $\Pr\{(Palas de Rei \rightarrow Santiago de Compostela \rightarrow Portmarin \rightarrow Sarria \rightarrow Pedrouzo \rightarrow Arzua)\}$



Trips are sequences too!

$\Pr\{(Sarria \rightarrow Portomarin \rightarrow Palas de Rei \rightarrow Arzua \rightarrow Pedrouzo \rightarrow Santiago de Compostela)\}$

> $\Pr\{(Palas de Rei \rightarrow Santiago de Compostela \rightarrow Portomarin \rightarrow Sarria \rightarrow Pedrouzo \rightarrow Arzua)\}$



Similarly:

$\Pr\{next_dest = \text{Santiago de Compostela} | so_far = (\text{Sarria}, \text{Portomarin}, \text{Palas de Rei}, \text{Arzua}, \text{Pedrouzo})\}$

> $\Pr\{next_dest = \text{Barcelona} | so_far = (\text{Sarria}, \text{Portomarin}, \text{Palas de Rei}, \text{Arzua}, \text{Pedrouzo})\}$

The multi-leg trip problem

- We have some censored information about the trip (e.g. first few legs)
- Want to predict missing legs
- This work: predicting forward legs only

The multi-leg trip problem

- We have some censored information about the trip (e.g. first few legs)
- Want to predict missing legs
- This work: predicting forward legs only

The problem:

Prediction

$$\operatorname{argmax}_{d_n \in D} P(d_n | d_1, \dots, d_{n-1}; \Theta)$$

Learning

$$\operatorname{argmax}_{\Theta} \mathbb{E}[P(d_k | d_1, \dots, d_{k-1}; \Theta)]$$

Notation:

D set of all known destinations

$\mathbf{d} = (d_1, d_2, \dots, d_n)$ trip (ordered list of destinations)

Θ model parameters

$P(\mathbf{d} | \Theta)$ probability of the trip/ likelihood of the model

$P(d_n | d_1, \dots, d_{n-1}; \Theta)$ next leg probability

The multi-leg trip problem

- We have some censored information about the trip (e.g. first few legs)
- Want to predict missing legs
- This work: predicting forward legs only

The problem:

Prediction $\operatorname{argmax}_{d_n \in D} P(d_n | d_1, \dots, d_{n-1}; \Theta)$

Learning $\operatorname{argmax}_{\Theta} \mathbb{E}[P(d_k | d_1, \dots, d_{k-1}; \Theta)]$

Notation:

D set of all known destinations

$\mathbf{d} = (d_1, d_2, \dots, d_n)$ trip (ordered list of destinations)

Θ model parameters

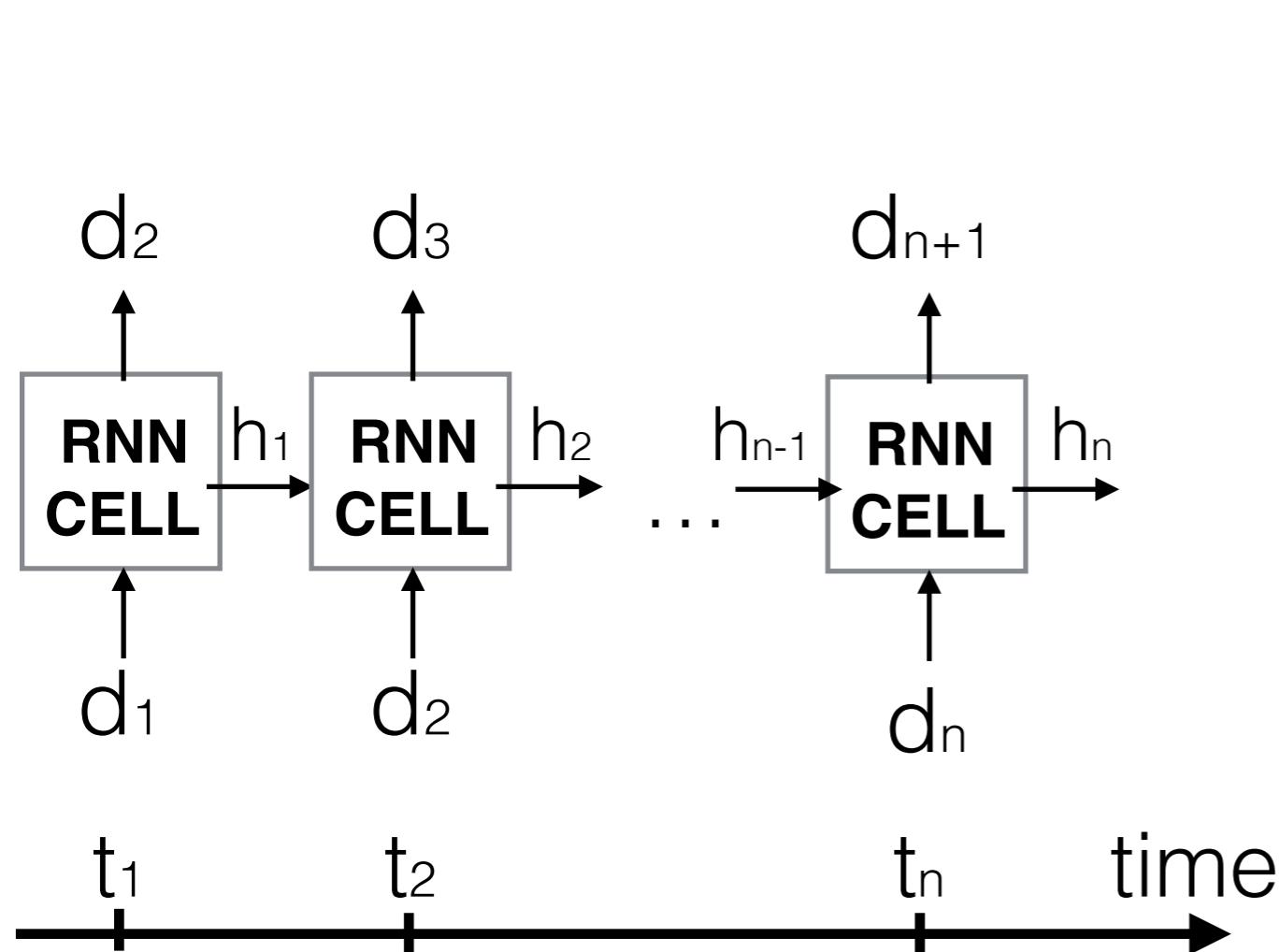
$P(\mathbf{d} | \Theta)$ probability of the trip/ likelihood of the model

$P(d_n | d_1, \dots, d_{n-1}; \Theta)$ next leg probability

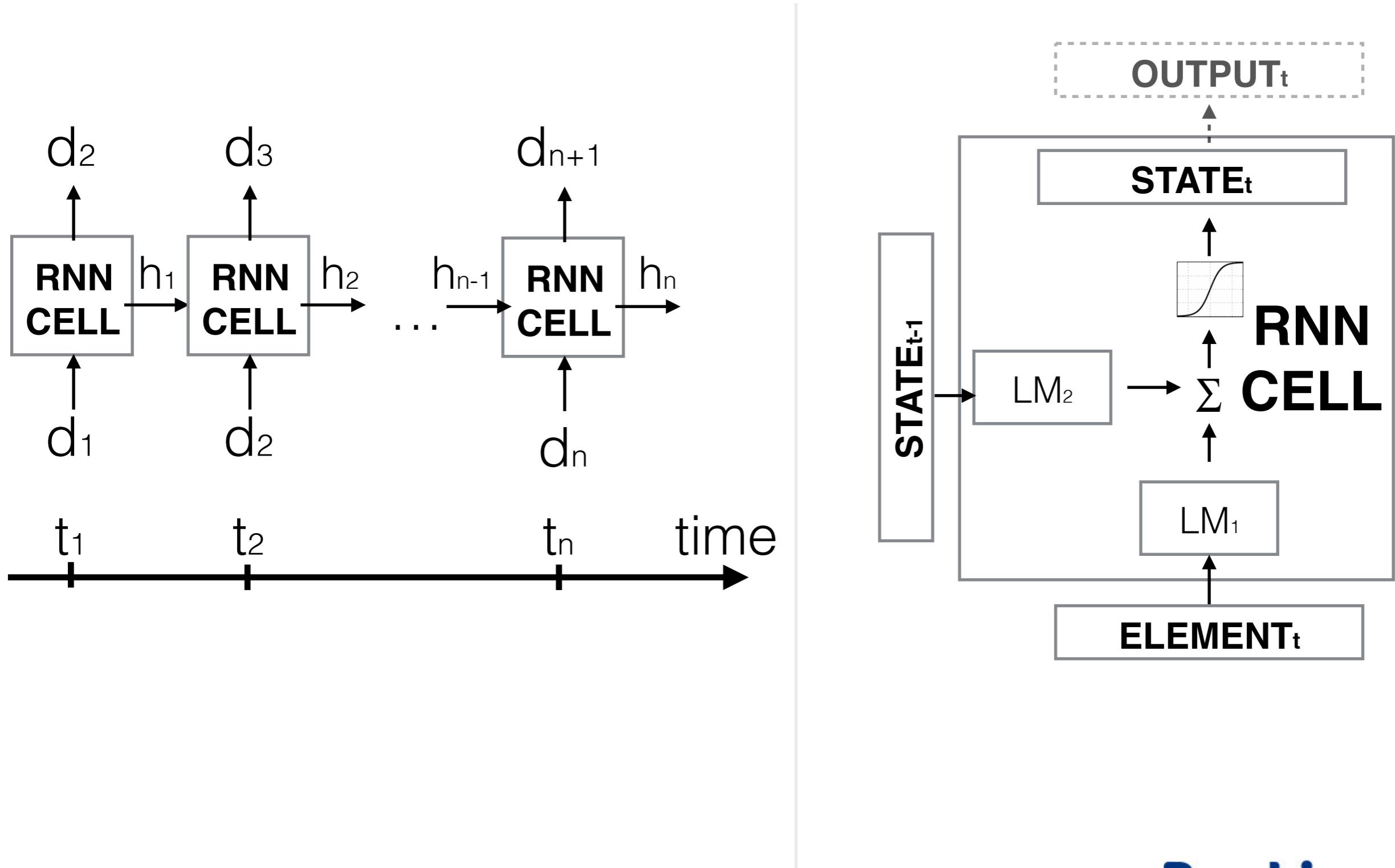
Can use RNNs to solve!

Booking.com

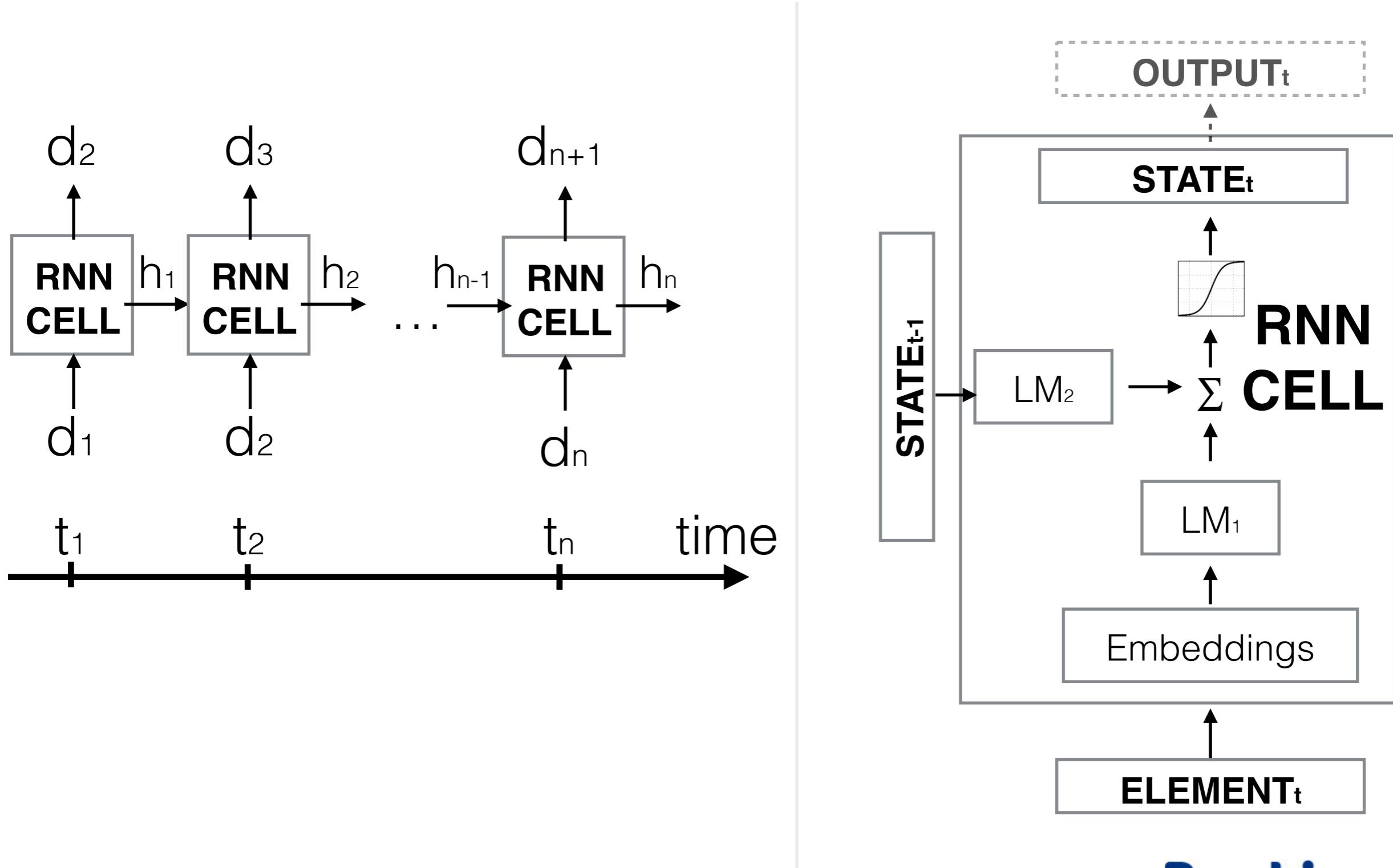
What's an RNN?



What's an RNN?



What's an RNN?



Learning RNN

Learning RNN

The RNN model:

$$x_t = \Theta_d d_t$$
$$h_t = \tanh(\Theta_x(x_t) + \Theta_h(h_{t-1}) + \theta_h)$$
$$y_t = \sigma(\Theta_y h_t + \theta_y)$$

Learning RNN

The RNN model:

$$x_t = \Theta_d d_t$$
$$h_t = \tanh(\Theta_x(x_t) + \Theta_h(h_{t-1}) + \theta_h)$$
$$y_t = \sigma(\Theta_y h_t + \theta_y)$$

What are we optimizing?

Learning RNN

The RNN model:

$$x_t = \Theta_d d_t$$
$$h_t = \tanh(\Theta_x(x_t) + \Theta_h(h_{t-1}) + \theta_h)$$
$$y_t = \sigma(\Theta_y h_t + \theta_y)$$

What are we optimizing?

For any given sequence \mathbf{d} , we want

$$-\frac{1}{|\mathbf{d}|} \sum_{t=1}^{|\mathbf{d}|} \log(P[RNN(d_t, h_{t-1}) = d_{t+1}])$$
 to be as small as possible

Learning RNN

The RNN model:

$$x_t = \Theta_d d_t$$
$$h_t = \tanh(\Theta_x(x_t) + \Theta_h(h_{t-1}) + \theta_h)$$
$$y_t = \sigma(\Theta_y h_t + \theta_y)$$

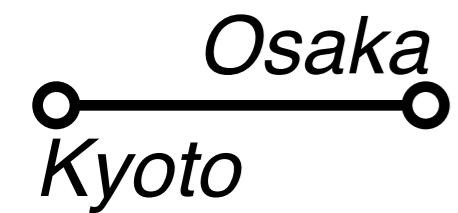
What are we optimizing?

For any given sequence \mathbf{d} , we want

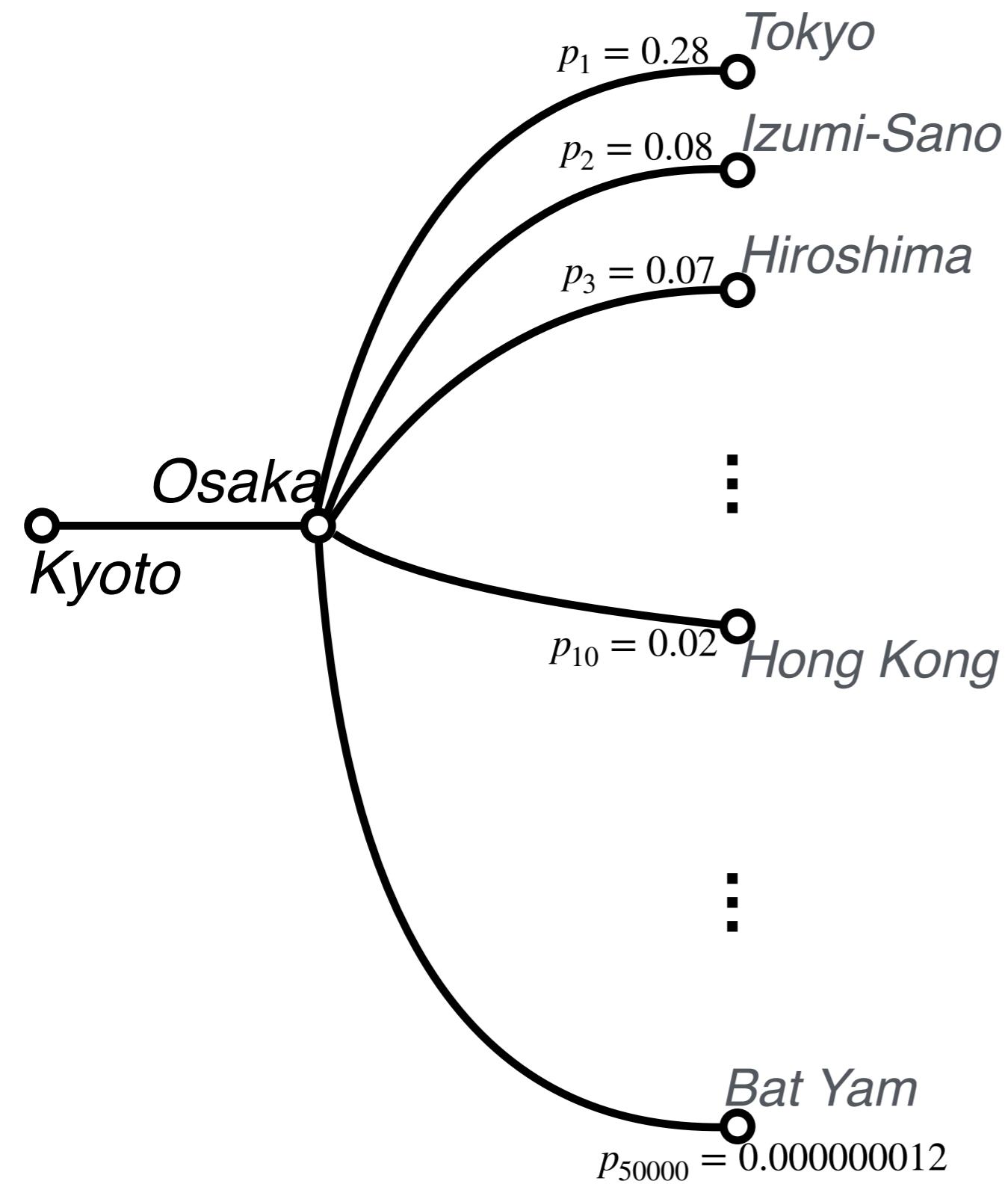
$$-\frac{1}{|\mathbf{d}|} \sum_{t=1}^{|\mathbf{d}|} \log(P[RNN(d_t, h_{t-1}) = d_{t+1}])$$
 to be as small as possible

i.e. minimize average **log loss** for next element classification
aka minimize average **Cross Entropy** loss
aka minimize **perplexity**

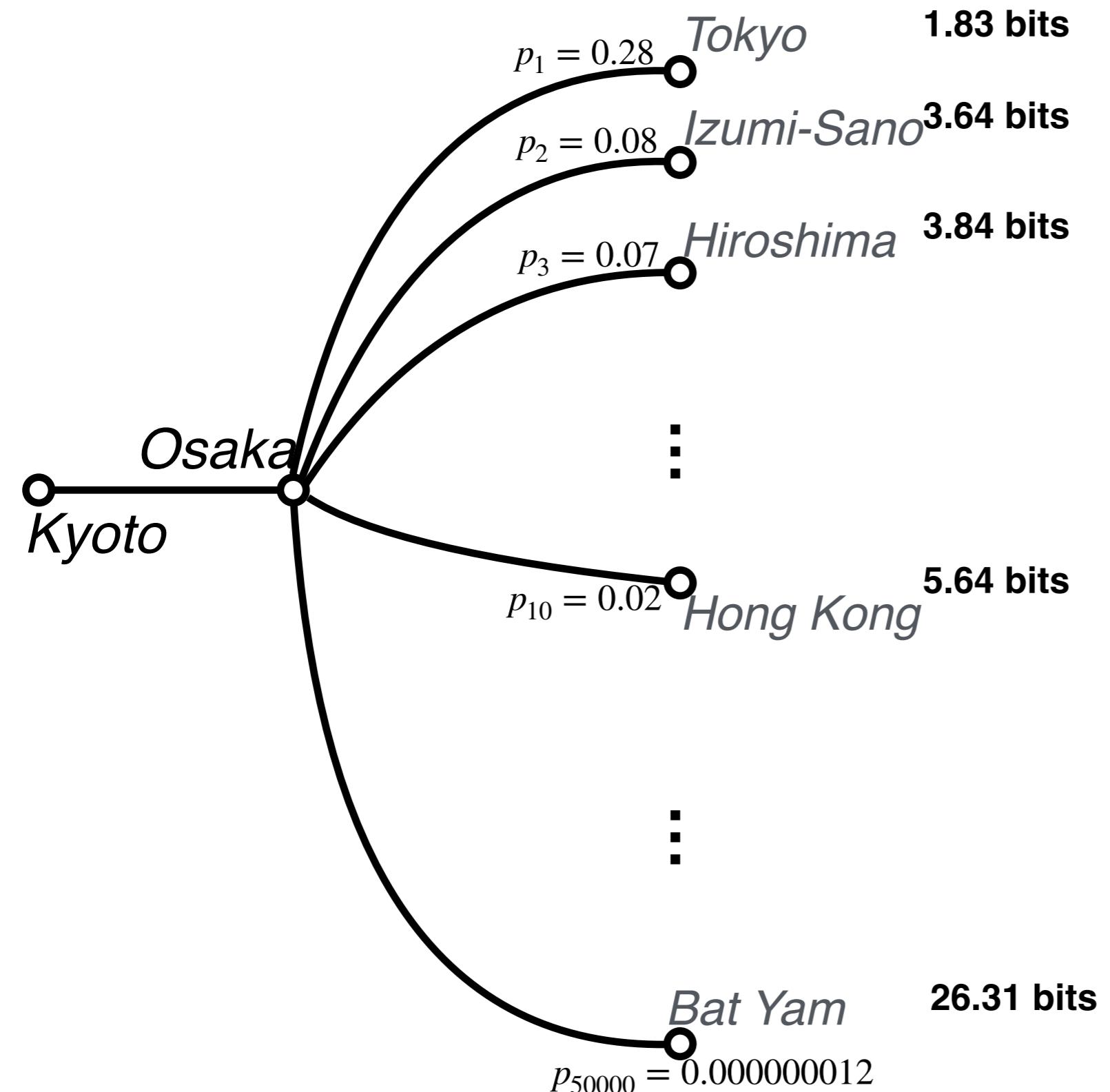
Perplexity: intuition



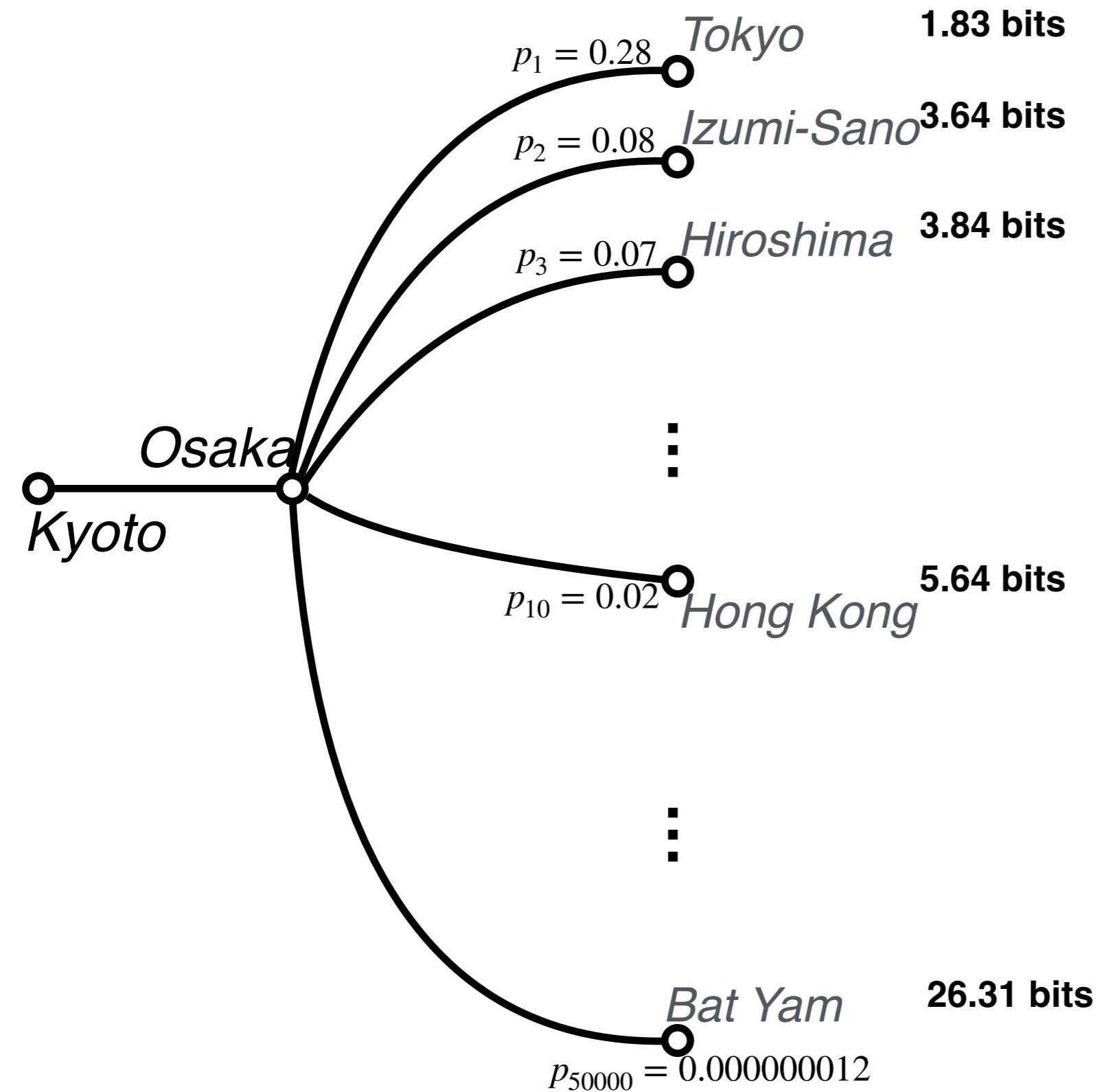
Perplexity: intuition



Perplexity: intuition

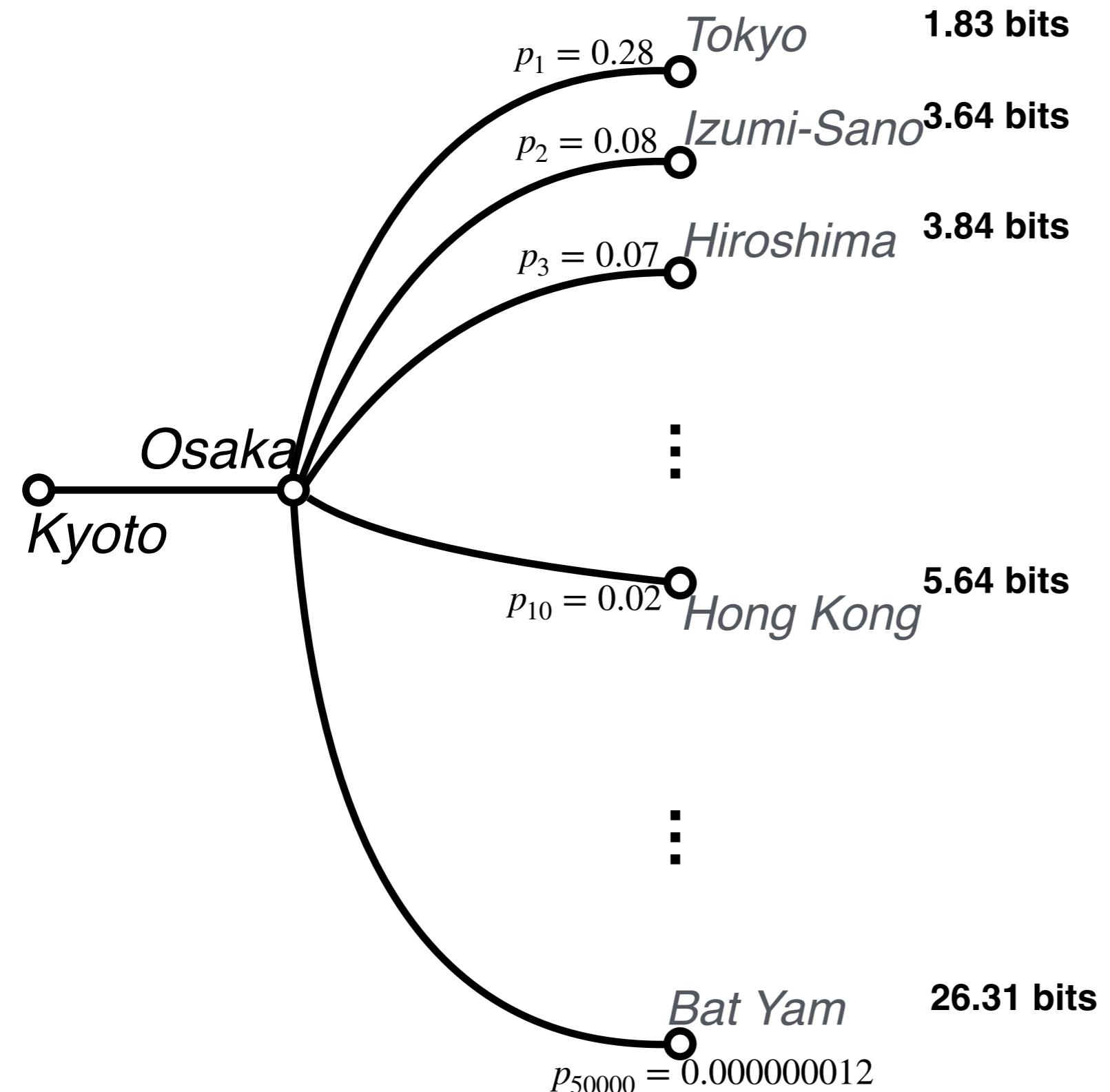


Perplexity: intuition



$$H = 0.28 \times 1.83 + 0.08 \times 3.64 + \dots + (1.2e - 8) * 26.31 = 5.26$$

Perplexity: intuition

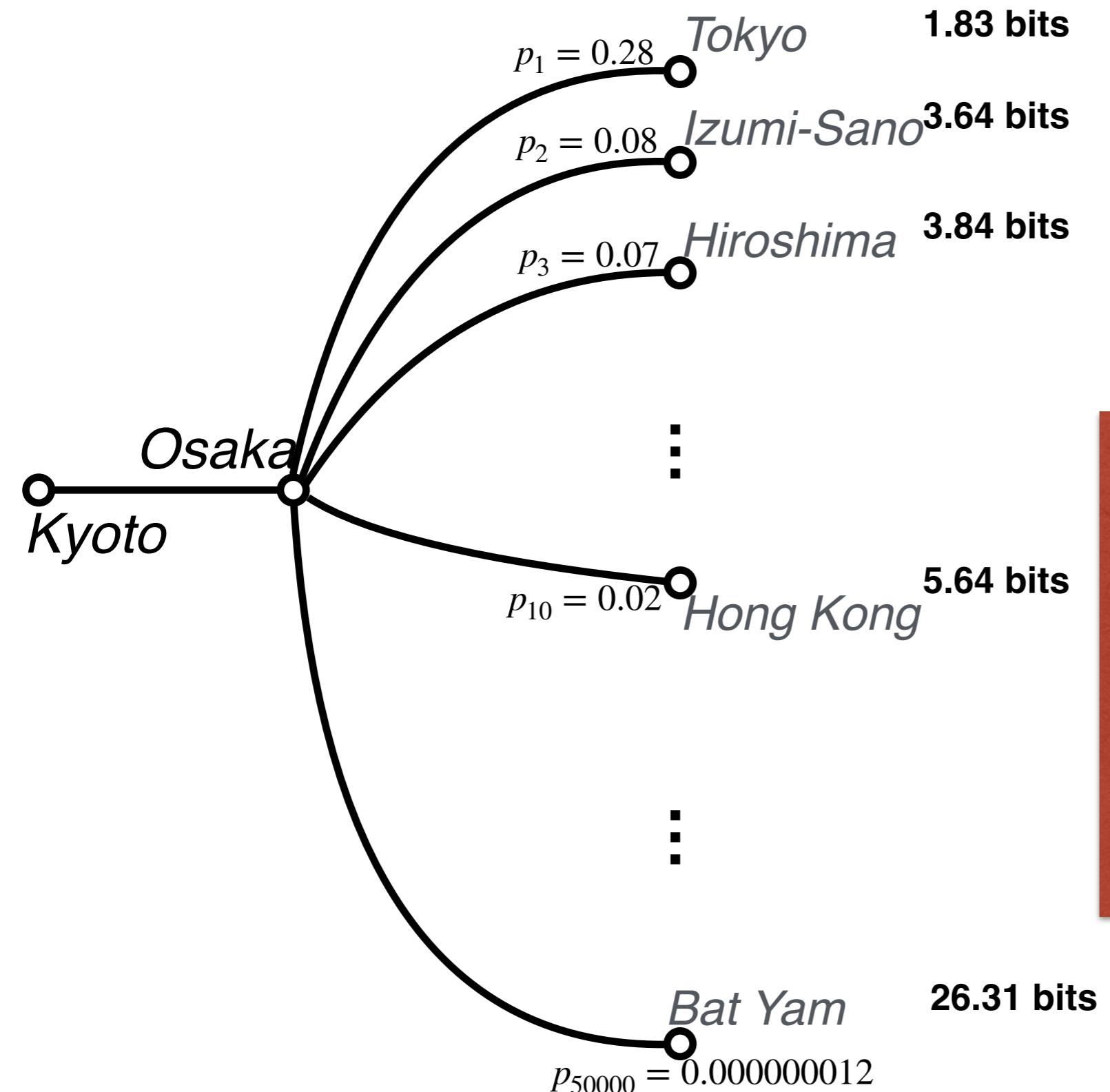


$$H = 0.28 \times 1.83 + 0.08 \times 3.64 + \dots$$

$$+(1.2e - 8) * 26.31 = 5.26$$

$$2^{5.26} = 38.4$$

Perplexity: intuition



$$H = 0.28 \times 1.83 + 0.08 \times 3.64 + \dots$$

$$+(1.2e - 8) * 26.31 = 5.26$$

$$2^{5.26} = 38.4$$

Uncertainty of
flipping a fair 38.4-
sided coin!

Perplexity: intuition

$$\text{PPL} = 2^H$$

$$H(\tilde{p}, q) = - \sum_x \tilde{p}(x) \log_2 q(x)$$

- How many typical likely next steps can I take?
- Average effective branching factor of the next leg probability tree
- Lower perplexity = more predictable sequences
- Find model which is the least perplexed by the data (likelihood principle)

Perplexity: intuition

$$\text{PPL} = 2^H$$

Our loss function!

$$H(\tilde{p}, q) = - \sum_x \tilde{p}(x) \log_2 q(x)$$

- How many typical likely next steps can I take?
- Average effective branching factor of the next leg probability tree
- Lower perplexity = more predictable sequences
- Find model which is the least perplexed by the data (likelihood principle)

Implementation

The category
Details

Data collection

- ~23 million trips from 2017
- Only look at top 50k destinations
- COLLECT() bookings into trips
- Remove consecutive duplicates (stays in different hotels of the same city)
- Randomize and dump into a text file

Data collection

- ~23 million trips from 2017

- Only look at top 50k destinations

- COLLECT() bookings into trips

- Remove consecutive duplicates (stays in different hotels of the same city)

- Randomize and dump into a text file

sketch in Spark

```
# from pyspark.sql import functions as f
# from pyspark.sql.functions import udf
# from pyspark.sql import Window
# from pyspark.sql.types import ArrayType, StringType
# from functools import reduce

(spark.sql('''SELECT user_id, checkin, checkout, destination_id from Reservations''')
.join(top_destinations, how='left_semi', on= 'destination_id') # DataFrame with top destinations
.withColumn('res_rank', f.row_number().over(book_window_checkin)) # Window.partitionBy('user_id').orderBy('checkin')
.withColumn('next_checkin', f.lead('checkin').over(book_window_checkin))
.withColumn('days_to_next', f.datediff(f.col('next_checkin'), f.col('checkout')))
.withColumn('prev_checkout', f.lag('checkout').over(book_w_checkin))
.withColumn('days_since_last', f.datediff(f.col('checkin'), f.col('prev_checkout')))

.filter('days_to_next <= %i' OR 'days_since_last <= %i')%(ALLOWED_GAP, ALLOWED_GAP) # ALLOWED_GAP=0: perfect trips only
.orderBy(['user_id', 'checkin'])
.withColumn('trip_id_', f.when(f.col('days_since_last')>ALLOWED_GAP, 1).otherwise(0))
.withColumn('trip_id', f.sum('trip_id_').over(book_window_checkin))
.drop('trip_id_')
.groupBy('user_id', 'trip_id')
.agg(remove_duplicates_udf(f.collect_list('destination_id')).alias('trip'))
.withColumn('trip_length', f.size('trip'))
.filter('trip_length > 1')
.orderBy(f.rand())
)

# remove_duplicates_udf = f.udf(lambda x:reduce(lambda x,y:((x if x[-1]==y else x+[y]) if isinstance(x, list) else ([x] if x==y else [x,y])), x)
# , ArrayType(StringType()))
```

Data collection

- ~23 million trips from 2017

- Only look at top 50k destinations

- COLLECT() bookings into trips

- Remove consecutive duplicates (stays in different hotels of the same city)

```
# from pyspark.sql import functions
# from pyspark.sql.functions import col
# from pyspark.sql import Window
# from pyspark.sql.types import ArrayType
# from functools import reduce

(spark.sql('''SELECT user_id, checkin
    .join(top_destinations, how='left')
    .withColumn('res_rank', f.row_number()
    .withColumn('next_checkin', f.lead(checkin))
    .withColumn('days_to_next', f.datediff(next_checkin, checkin))
    .withColumn('prev_checkout', f.lag(checkout))
    .withColumn('days_since_last', f.datediff(prev_checkout, checkin))

    .filter('(days_to_next <= %i) OR (days_since_last >= %i)' % (ALLOWED_GAP, ALLOWED_GAP)) # ALLOWED_GAP=0: perfect trips only
    .orderBy(['user_id', 'checkin'])
    .withColumn('trip_id_', f.when(f.col('days_since_last')>ALLOWED_GAP, 1).otherwise(0))
    .withColumn('trip_id', f.sum('trip_id_').over(book_window_checkin))
    .drop('trip_id_')
    .groupBy('user_id', 'trip_id')
    .agg(remove_duplicates_udf(f.collect_list('destination_id')).alias('trip'))
    .withColumn('trip_length', f.size('trip'))
    .filter('trip_length > 1')
    .orderBy(f.rand())
)

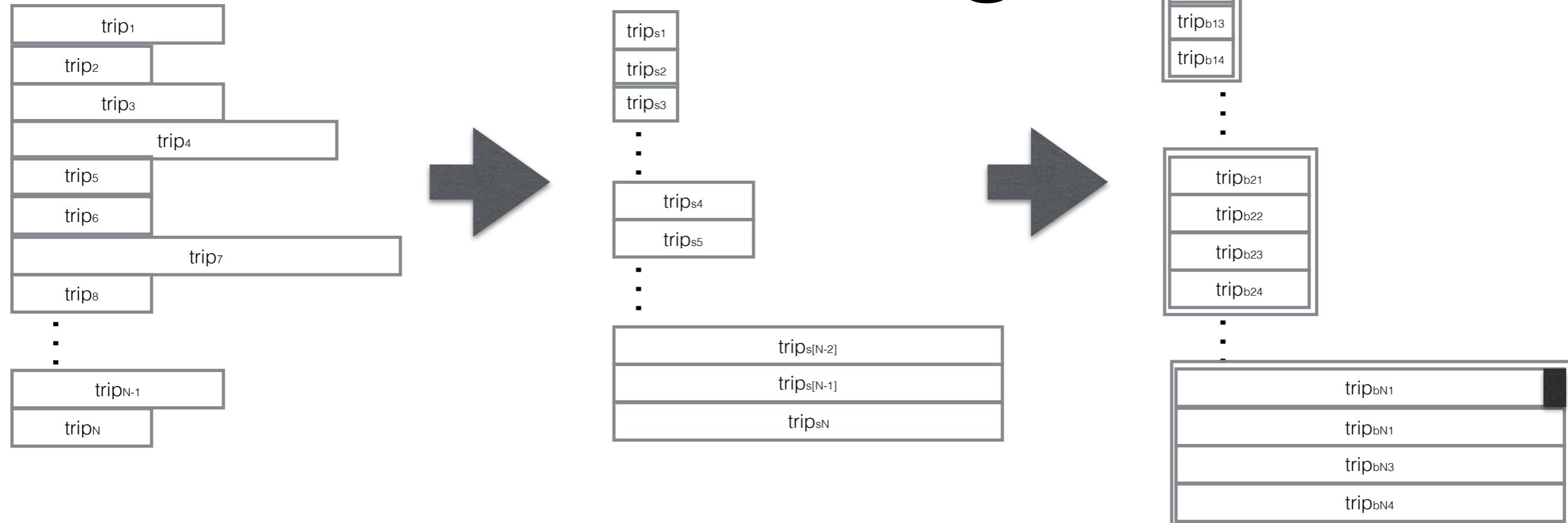
# remove_duplicates_udf = f.udf(lambda x:reduce(lambda x,y:((x if x[-1]==y else x+[y]) if isinstance(x, list) else ([x] if x==y else [x,y])), x),
#                                ArrayType(StringType()))
```

Approximately:
23 million trips

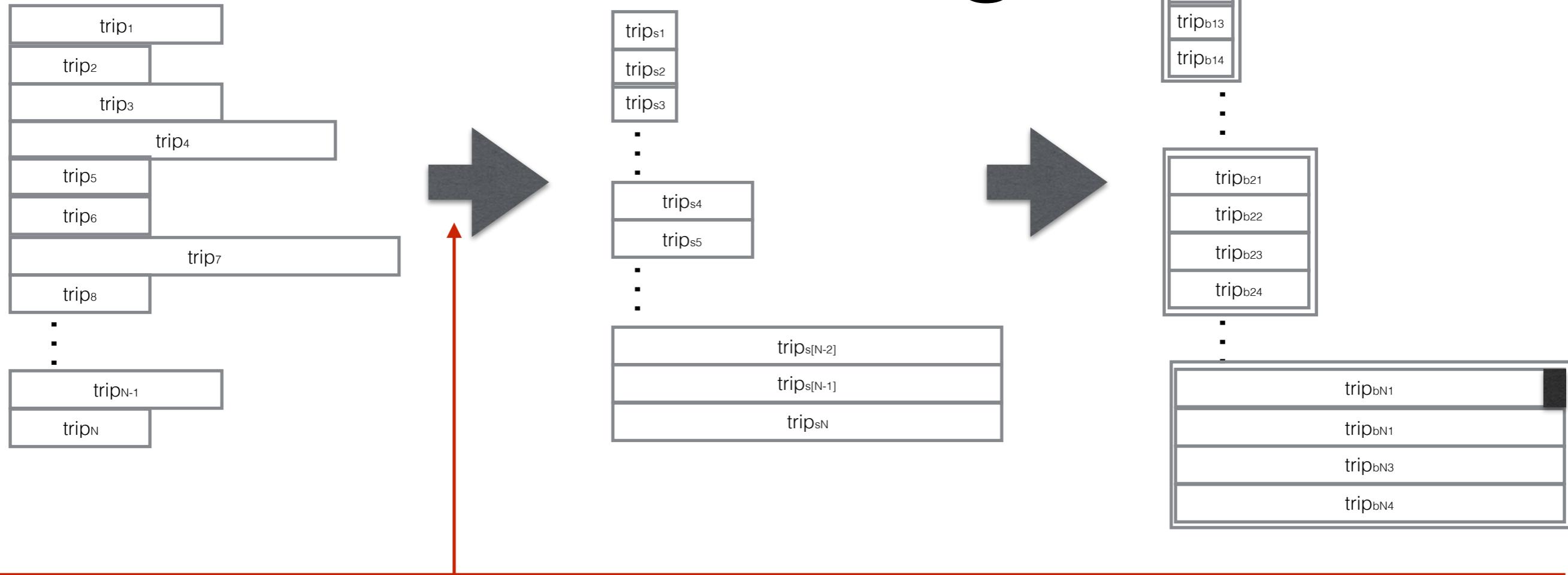
90% <5 legs
10% 5+ legs

- Randomize and dump into a text file

Batching

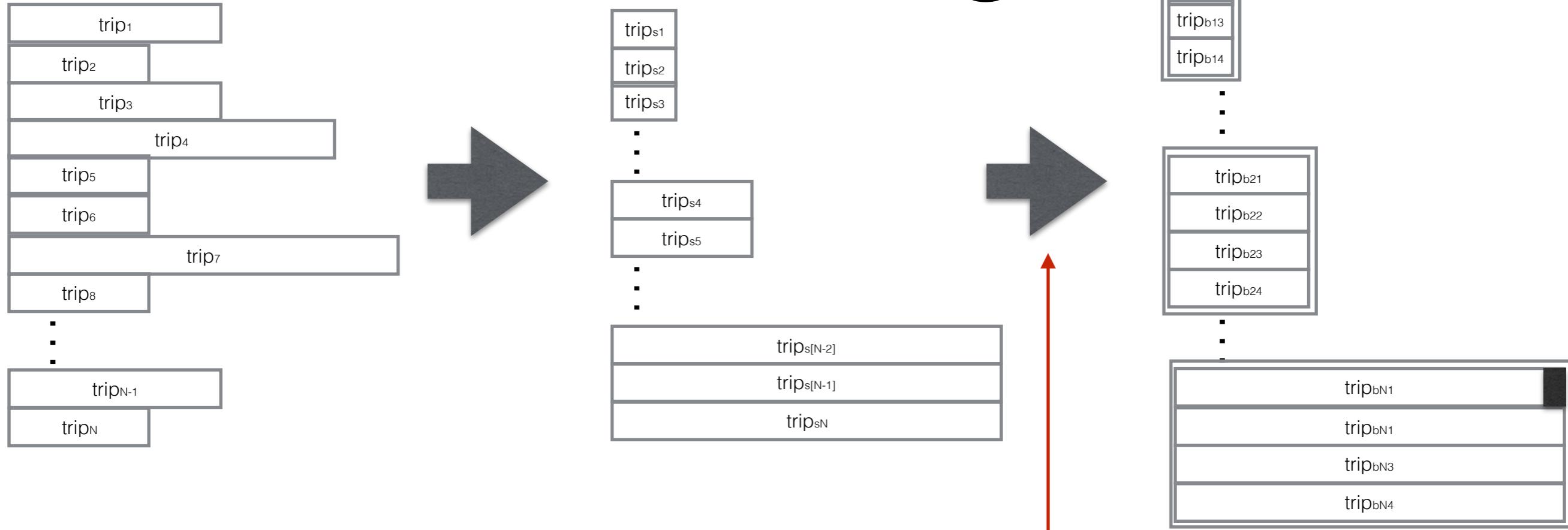


Batching



```
with open(DATA_FILE) as f:
    for line in f:
        lines.append([word2id.get(_, word2id['<UNK>']) for _ in (line.split() + ['<EOS>'])])
    length = len(lines[-1])
    if length > MAX_SEQ_LEN:
        lines[-1] = lines[-1][:MAX_SEQ_LEN]
        length = MAX_SEQ_LEN
    lens.append(length)
    order = np.argsort(lens)
lines_sorted = sorted(lines, key=len)
```

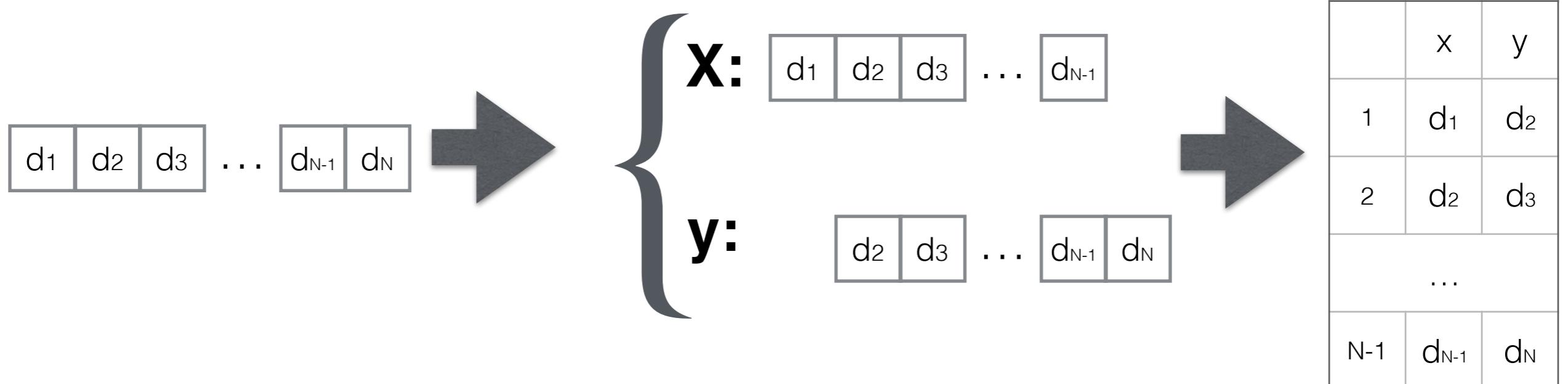
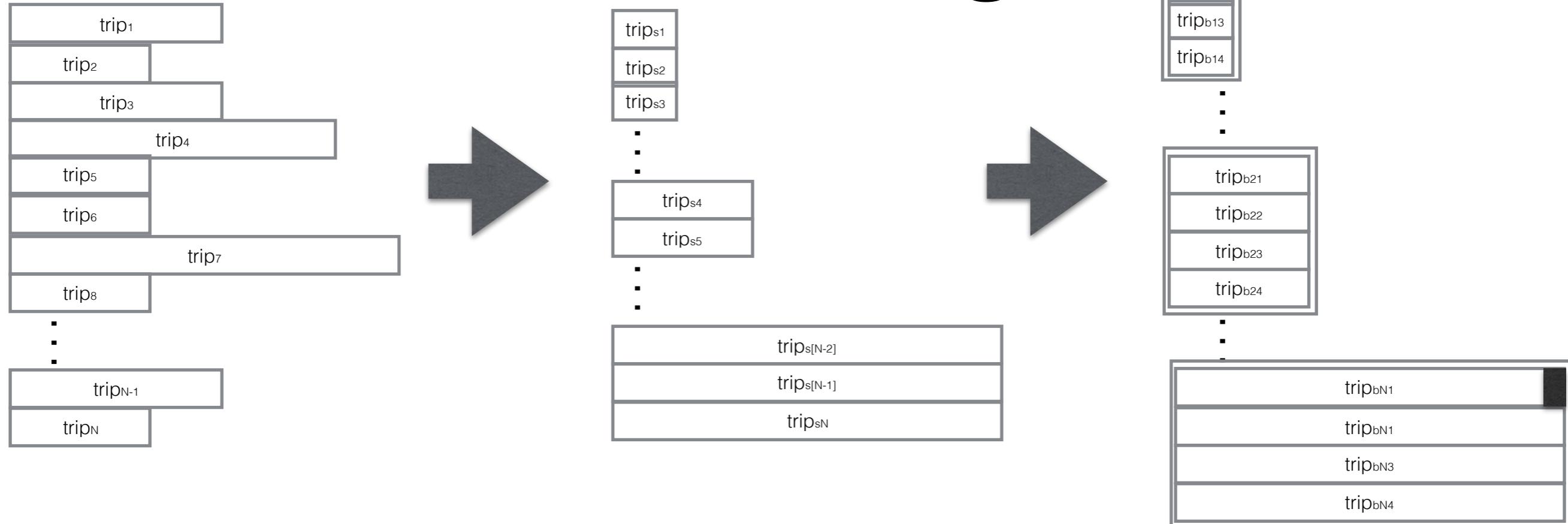
Batching



```
# from torch.autograd import Variable
# if CUDA_ENABLED:
#     from torch.cuda import LongTensor
# else:
#     from torch import LongTensor

# GETTING BATCH i
lines = lines_sorted[i:i+batch_size]
max_len = np.max([len(_) for _ in lines])
data = LongTensor([pad_sequence(_, max_len) for _ in lines])
source, target = (Variable(data.narrow(1, 0, data.size(1)-1).t()),
                  Variable(data.narrow(1, 1, data.size(1)-1).t().contiguous().view(-1)))
)
```

Batching



The model

```
class RNNLM(nn.Module):
    def __init__(self, vocab_size, emb_dim=500, rnn_dim=500, n_layers=2, dropout=0.3, rnn_dropout=0.3, tie_weights=False):
        super(RNNLM, self).__init__()
        self.drop = nn.Dropout(dropout)
        self.embedding = nn.Embedding(vocab_size, emb_dim)
        self.rnn = nn.RNN(input_size=emb_dim, hidden_size=rnn_dim, num_layers=n_layers, dropout=rnn_dropout)
        self.dense = nn.Linear(rnn_dim, vocab_size)

        if tie_weights:
            self.dense.weight = self.embedding.weight
        self.vocab_size = vocab_size
        self.emb_dim = emb_dim
        self.rnn_dim = rnn_dim
        self.n_layers = n_layers
        self.dropout = dropout
        self.rnn_dropout = rnn_dropout
        self.tie_weights = tie_weights

    def forward(self, input, h0=None):
        out, hidden = self.rnn(self.drop(self.embedding(input)), h0)
        out = self.dense(self.drop(out.view(out.size(0)*out.size(1), out.size(2))))
        return out, hidden
```

The model

```
class RNNLM(nn.Module):
    def __init__(self, vocab_size, emb_dim=500, rnn_dim=500, n_layers=2, dropout=0.3, rnn_dropout=0.3, tie_weights=False):
        super(RNNLM, self).__init__()
        self.drop = nn.Dropout(dropout)
        self.embedding = nn.Embedding(vocab_size, emb_dim)
        self.rnn = nn.RNN(input_size=emb_dim, hidden_size=rnn_dim, num_layers=n_layers, dropout=rnn_dropout)
        self.dense = nn.Linear(rnn_dim, vocab_size)

        if tie_weights:
            self.dense.weight = self.embedding.weight
        self.vocab_size = vocab_size
        self.emb_dim = emb_dim
        self.rnn_dim = rnn_dim
        self.n_layers = n_layers
        self.dropout = dropout
        self.rnn_dropout = rnn_dropout
        self.tie_weights = tie_weights

    def forward(self, input, h0=None):
        out, hidden = self.rnn(self.drop(self.embedding(input)), h0)
        out = self.dense(self.drop(out.view(out.size(0)*out.size(1), out.size(2))))
        return out, hidden
```

The model

```
class RNNLM(nn.Module):
    def __init__(self, vocab_size, emb_dim=500, rnn_dim=500, n_layers=2, dropout=0.3, rnn_dropout=0.3, tie_weights=False):
        super(RNNLM, self).__init__()
        self.drop = nn.Dropout(dropout)
        self.embedding = nn.Embedding(vocab_size, emb_dim)
        self.rnn = nn.RNN(input_size=emb_dim, hidden_size=rnn_dim, num_layers=n_layers, dropout=rnn_dropout)
        self.dense = nn.Linear(rnn_dim, vocab_size)

        if tie_weights:
            self.dense.weight = self.embedding.weight
        self.vocab_size = vocab_size
        self.emb_dim = emb_dim
        self.rnn_dim = rnn_dim
        self.n_layers = n_layers
        self.dropout = dropout
        self.rnn_dropout = rnn_dropout
        self.tie_weights = tie_weights

    def forward(self, input, h0=None):
        out, hidden = self.rnn(self.drop(self.embedding(input)), h0)
        out = self.dense(self.drop(out.view(out.size(0)*out.size(1), out.size(2))))
        return out, hidden
```

$$\Theta_d = \Theta_y$$

Published as a conference paper at ICLR 2017

TYING WORD VECTORS AND WORD CLASSIFIERS: A LOSS FRAMEWORK FOR LANGUAGE MODELING

Hakan Inan, Khashayar Khosravi
Stanford University
Stanford, CA, USA
{inan,hkhosravi}@stanford.edu

Richard Socher
Salesforce Research
Palo Alto, CA, USA
rsocher@salesforce.com

ABSTRACT

Recurrent neural networks have been very successful at predicting sequences of words in tasks such as language modeling. However, all such models are based on the assumption that words share the same hidden state.

Training

- Get fresh batch
(inputs, outputs)
- Get predictions
- Calculate loss on
the predictions
- Backpropagate
- Update model
parameters

Training

- Get fresh batch
(inputs, outputs)
- Get predictions
- Calculate loss on
the predictions
- Backpropagate
- Update model
parameters

```
for epoch in range(1, TOTAL_EPOCHS):           pytorch sketch
    for i, batch_n in enumerate(shuffled_batches):
        data_in, targets = get_batch(batch_n, batch_size)
        model.zero_grad()
        output = model(data_in)
        loss = criterion(output.view(-1, n_destinations), targets)
        loss.backward()
        optimizer.step()
        losses.append(loss.data[0])
```

Training demo

http://localhost:8888/notebooks/jerusalem_demo.ipynb

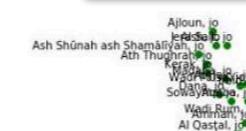
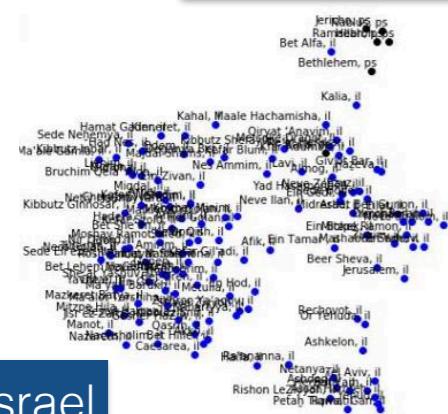
Training demo

http://localhost:8888/notebooks/jerusalem_demo.ipynb

Similar model open source (from official pytorch examples):

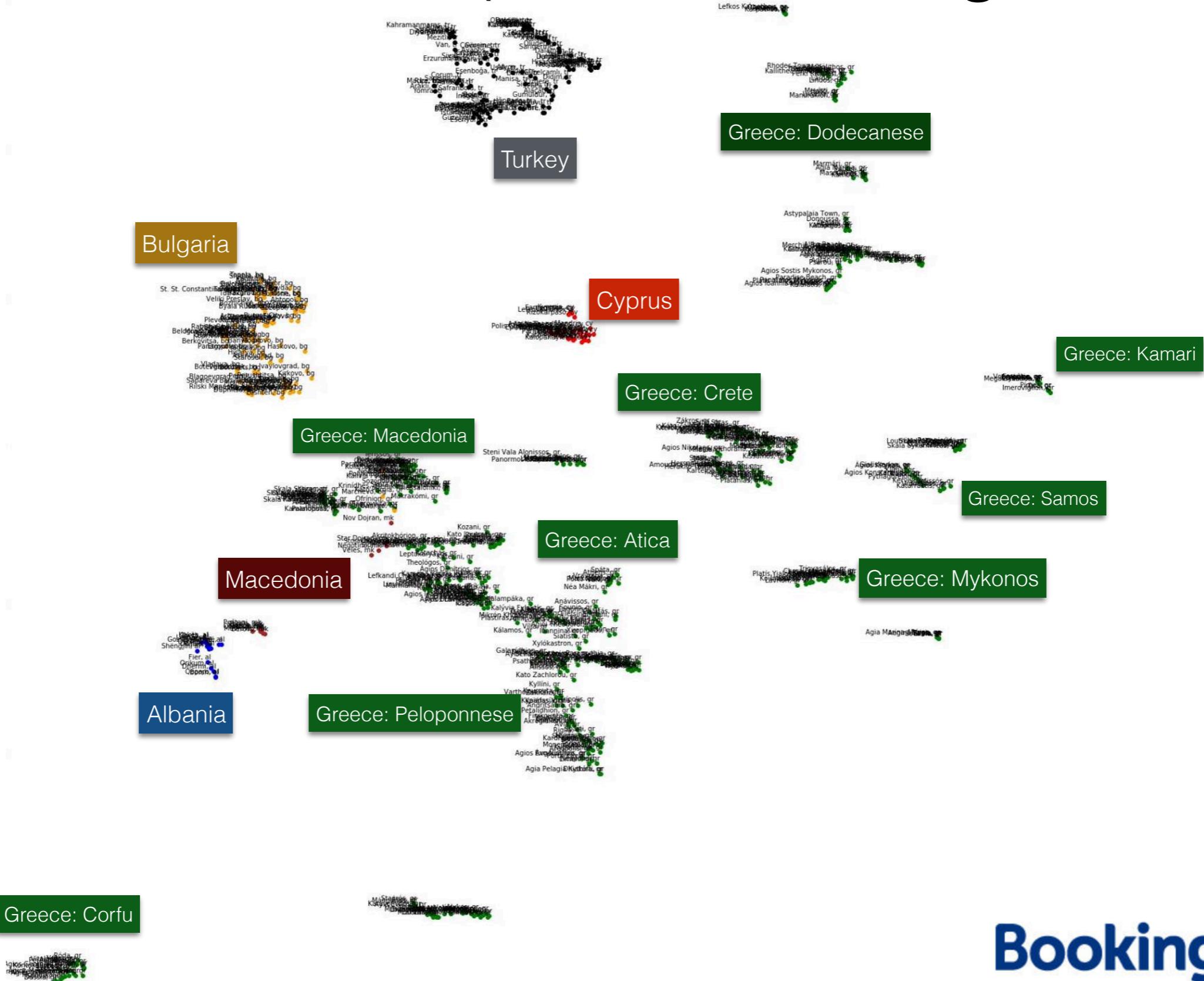
https://github.com/pytorch/examples/tree/master/word_language_model

Destination embeddings learned by the RNN model (middle east)



Booking.com

Destination embeddings learned by the RNN model (Greece + neighbors)



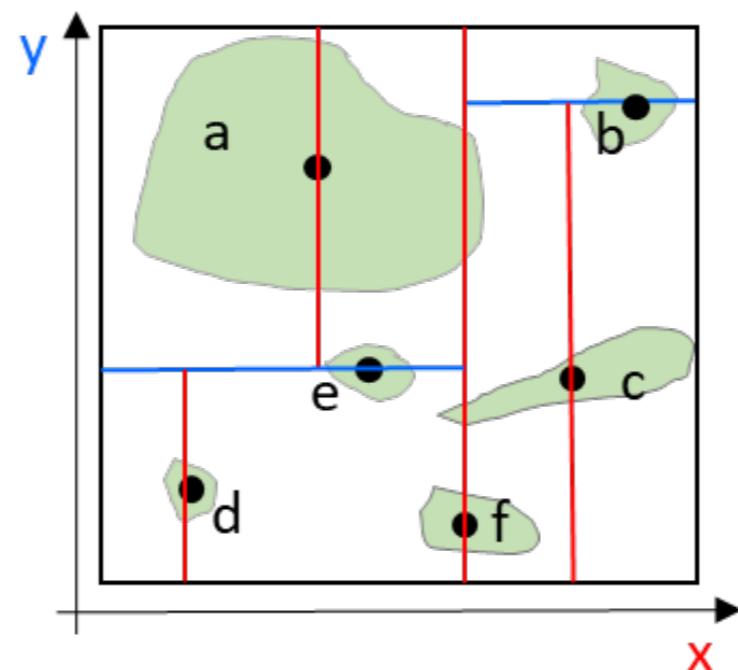
Booking.com

Handling unknown input

- Maintain destination mapping (MySQL table)
- Map all known destinations to the 50k most popular ones based on geography
- use kd-Tree for fast nearest neighbor search

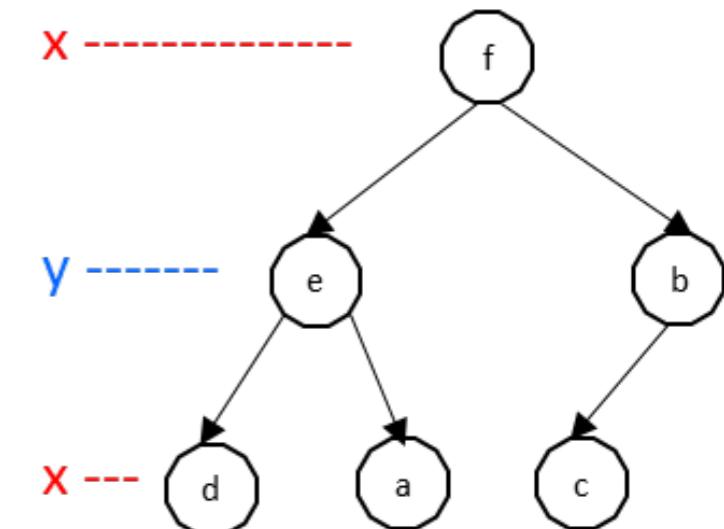
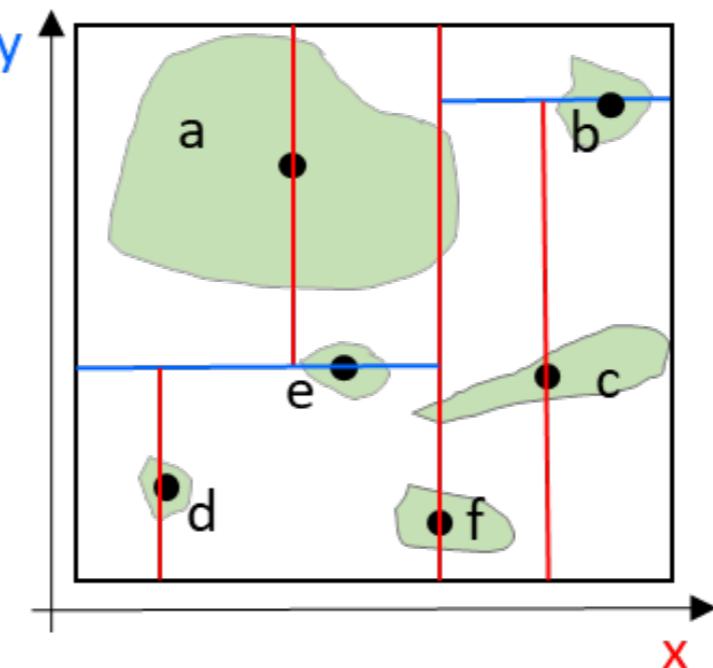
Handling unknown input

- Maintain destination mapping (MySQL table)
- Map all known destinations to the 50k most popular ones based on geography
- use kd-Tree for fast nearest neighbor search



Handling unknown input

- Maintain destination mapping (MySQL table)
- Map all known destinations to the 50k most popular ones based on geography
- use kd-Tree for fast nearest neighbor search



```
from scipy.spatial import cKDTree
dest_coords = (destination_data
                [['destination_id', 'latitude', 'longitude']]
                .values
                )
coords = [[lat, lon] for destination_id, lat, lon in dest_coords]
kdtree = cKDTree(coords)

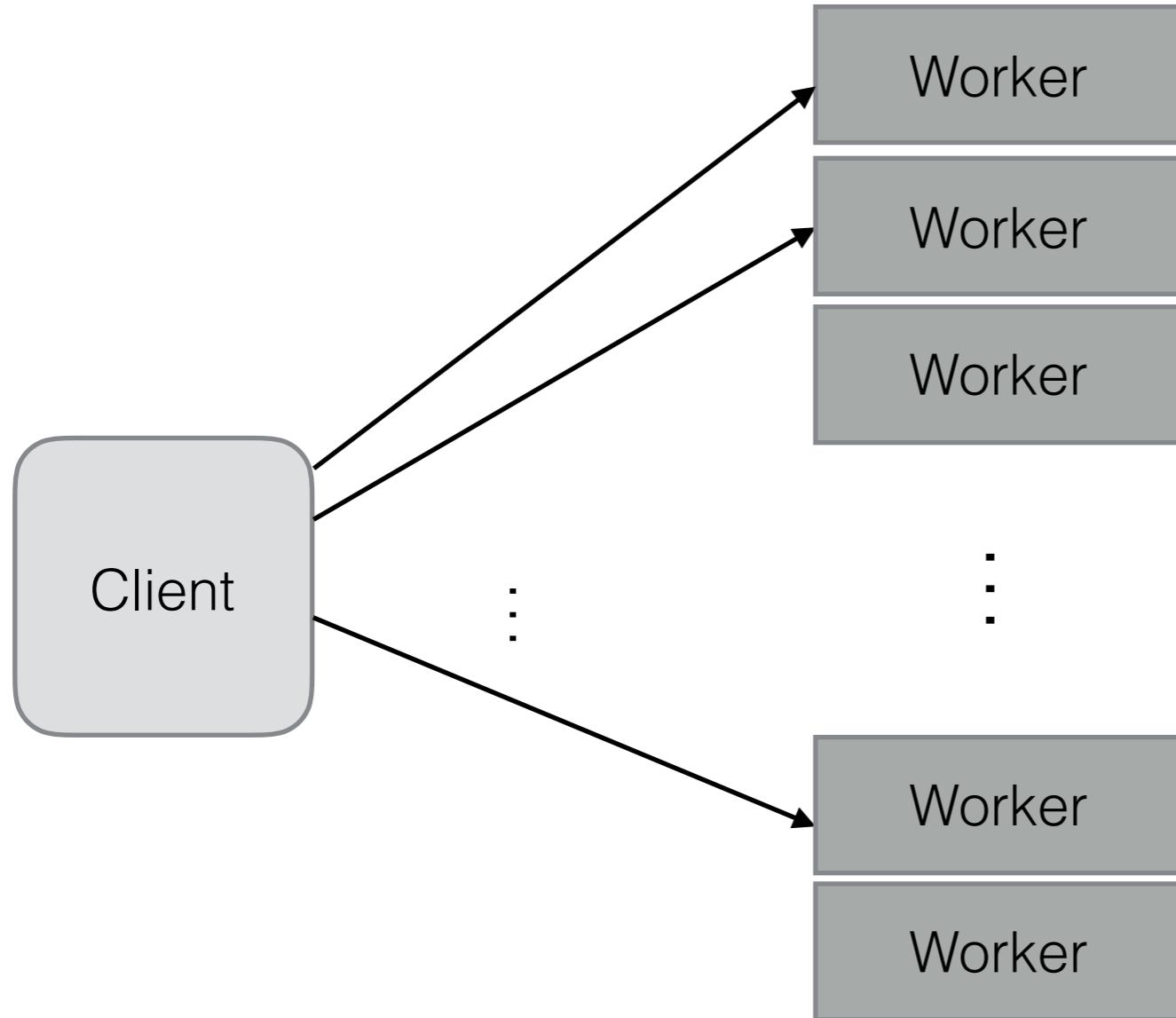
all_known_destinations['closest'] = (
    all_known_destinations
    .apply(lambda x: dest_coords[kdtree.query([x.latitude, x.longitude]), 0]
          , axis=1)
)
```

Python sketch

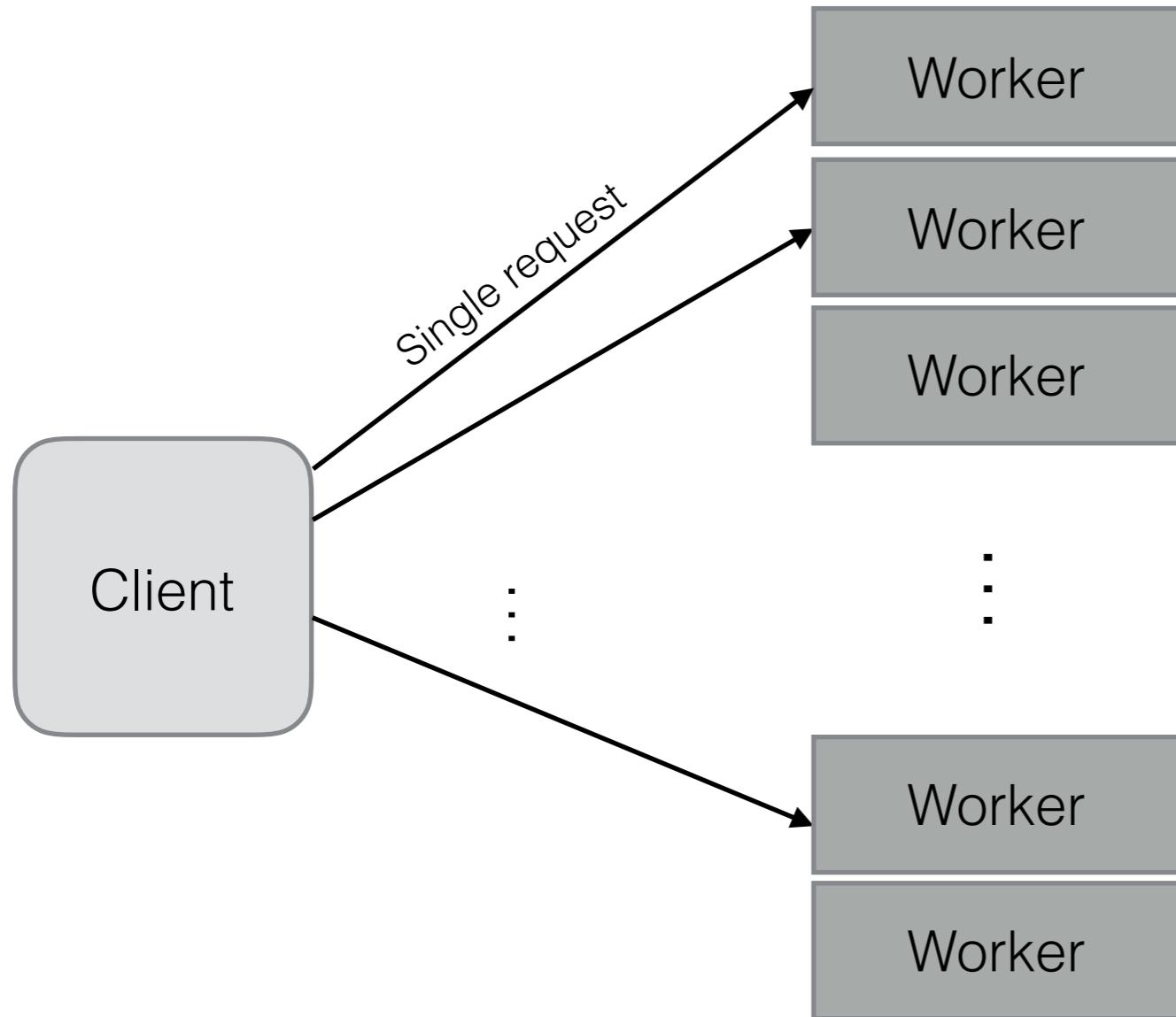
Putting it in production



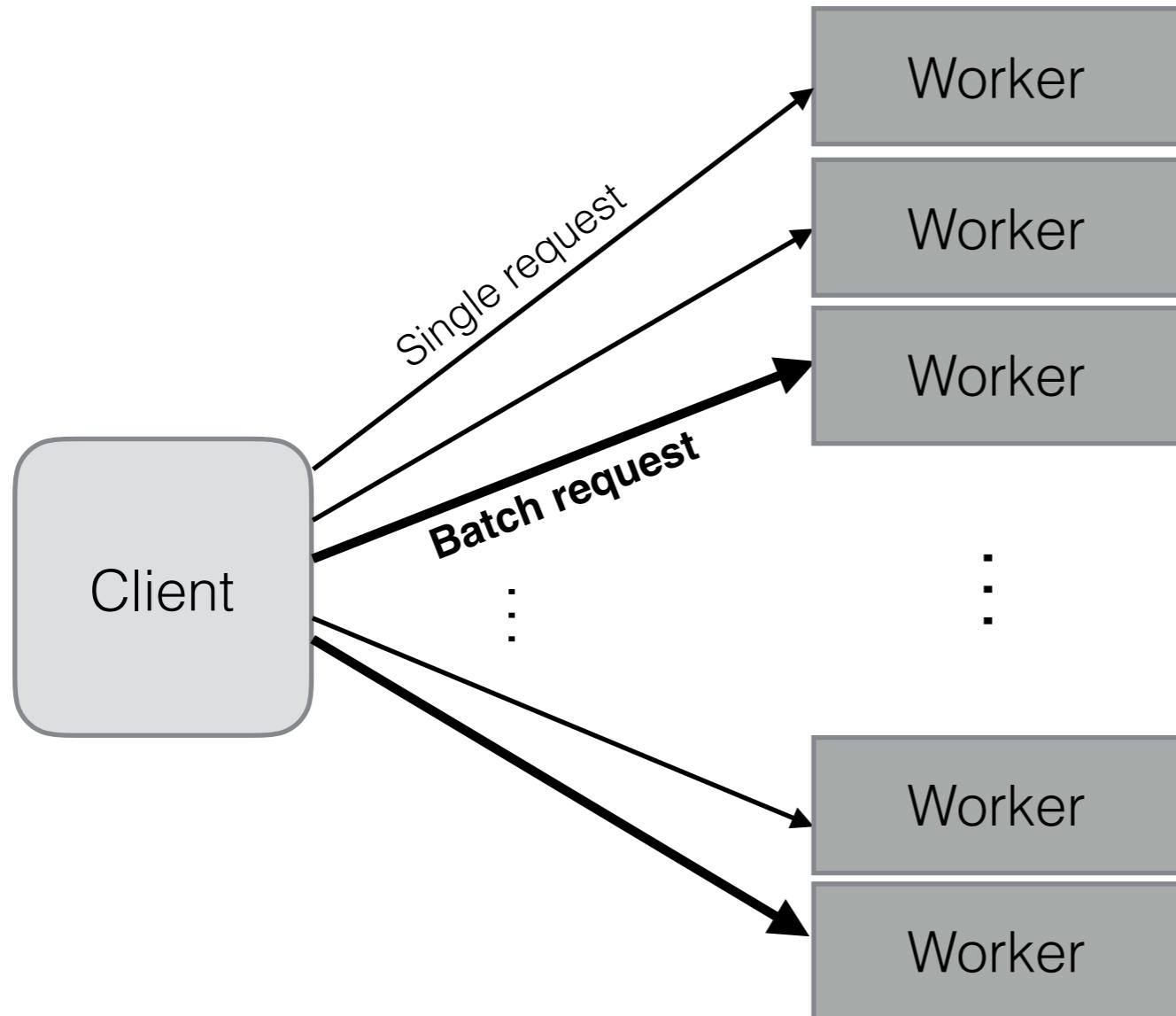
General architecture



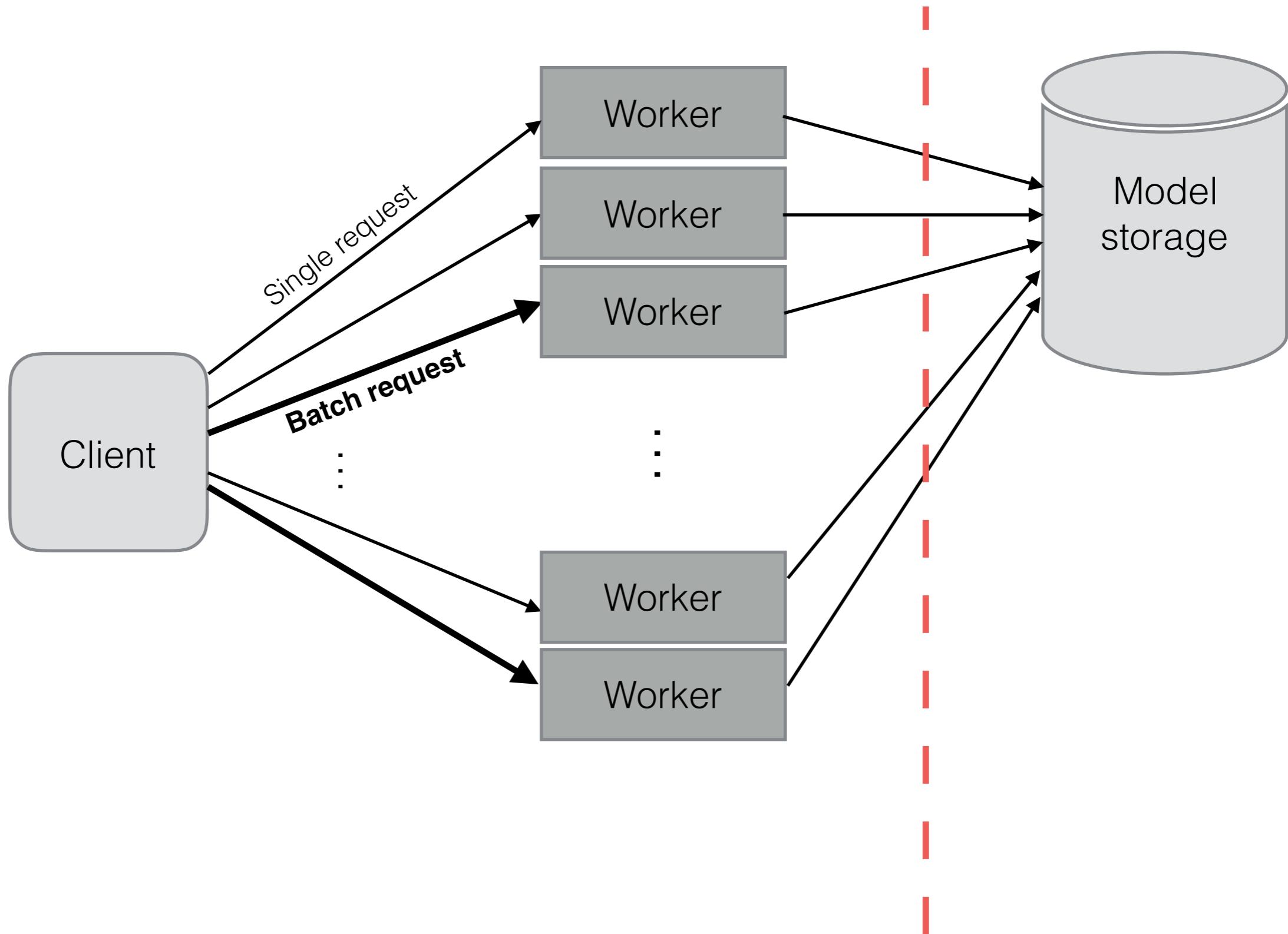
General architecture



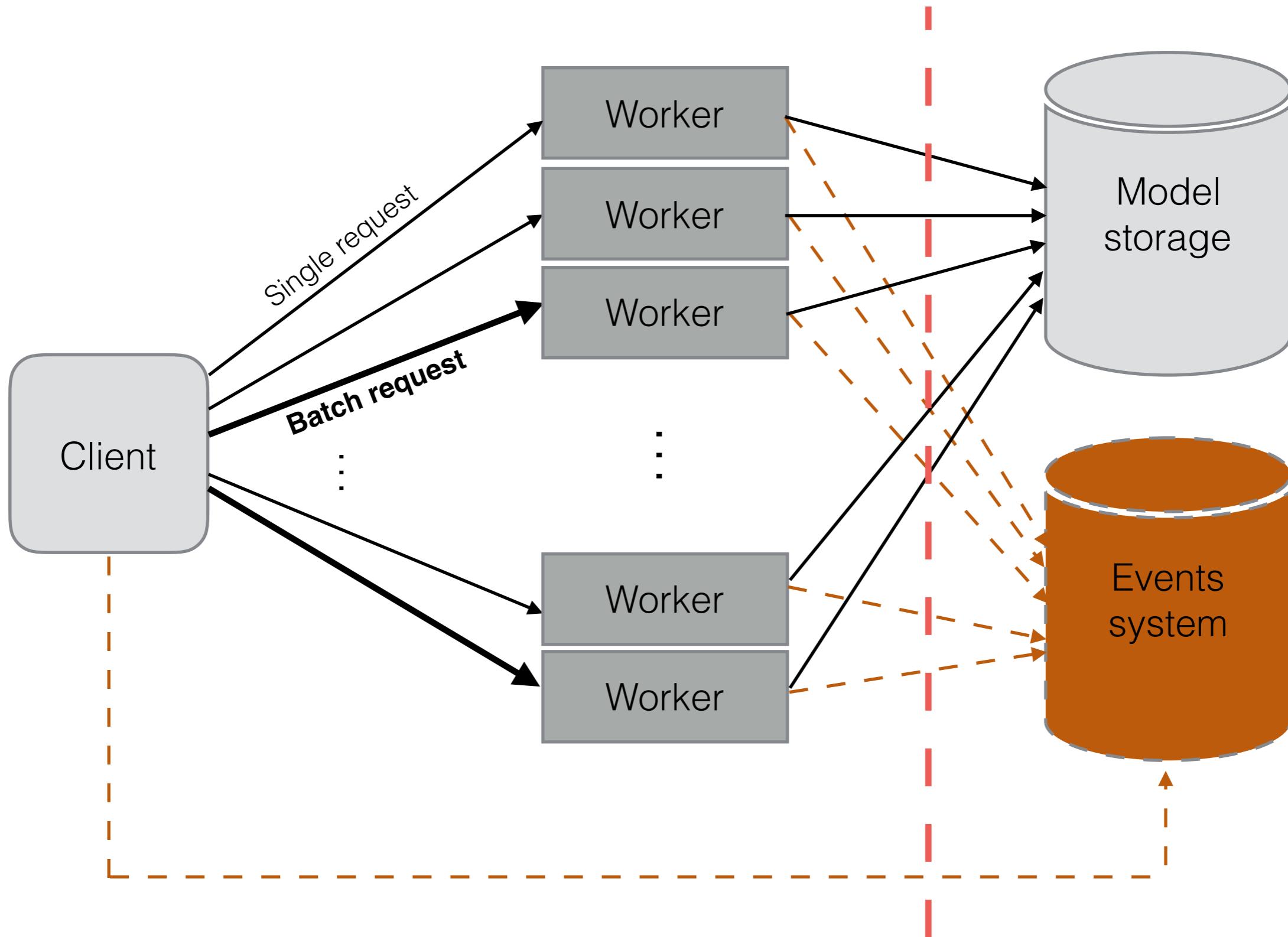
General architecture



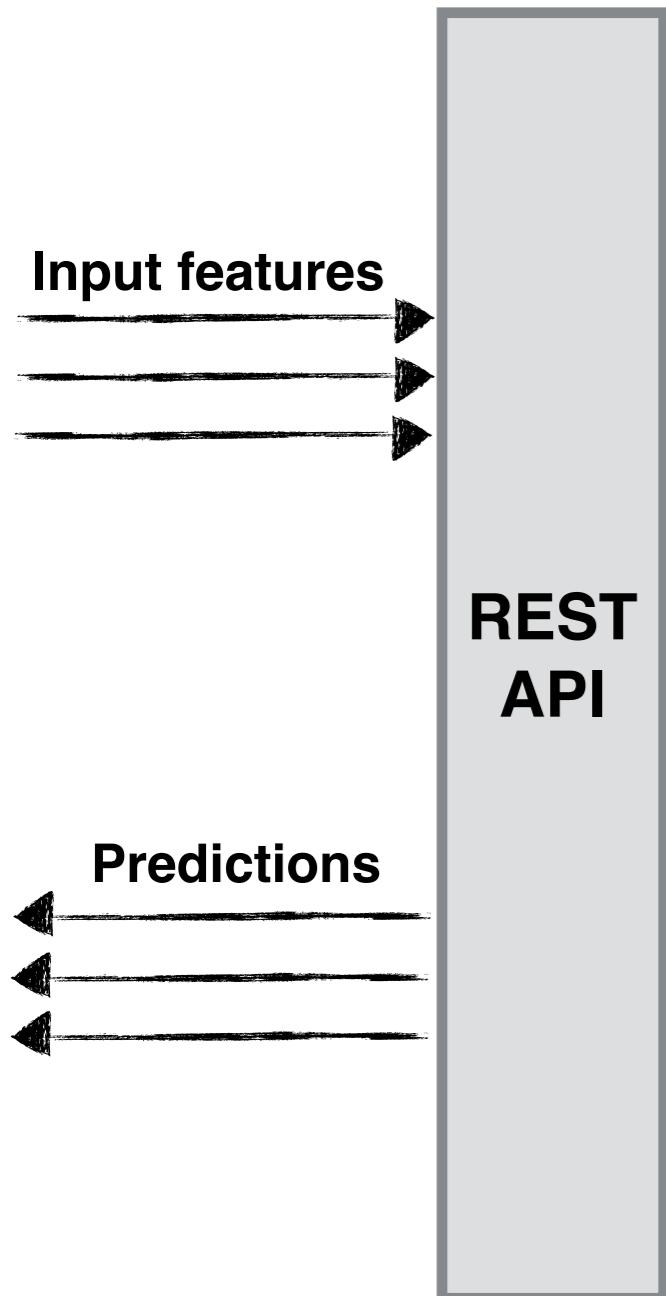
General architecture



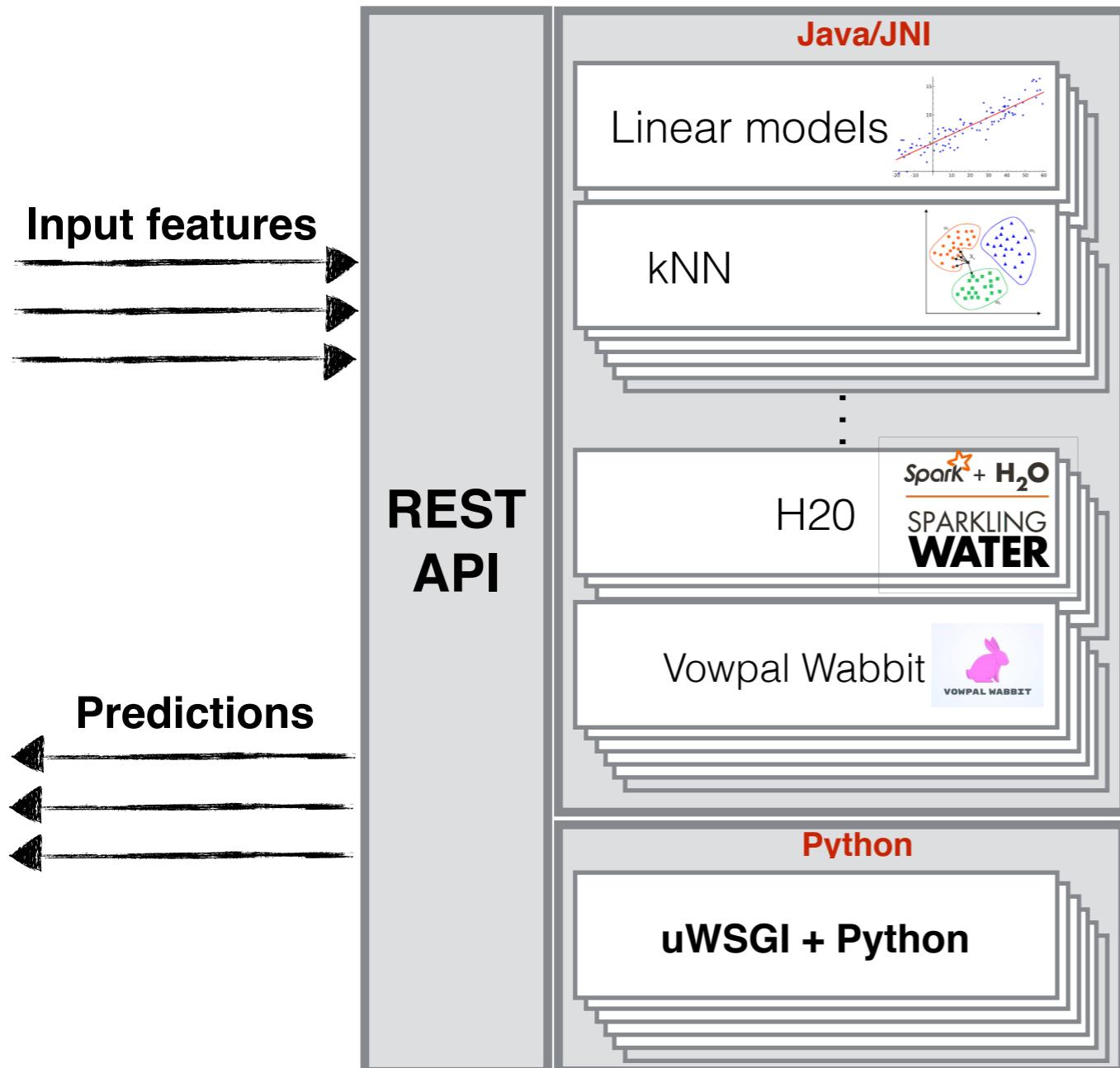
General architecture



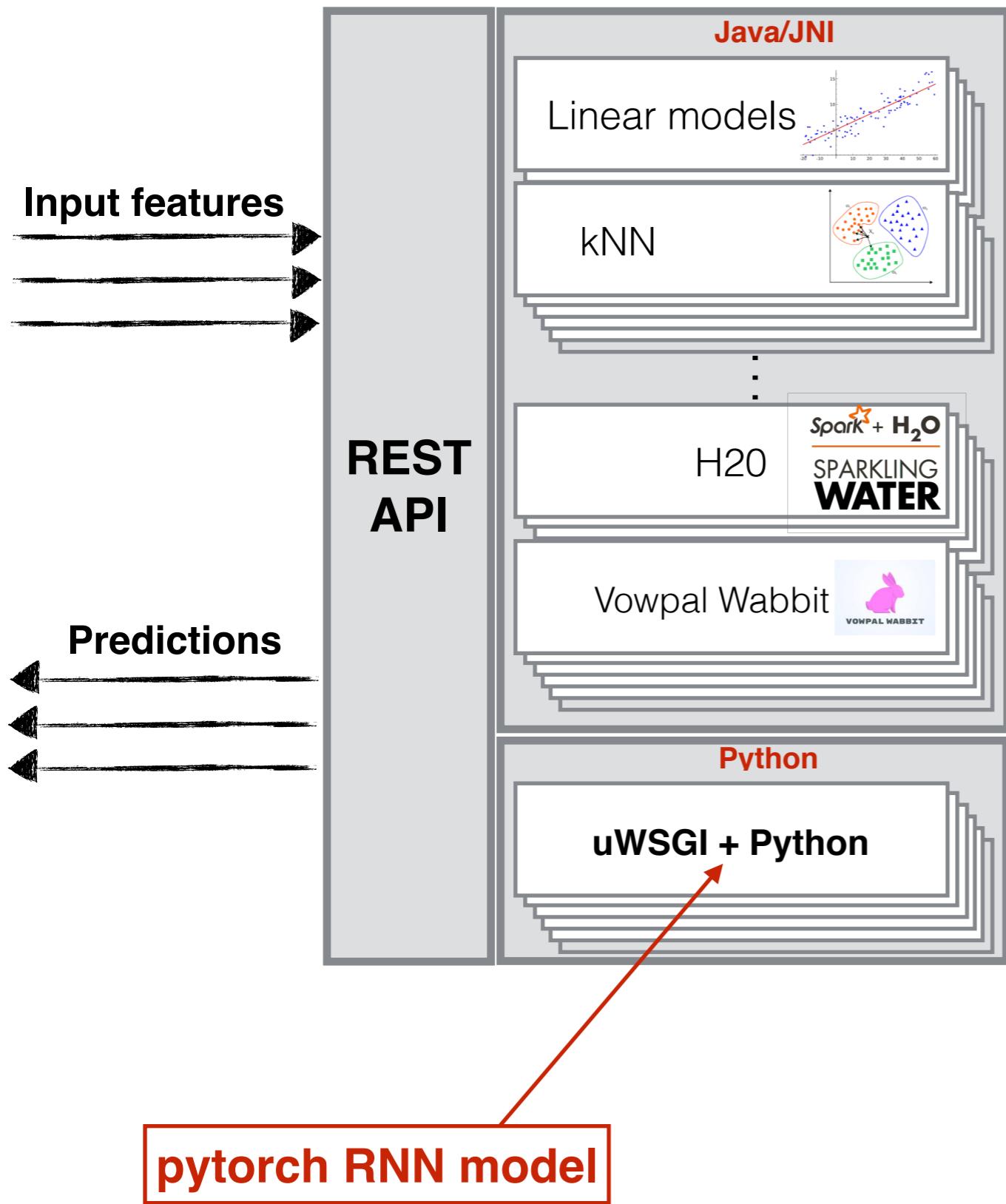
Worker structure



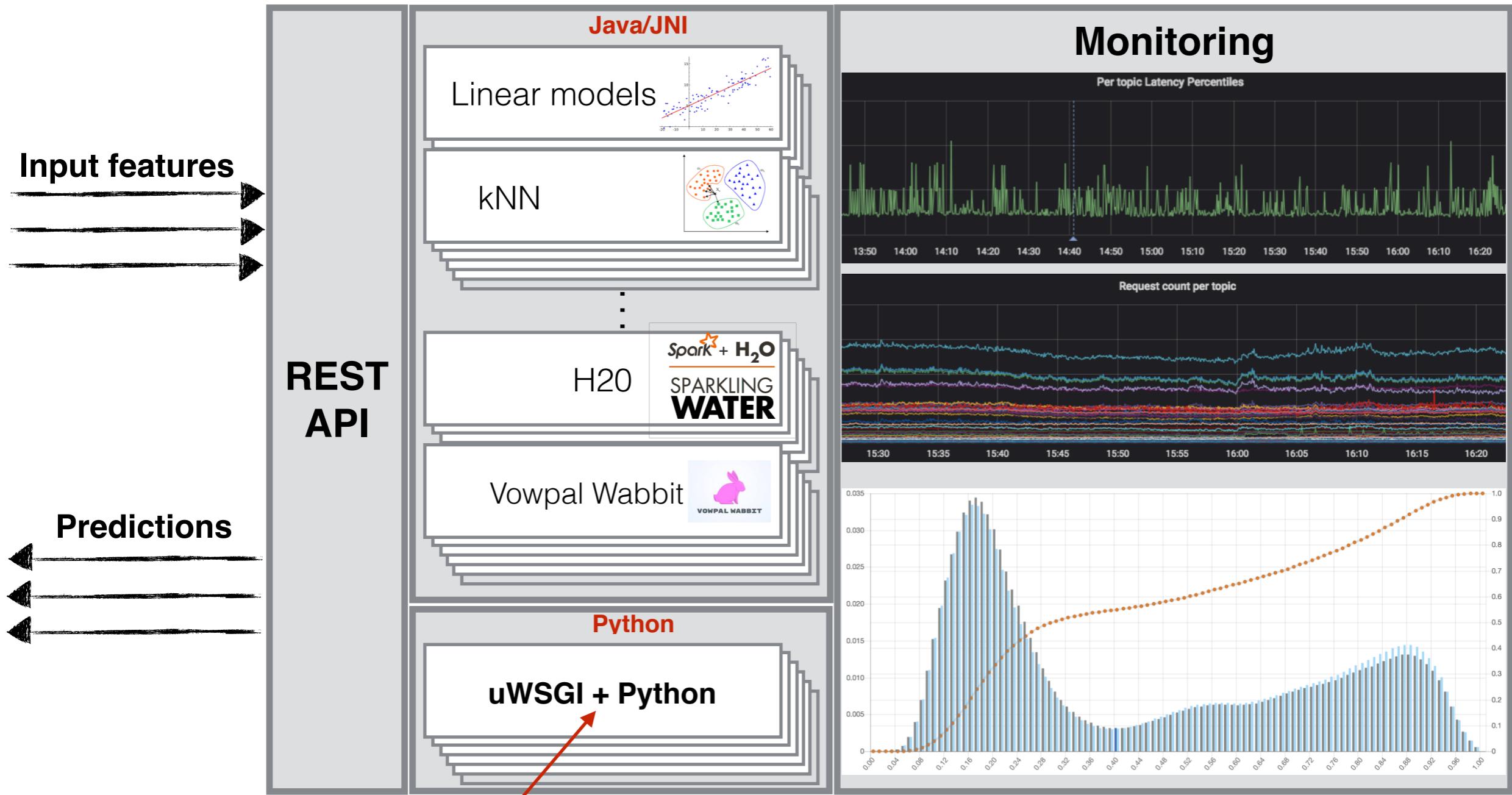
Worker structure



Worker structure



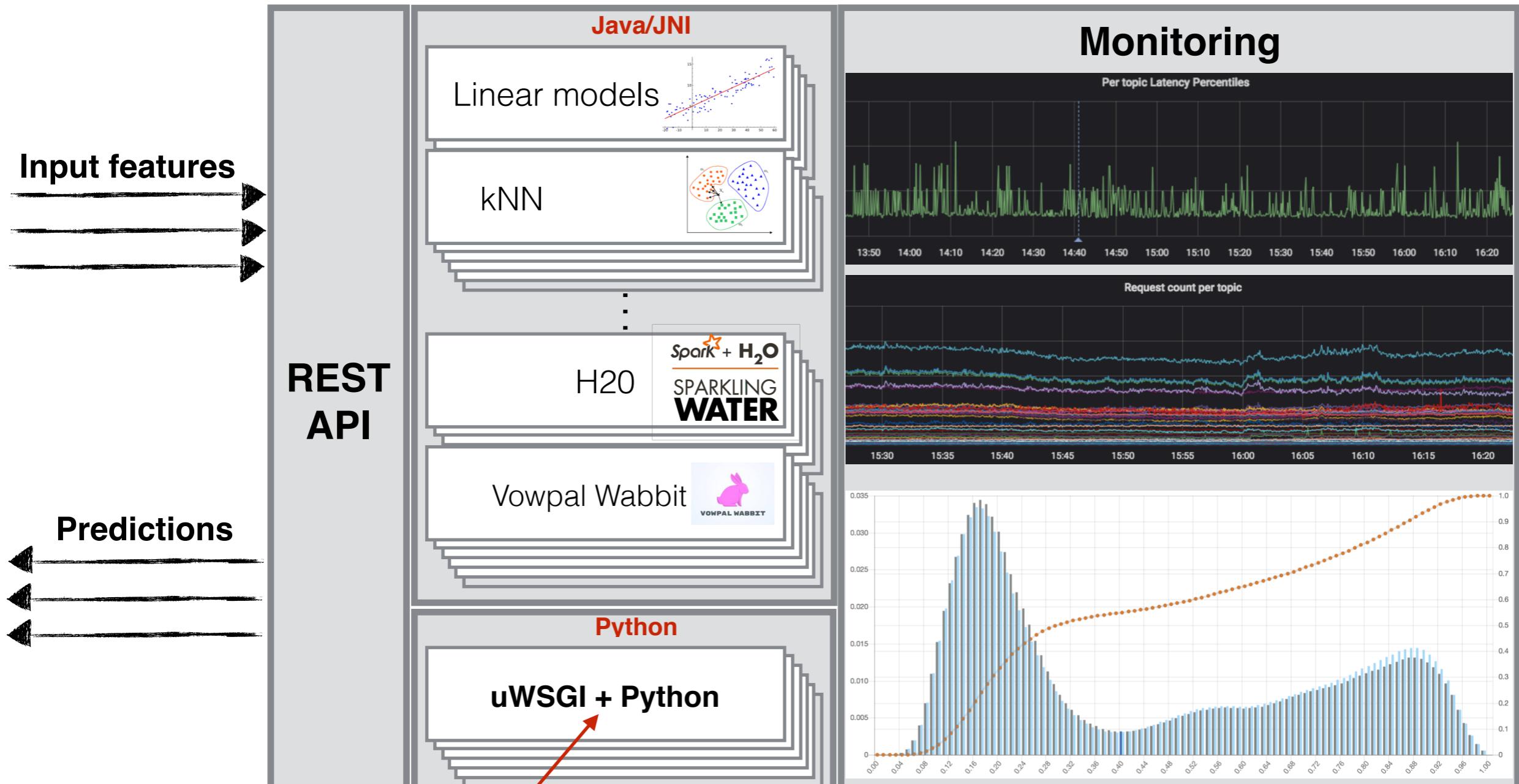
Worker structure



pytorch RNN model

Booking.com

Worker structure

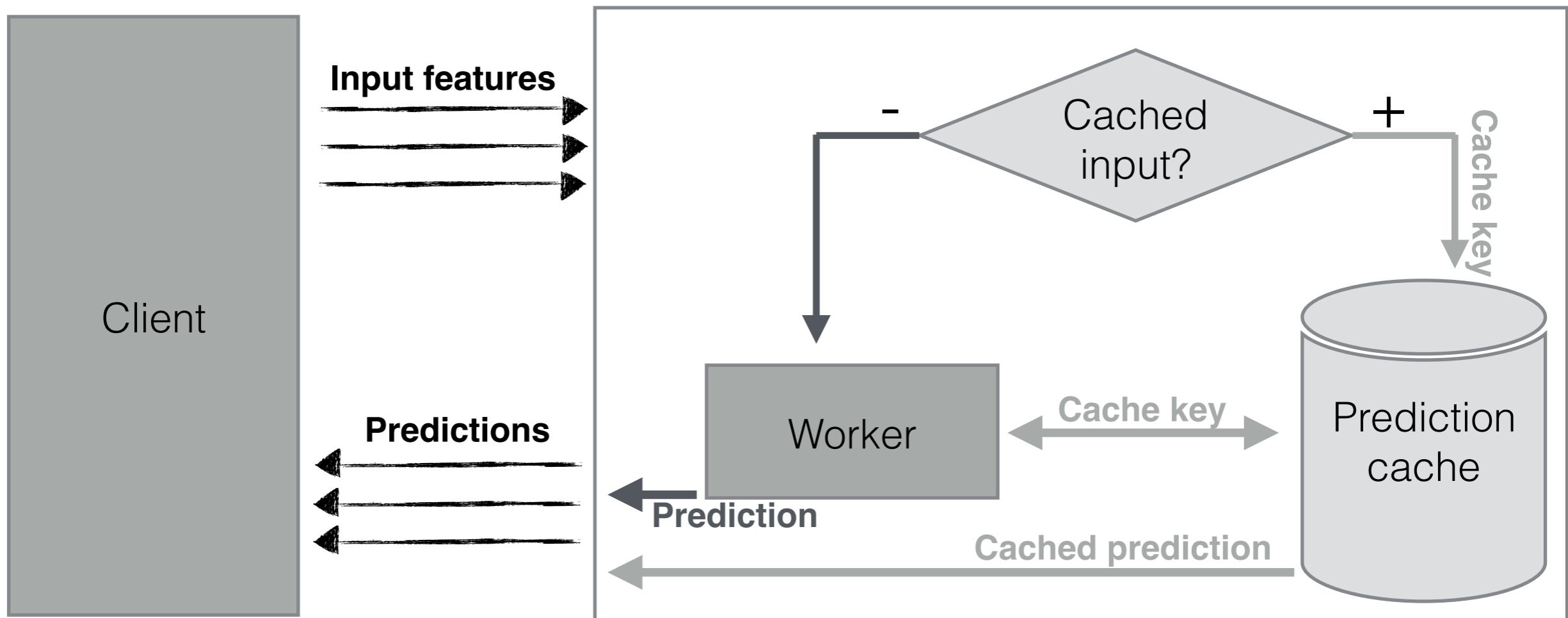


pytorch RNN model

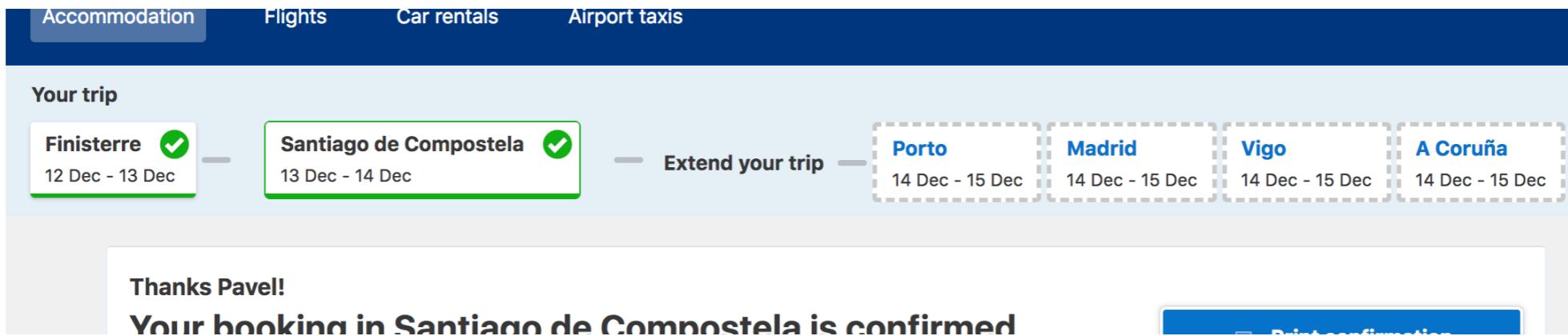


Booking.com

Caching common inputs



Some use cases



- Website: trip extension bar
- Email Marketing: personalized destination suggestions
- Mobile notifications [WIP]

Beyond single-step predictions

- Beam decoding for multiple steps
- Handling gaps properly
 - Sequence scoring
 - Direct modeling
- Additional user/trip/city-level features
- Hierarchical encoders to model multiple trips

Beyond single-step predictions

- Beam decoding for multiple steps
- Handling gaps properly
 - Sequence scoring
 - Direct modeling
- Additional user/trip/city-level features
- Hierarchical encoders to model multiple trips

e.g.

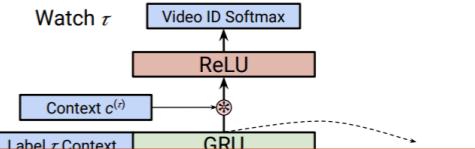
Latent Cross: Making Use of Context in Recurrent Recommender Systems

Alex Beutel, Paul Covington, Sagar Jain, Can Xu, Jia Li*, Vince Gatto, Ed H. Chi
Google, Inc.

Mountain View, California
{alexbeutel, pcovington, sagarj, canxu, vgatto, edchi}@google.com, vena900620@gmail.com

ABSTRACT

The success of recommender systems often depends on their ability to understand and make use of the context of the recommendation request. Significant research has focused on how time, location, interfaces, and a plethora of other contextual features affect recommendations. However, in using deep neural networks for recommender systems, researchers often ignore these contexts or



Beyond single-step predictions

- Beam decoding for multiple steps
- Handling gaps properly
 - Sequence scoring
 - Direct modeling
- Additional user/trip/city-level features
- Hierarchical encoders to model multiple trips

e.g. **Latent Cross: Making Use of Context in Recurrent Recommender Systems**

Alex Beutel, Paul Covington, Sagar Jain, Can Xu, Jia Li*, Vince Gatto, Ed H. Chi
Google, Inc.
Mountain View, California
{alexbeutel, pcovington, sagarj, canxu, vgatto, edchi}@google.com, vena900620@gmail.com

ABSTRACT

The success of recommendation systems depends on their ability to understand and make recommendations given a user's context and a query. Significant work has been done on improving the quality of recommendations. However, most of this work has focused on improving the quality of recommendations given a user's context and a query. In this paper, we propose a new approach to recommendation systems that takes into account the user's context and the query. We show that our approach can improve the quality of recommendations given a user's context and a query.

Watch → Video ID: Softmax

e.g. **A Hierarchical Neural Autoencoder for Paragraphs and Documents**

Jiwei Li, Minh-Thang Luong and Dan Jurafsky
Computer Science Department, Stanford University, Stanford, CA 94305, USA
jiweil, lmthang, jurafsky@stanford.edu

Abstract

Natural language generation of coherent long texts like news articles or longer doc-

2004; Elsner and Charniak, 2008; Li and Hovy, 2014, inter alia). However, applying these to text generation remains difficult. To understand how

Thank you!

We're hiring!

<https://workingatbooking.com>

<https://booking.ai>