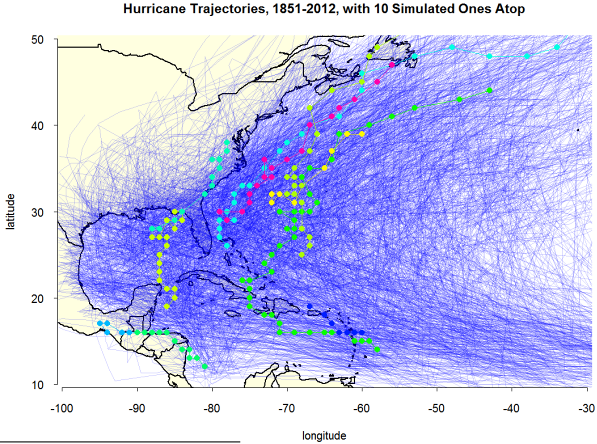
Towards Parallel Detection of Moving Flock Patterns in Large Spatiotemporal Datasets

Andres Calderon

December 1, 2016

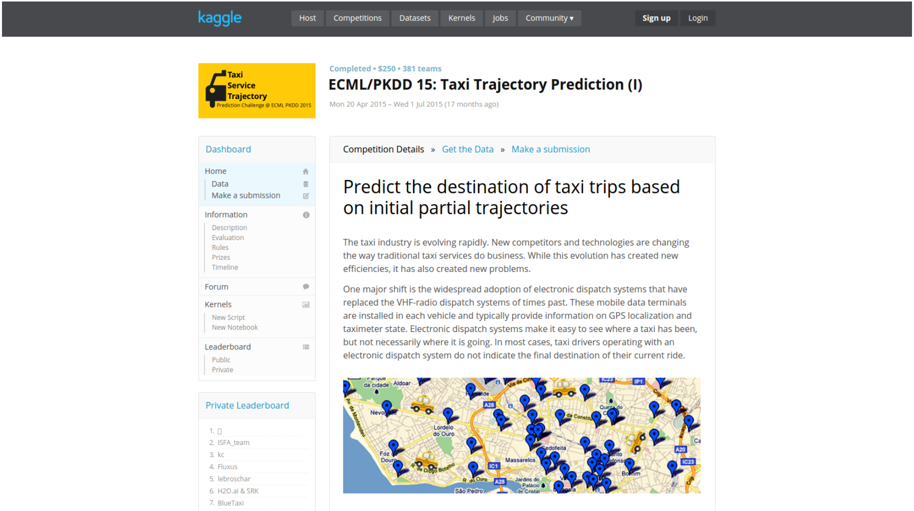
Trajectory Datasets

- Sensors, sensors everywhere...
    - Smart phones, GPS, RFID, WiFi, Bluetooth, IoT, Remote sensing...


Hurricane Trajectories, 1851-2012, with 10 Simulated Ones Atop
latitude
50
40
30
20
10
longitude
-100
-90
-80
-70
-60
-50
-40
-30

http://tinyurl.com/h4mbvxz

Applications



Completed • $250 • 381 teams

ECML/PKDD 15: Taxi Trajectory Prediction (I)

Mon 20 Apr 2015 – Wed 1 Jul 2015 (17 months ago)

1. ▯
2. ISFA_team
3. kc
4. Fluxus
5. lebroschar
6. H2O.ai & SRK
7. BlueTaxi

Competition Details » Get the Data » Make a submission

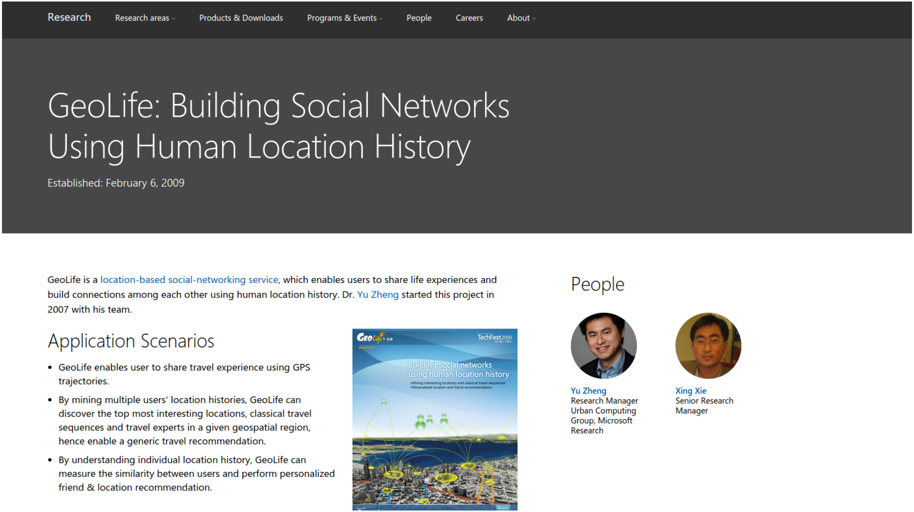
Predict the destination of taxi trips based on initial partial trajectories

The taxi industry is evolving rapidly. New competitors and technologies are changing the way traditional taxi services do business. While this evolution has created new efficiencies, it has also created new problems.

One major shift is the widespread adoption of electronic dispatch systems that have replaced the VHF-radio dispatch systems of times past. These mobile data terminals are installed in each vehicle and typically provide information on GPS localization and taximeter state. Electronic dispatch systems make it easy to see where a taxi has been, but not necessarily where it is going. In most cases, taxi drivers operating with an electronic dispatch system do not indicate the final destination of their current ride.

http://tinyurl.com/jfm8qfu

Applications



Research   Research areas   Products & Downloads   Programs & Events   People   Careers   About

GeoLife: Building Social Networks Using Human Location History

Established: February 6, 2009

GeoLife is a location-based social-networking service, which enables users to share life experiences and build connections among each other using human location history. Dr. Yu Zheng started this project in 2007 with his team.

Application Scenarios

- GeoLife enables user to share travel experience using GPS trajectories.
- By mining multiple users' location histories, GeoLife can discover the top most interesting locations, classical travel sequences and travel experts in a given geospatial region, hence enable a generic travel recommendation.
- By understanding individual location history, GeoLife can measure the similarity between users and perform personalized friend & location recommendation.

GeoLife 2.0
TechFest 2009
Building social networks using human location history
•Mining interesting locations and classical travel sequences
•Personalized location and friend recommendation

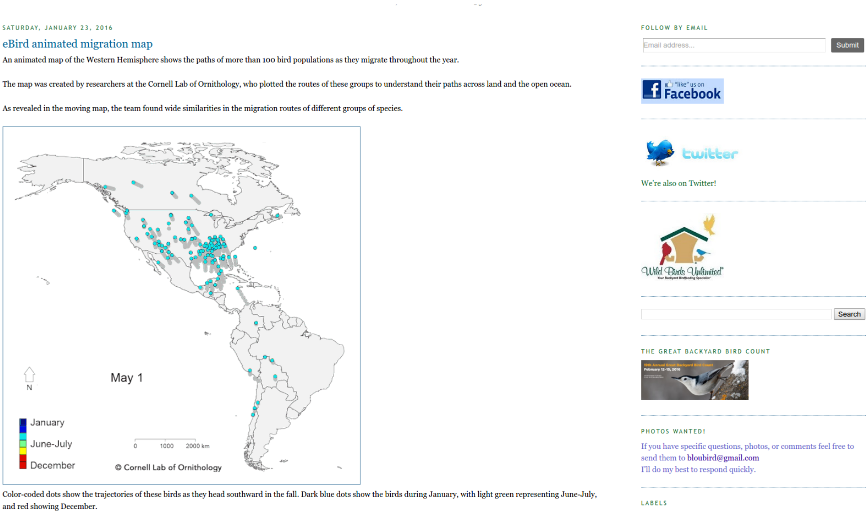
People

Yu Zheng
Research Manager
Urban Computing Group, Microsoft Research

Xing Xie
Senior Research Manager

http://tinyurl.com/hpd4nxl

Applications



SATURDAY, JANUARY 23, 2016

eBird animated migration map

An animated map of the Western Hemisphere shows the paths of more than 100 bird populations as they migrate throughout the year.

The map was created by researchers at the Cornell Lab of Ornithology, who plotted the routes of these groups to understand their paths across land and the open ocean.

As revealed in the moving map, the team found wide similarities in the migration routes of different groups of species.

Color-coded dots show the trajectories of these birds as they head southward in the fall. Dark blue dots show the birds during January, with light green representing June-July, and red showing December.

FOLLOW BY EMAIL

We're also on Twitter!

THE GREAT BACKYARD BIRD COUNT

PHOTOS WANTED!

If you have specific questions, photos, or comments feel free to send them to bloubird@gmail.com
I'll do my best to respond quickly.

LABELS

http://tinyurl.com/hkc6ahl

Outline

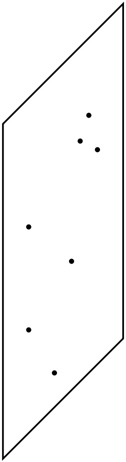
1. Moving Flock Patterns
2. Implementation
3. Experiments
4. Conclusions

What is a flock???

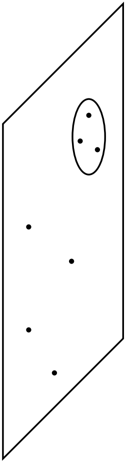
Definition ($(\mu, \epsilon, \delta) - flock$)

Sets of at least μ objects moving close enough (ε) for at least δ time intervals (Benkert et al, 2008).

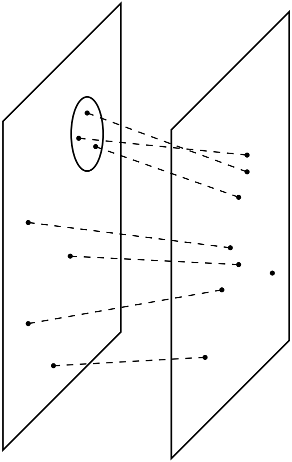
BFE algorithm (Vieira et al, 2009)



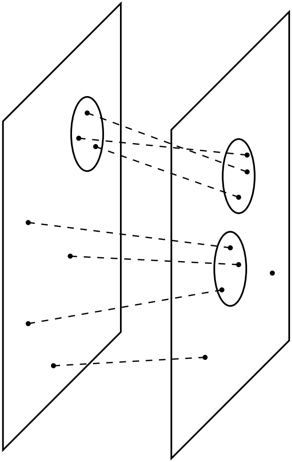
BFE algorithm (Vieira et al, 2009)



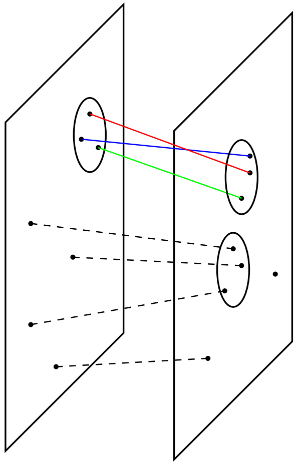
BFE algorithm (Vieira et al, 2009)



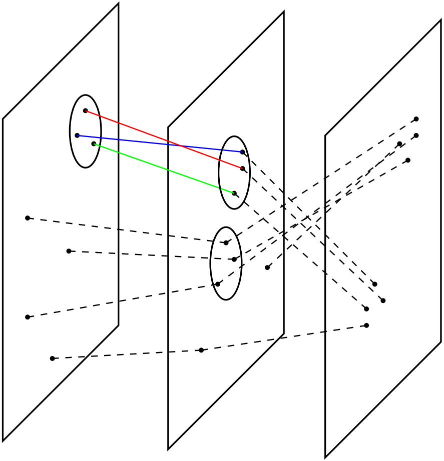
BFE algorithm (Vieira et al, 2009)



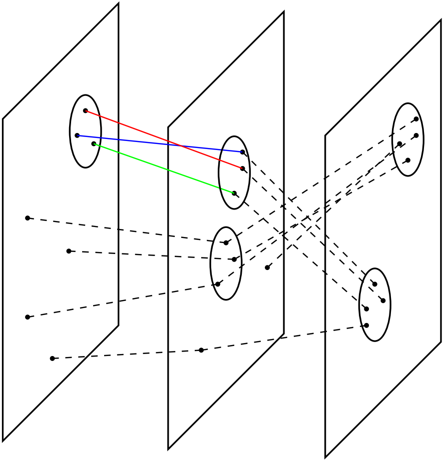
BFE algorithm (Vieira et al, 2009)



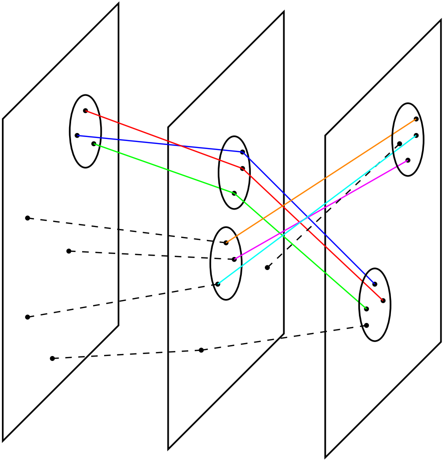
BFE algorithm (Vieira et al, 2009)



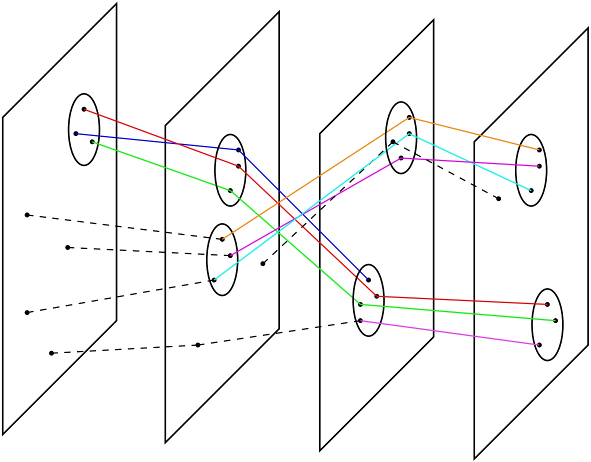
BFE algorithm (Vieira et al, 2009)



BFE algorithm (Vieira et al, 2009)



BFE algorithm (Vieira et al, 2009)



Why am I doing this???

- Why are moving flock patterns important?
    - They capture the collective behavior of trajectories as groups.
- Why is the finding of disks important?
    - It is the base of the algorithm but it has a high complexity ($\mathcal{O}(2n^2)$).
    - It is no trivial, disks can be at any location.

Outline

1. Moving Flock Patterns

2. Implementation

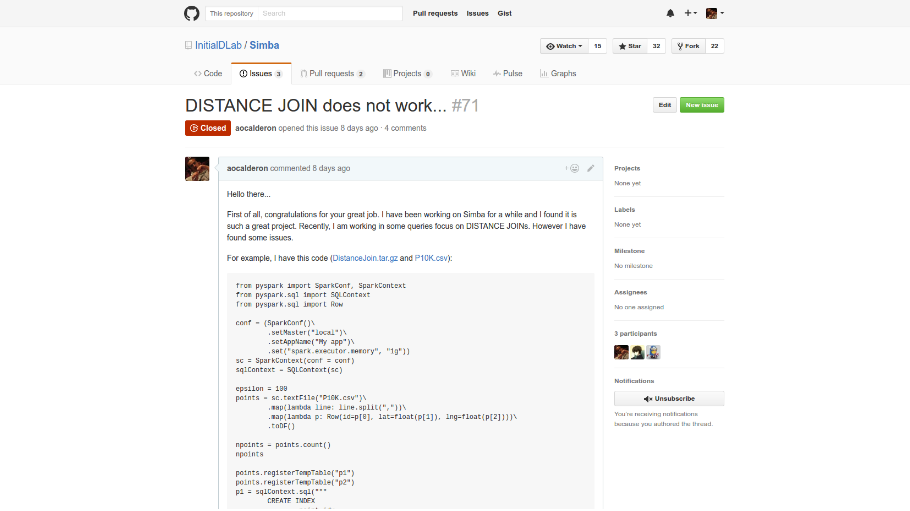
3. Experiments

4. Conclusions

Demo

- Demo time:
  - `http://tinyurl.com/jl55849`.

Bug report



DISTANCE JOIN does not work... #71

Closed **aocalderon** opened this issue 8 days ago · 4 comments

aocalderon commented 8 days ago

Hello there...

First of all, congratulations for your great job. I have been working on Simba for a while and I found it is such a great project. Recently, I am working in some queries focus on DISTANCE JOINs. However I have found some issues.

For example, I have this code (DistanceJoin.tar.gz and P10K.csv):

```
from pyspark import SparkConf, SparkContext
from pyspark.sql import SQLContext
from pyspark.sql import Row

conf = (SparkConf()\
        .setMaster("local")\
        .setAppName("My app")\
        .set("spark.executor.memory", "1g"))
sc = SparkContext(conf = conf)
sqlContext = SQLContext(sc)

epsilon = 100
points = sc.textFile("P10K.csv")\
        .map(lambda line: line.split(","))\
        .map(lambda p: Row(id=p[0], lat=float(p[1]), lng=float(p[2])))\
        .toDF()

npoints = points.count()
npoints

points.registerTempTable("p1")
points.registerTempTable("p2")
p1 = sqlContext.sql("""
        CREATE INDEX
```

Projects

None yet

Labels

None yet

Milestone

No milestone

Assignees

No one assigned

3 participants

Notifications

Unsubscribe

You're receiving notifications because you authored the thread.

https://github.com/InitialDLab/Simba/issues/71

Outline

1 Moving Flock Patterns

2 Implementation

3 Experiments

4 Conclusions

Dataset

- **Beijing** from Geolife project[1].
  - 182 users in a period of over three years (from April 2007 to August 2012).
  - 17,621 trajectories.
  - ≈18 million points (no duplicates).

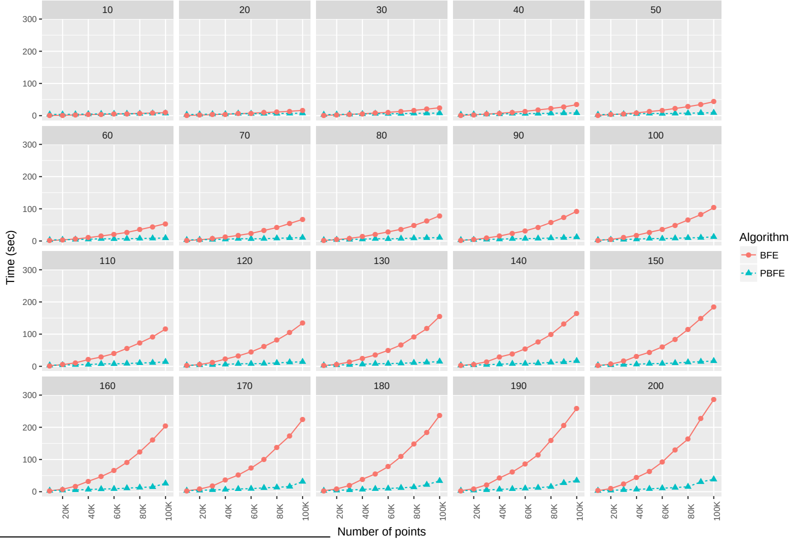
[1] http://tinyurl.com/j7t2cao

Setup

- Single node.
- Processor: 4-core Intel(R) Core(TM) i5-2400S CPU @ 2.50GHz
- RAM: 8 GB.
- Ubuntu 16.04 LTS, Simba/Spark 1.6.0.

Beijing [N = 10K - 100K; ε = 10 - 200 (mts)]

Execution time by ε (radius of disk in mts) in Beijing dataset.



http://tinyurl.com/js6us8g

Dataset

- **Porto** from ECML/PKDD 15 Taxi Trajectory Prediction Challenge[2].
    - A complete year (from 01/07/2013 to 30/06/2014).
    - Trajectories for all the 442 taxis running in the city of Porto, in Portugal.
    - ≈17.7 million points (no duplicates).

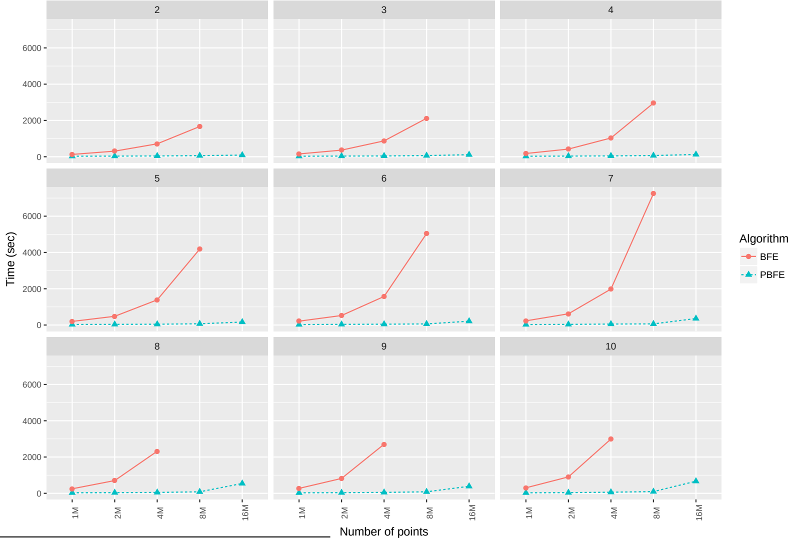
[2] http://tinyurl.com/zzbtlt9

Setup

- 4-node cluster at DBLab.
- Processors: 8-core Intel(R) Xeon(R) CPU E3-1230 V2 @ 3.30GHz
- RAM: 15.5 GB.
- Centos 6.8, Simba/Spark 1.6.0.

Porto [N = 1M - 16M; ε = 2 - 10 (mts)]

Execution time by ε (radius of disk in mts) in Porto dataset.



http://tinyurl.com/j9u9c7h

Outline

1. Moving Flock Patterns
2. Implementation
3. Experiments
4. Conclusions

Conclusions

Cooming soon...

Thank you!!!

Do you have any question?