Learning Networks of Places and People Using Location Data

Tony Jebara

Sense Networks

Columbia University
WWW

A network of online places
facebook

A network of online people
From online to real networks?

What’s next?

- a network of real places
- a network of real people

Online data is easy to get, what about the real world?
From online to real networks?

What’s next?

GPS

Online data is easy to get, what about the real world?
GPS and location data

Collaborative Filtering  
Marketing  
Advertising  
Search  
Social Recommendation

SENSE NETWORKS
ANALYSIS, CITYSENSE, NETWORKS OF PLACES & PEOPLE

GPS  
VEHICLES  
APPS  
MAPS  
CARRIERS
GPS and location data
CitySense: where is everyone

- Citysense: real-time density of users at every street corner
- Poisson models find most active bars/restaurants
Next: where’s everyone like me

Need to have a network of people

Each dot is a user

Dot’s color is user’s social cluster
Network of People
who is like whom? who collocates with whom?
Network of People

Hard to say if User A is like User B...

... don’t just see if they colocate physically
... do they overlap semantically (network of places)
Network of Places

Is place A like place B?
Look at each place’s Flow, Commerce & Demographics
Network of Places: Flow

Look at flow A to B

Markov transition

Minimum Volume Embedding (MVE)

Color code clusters
Network of Places: Commerce

Get each block’s SIC (standard industrial categorization) code & cluster
Network of Places: Demographics

Get each block’s census demographic data & cluster

Block ID: 60750302021
1264 8th Ave
San Francisco, California 94122
Psychographic cluster: Urban Achievers

- Higher Population Density: (38250;1511)
- Smaller Household Size: (2;76)
- Higher Household Income: ($58,000.00;143)
- Higher Household Expenditures: $49,288.00;123
- Underindex: Black: (10;10)
## Encoding people

For each user, convert GPS trail into matrix of probabilities for week hour probability of being in:

1) flow cluster
2) sic cluster
3) demographic cluster

| Week Hour | FLO 1 | FLO 2 | ... | FLO 20 | SIC 1 | SIC 2 | ... | SIC 97 | DEM 1 | DEM 2 | ... | DEM 78 |
|-----------|-------|-------|-----|--------|-------|-------|-----|-------|-------|-------|-----|-------|-------|
| 1         | .03   | .31   |     | .14    | .03   | .05   |     | .41   | .11   | .04   |     | .01   |       |
| 2         | .14   | .34   |     | .02    | .04   | .05   |     | .52   | .01   | .01   |     | .00   |       |
| ...       |       |       |     |        |       |       |     |       |       |       |     |       |       |
| 168       | .07   | .34   |     | .51    | .02   | .06   |     | .48   | .02   | .01   |     | .00   |       |
Encoding people

9 example users
compute pair-wise overlap from weekly exposure matrices
real friends should colocate!
Connecting Similar People
Network of People

Churn
Advertising
Marketing
Collaborative Filtering
Demographics
Network of People → Tribes

How often do they go out each day of the week?
- "Young & Edgy" - Out every night in young, racially diverse, low income neighborhoods
- "Weekend Mole" - Out occasionally on weeknights, typically middle-aged, Latino, middle-income neighborhoods
- "Mature Homebody" - Rarely goes out, typically spends nights in mature, white, higher income neighborhoods

Where do they hang out?

What is the avg age of most people in the neighborhoods they spend time in?

How racially diverse are the neighborhoods they spend time in?

Are the places they spend time in rich neighborhoods or poor neighborhoods?
Network of People Predictions

Percent Improvement

- baseline
- allTimeOfDayUsageFtrs
- sicDemogTravel

Percent Increase vs. Baseline

Number of Impressions

0 5000 10000 15000
The Next Net

(Stephen Baker, BusinessWeek)

Indexing the real world using location data for predictive analytics.