Finding Your Friends and Following Them to Where You Are

Adam Sadilek
Henry Kautz
Jeffrey P. Bigham
There was a man dancing on the train. He kept swinging his bookbag. Now I have a bruise on my leg. -_-
There was a man dancing on the train. He kept swinging his bookbag. Now I have a bruise on my leg. -_-
The Data

- New York City & Los Angeles
The Data

• New York City & Los Angeles
• 26M tweets over a month
The Data

- **New York City & Los Angeles**
- **26M tweets** over a month
- **1.2M unique users**
The Data

• **New York City & Los Angeles**
• **26M tweets** over a month
  • 1.2M unique **users**
• **7.6M geo-tagged tweets**
The Data

- 11K geo-active users
The Data

- 11K geo-active users
- 4M tweets by geo-active users
The Data

- 11K geo-active users
- 4M tweets by geo-active users
- 123K “follows” relationships
The Data

- 11K geoc-active users
- 4M tweets by geo-active users
- 123K “follows” relationships
- 52K “friends” relationships
Predicting Social Ties
Predicting Social Ties
Predicting Social Ties

Location

Text

Twitter Feed

© 2012 Adam Sadilek
Predicting Social Ties

CL = 87
TS = 75

CL = 64
TS = 50

CL = 100
TS = 96

Location

Text

Twitter Feed

© 2012 Adam Sadilek
Predicting Social Ties

CL = 87
TS = 75

CL = 64
TS = 50

CL = 100
TS = 96

Learn Decision Tree

Belief Propagation

Predicted Friendships

Location

Text

Twitter Feed

© 2012 Adam Sadilek
Predicting Social Ties

CL = 87
TS = 75

CL = 64
TS = 50

CL = 100
TS = 96

Learn Decision Tree
Belief Propagation
Predicted Friendships

Location
Text
Network Structure

Twitter Feed

© 2012 Adam Sadilek
Predicting Social Ties

Learn Decision Tree
Belief Propagation
Predicted Friendships

Location
Text
Network Structure
Twitter Feed

CL = 87
TS = 75

CL = 64
TS = 50

CL = 100
TS = 96

© 2012 Adam Sadilek
Predicting Social Ties

Learn Decision Tree
Belief Propagation
Predicted Friendships

Location
Text
Network Structure

Twitter Feed

CL = 87
TS = 75

CL = 64
TS = 50

CL = 100
TS = 96

0.7

0.8

0.9

© 2012 Adam Sadilek
Leveraging Structure
Evaluation

- Cross-validation
  - New York City ➔ Los Angeles
  - Los Angeles ➔ New York City
Predicting Social Ties

© 2012 Adam Sadilek
Predicting Location
Predicting Location

- Friends ~ noisy “sensors” of one’s location

© 2012 Adam Sadilek
Predicting Location

• Find significant locations
Predicting Location

- Find significant locations
- 90K with 100m granularity
Predicting Location

- Find significant locations
- 90K with 100m granularity
- 26K visited >4 times

© 2012 Adam Sadilek
Evaluation
Evaluation

- Cross-validation over all users
- Train on first 3 weeks, test on the 4th
Predicting Location

© 2012 Adam Sadilek
Summary

• Predicting **social ties** and **location**
Summary

• Predicting social ties and location
  – Large scale
Summary

- Predicting **social ties** and **location**
  - Large scale
- Evaluation on Twitter data from NYC and LA
Future Work

• **Unify** our two approaches
• **Fine-grained** modeling
  • General
  • Fashion, gossip, epidemics
Finding Your Friends and Following Them to Where You Are

Adam Sadilek
Henry Kautz
Jeffrey P. Bigham
\[ y_{1:t}^* = \arg\max_{y_{1:t}} \log \left( \Pr(y_{1:t} | x_{1:t}) \right) \]

\[ \theta^* = \arg\max_\theta \log \left( \Pr(x_{1:t}, y_{1:t} | \theta) \right) \]

\[ \theta^* = \arg\max_\theta \log \sum_{y_{1:t}} \Pr(x_{1:t}, y_{1:t} | \theta) \]