13.5 Supervised Random Walk Case Studies

DATASCI W261

Machine Learning at Scale

Case Study: Coauthorship Networks

Case Study: Coauthorship Networks

- From arXiv e-print archive

Case Study: Coauthorship Networks

- From arXiv e-print archive
- Contains time-stamped papers with authors' names and titles from 1992–2002

Case Study: Coauthorship Networks

- From arXiv e-print archive
- Contains time-stamped papers with authors' names and titles from 1992–2002
- Selected from four areas of physics

Results: Predicting Coauthorship

Learning Method	AUC	Prec@20
Random Walk With Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node + Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node + Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: One edge type	0.69996	4.24
SRW: Multiple edge types	0.71238	4.25

Table 2. Hep-Ph coauthorship network. DT: decision tree, LR: logistic regression, and SRW: supervised random walks.

Results: Predicting Coauthorship

Learning Method	AUC	Prec@20
Random Walk With Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node + Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node + Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: One edge type	0.69996	4.24
SRW: Multiple edge types	0.71238	4.25

Table 2. Hep-Ph coauthorship network. DT: decision tree, LR: logistic regression, and SRW: supervised random walks.

Results: Predicting Coauthorship

Learning Method	AUC	Prec@20
Random Walk With Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node + Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node + Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: One edge type	0.69996	4.24
SRW: Multiple edge types	0.71238	4.25

Table 2. Hep-Ph coauthorship network. DT: decision tree, LR: logistic regression, and SRW: supervised random walks.

Results: Predicting Coauthorship

Learning Method	AUC	Prec@20
Random Walk With Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node + Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node + Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: One edge type	0.69996	4.24
SRW: Multiple edge types	0.71238	4.25

Table 2. Hep-Ph coauthorship network. DT: decision tree, LR: logistic regression, and SRW: supervised random walks.

Results: Predicting Coauthorship

Learning Method	AUC	Prec@20
Random Walk With Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node + Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node + Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: One edge type	0.69996	4.24
SRW: Multiple edge types	0.71238	4.25

Table 2. Hep-Ph coauthorship network. DT: decision tree, LR: logistic regression, and SRW: supervised random walks.

Results: Predicting Coauthorship

Learning Method	AUC	Prec@20
Random Walk With Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node + Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node + Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: One edge type	0.69996	4.24
SRW: Multiple edge types	0.71238	4.25

Table 2. Hep-Ph coauthorship network. DT: decision tree, LR: logistic regression, and SRW: supervised random walks.

Results: Predicting Coauthorship

Learning Method	AUC	Prec@20
Random Walk With Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node + Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node + Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: One edge type	0.69996	4.24
SRW: Multiple edge types	0.71238	4.25

Table 2. Hep-Ph coauthorship network. DT: decision tree, LR: logistic regression, and SRW: supervised random walks.

Results: Predicting Coauthorship

Learning Method	AUC	Prec@20
Random Walk With Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node + Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node + Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: One edge type	0.69996	4.24
SRW: Multiple edge types	0.71238	4.25

Table 2. Hep-Ph coauthorship network. DT: decision tree, LR: logistic regression, and SRW: supervised random walks.

Results: Predicting Coauthorship

Learning Method	AUC	Prec@20
Random Walk With Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node + Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node + Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: One edge type	0.69996	4.24
SRW: Multiple edge types	0.71238	4.25

Table 2. Hep-Ph coauthorship network. DT: decision tree, LR: logistic regression, and SRW: supervised random walks.

Results: Predicting Coauthorship

Learning Method	AUC	Prec@20
Random Walk With Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node + Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node + Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: One edge type	0.69996	4.24
SRW: Multiple edge types	0.71238	4.25

Table 2. Hep-Ph coauthorship network. DT: decision tree, LR: logistic regression, and SRW: supervised random walks.

Results: Predicting Coauthorship

Learning Method	AUC	Prec@20
Random Walk With Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node + Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node + Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: One edge type	0.69996	4.24
SRW: Multiple edge types	0.71238	4.25

Table 2. Hep-Ph coauthorship network. DT: decision tree, LR: logistic regression, and SRW: supervised random walks.

Out of 20 authors recommended, over 20% (4.25) happen in the near future.

Case Study: Facebook

Case Study: Facebook

- First study: 100 individuals for training, 100 for testing

Case Study: Facebook

- First study: 100 individuals for training, 100 for testing
- Supervised random walk trained using 100 training examples

Case Study: Facebook

- First study: 100 individuals for training, 100 for testing
- Supervised random walk trained using 100 training examples
- Evaluated on 100 test users

Results: Facebook Study

Learning Method	AUC	Prec@20
Random Walk With Restart	0.81725	6.80
Degree	0.58535	3.25
DT: Node features	0.59248	2.38
DT: Path features	0.62836	2.46
DT: All features	0.72986	5.34
LR: Node features	0.54134	1.38
LR: Path features	0.51418	0.74
LR: All features	0.81681	7.52
SRW: One edge type	0.82502	6.87
SRW: Multiple edge types	0.82799	7.57

Table 3. Results for the Facebook data set.

Results: Facebook Study

Learning Method	AUC	Prec@20
Random Walk With Restart	0.81725	6.80
Degree	0.58535	3.25
DT: Node features	0.59248	2.38
DT: Path features	0.62836	2.46
DT: All features	0.72986	5.34
LR: Node features	0.54134	1.38
LR: Path features	0.51418	0.74
LR: All features	0.81681	7.52
SRW: One edge type	0.82502	6.87
SRW: Multiple edge types	0.82799	7.57

Table 3. Results for the Facebook data set.

Results: Facebook Study

Learning Method	AUC	Prec@20
Random Walk With Restart	0.81725	6.80
Degree	0.58535	3.25
DT: Node features	0.59248	2.38
DT: Path features	0.62836	2.46
DT: All features	0.72986	5.34
LR: Node features	0.54134	1.38
LR: Path features	0.51418	0.74
LR: All features	0.81681	7.52
SRW: One edge type	0.82502	6.87
SRW: Multiple edge types	0.82799	7.57

Table 3. Results for the Facebook data set.

Out of 20 friendships Facebook recommends, nearly 40% are realized in the near future.

Algorithm

Extension: Global W for 100 training nodes. Test on 100 testing nodes.

- Algorithm works only for recommendation for one node or user.
  - If we want to give recommendations to multiple users, we need to run it multiple times.
- How to extend it to multiple nodes or users?
- Answer: Find best θ based on training data.

$$arg\ min_{\theta} \sum_{s \in S} \sum_{p \in P} \sum_{n \in N} \delta(r_p < r_n) + \lambda||\theta||^2$$

Penalty for violating constraint $r_p > r_n$

S are the nodes we want to give recommendation

Positive nodes

Negative nodes

r_x … score of node x on a weighted graph with edge weights $f_\theta(x, y)$

Live AB Test: Facebook Iceland Study

Live AB Test: Facebook Iceland Study

- Node and edge attributes

Live AB Test: Facebook Iceland Study

- Node and edge attributes
    - Node: Age, gender, school

Live AB Test: Facebook Iceland Study

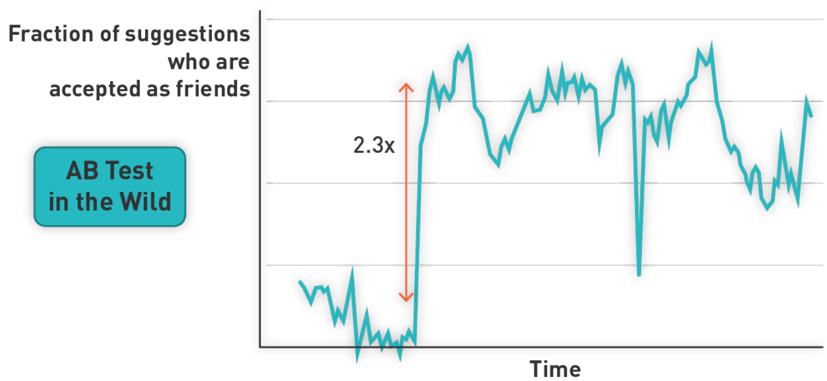
- Node and edge attributes
    - Node: Age, gender, school
    - Edge: Age of an edge, communication, profile visits, cotagged photos

Live AB Test: Facebook Iceland Study

- Node and edge attributes
    - Node: Age, gender, school
    - Edge: Age of an edge, communication, profile visits, cotagged photos
- Compared to baseline of People You May Know (PYMK)

Live AB Test: Facebook Iceland Study

- Node and edge attributes
  - Node: Age, gender, school
  - Edge: Age of an edge, communication, profile visits, cotagged photos
- Compared to baseline of People You May Know (PYMK)



Conclusion

Conclusion

- Supervised random walks: General framework for generating links on a graph

Conclusion

- Supervised random walks: General framework for generating links on a graph
  - Not specific to link prediction

Conclusion

- Supervised random walks: General framework for generating links on a graph
  - Not specific to link prediction
- Applications: Recommending experts, etc.

Conclusion

- Supervised random walks: General framework for generating links on a graph
  - Not specific to link prediction
- Applications: Recommending experts, etc.
  - Link sentiment (positive vs. negative)

Conclusion

- Supervised random walks: General framework for generating links on a graph
  - Not specific to link prediction
- Applications: Recommending experts, etc.
  - Link sentiment (positive vs. negative)
- Impressive precision at 20 but expensive to compute RWR for each user

Conclusion

- Supervised random walks: General framework for generating links on a graph
  - Not specific to link prediction
- Applications: Recommending experts, etc.
  - Link sentiment (positive vs. negative)
- Impressive precision at 20 but expensive to compute RWR for each user
- Fertile area for research and development