Enron Mention
Network

Agarwal et al.

Enron Corpus
Related work

Mention Network
Intuition
Name resolution
Types of network
Results

Thank You

References

Enron Corporation: You're the Boss if People Get Mentioned to You

Apoorv Agarwal[1]   Adinoyi Omuya[1]   Jingwei Zhang[1]
Owen Rambow[2]
Presenter: Jiehan Zheng (郑界涵)[3]

[1]Department of Computer Science, Columbia University

[2]Center for Computational Learning Systems, Columbia University

[3]Trinity College of Arts and Sciences, Duke University

August 5, 2014

Please direct questions to the first author, Apoorv Agarwal:
apoorv@cs.columbia.edu

The Enron email corpus [Klimt and Yang, 2004]

- ▶ 279,844 emails from/to 93,421 people
- ▶ Standard task: predicting dominance relations given the emails of employees

Previous work — others

- ▶ Social network analysis-based

- ▶ Natural language processing-based

Previous work — others

- ▶ Social network analysis-based
  - ▶ Rowe et al. [2007]
    - ▶ No quantitative evaluation
- ▶ Natural language processing-based

Previous work — others

- Social network analysis-based
  - Rowe et al. [2007]
    - No quantitative evaluation
- Natural language processing-based
  - Bramsen et al. [2011]
    - Test set not available to compare with
  - Gilbert [2012]
    - The use of NLP features requires complete email exchanges (covers only 19% of the gold)

Previous work — ours

- Agarwal et al. [2012]
  - Much larger gold standard (1,518 employees, 13,724 dominance pairs)
  - A simple social network analysis technique
    - Outperforms Gilbert's techniques by 5.2%

Previous work — ours

Enron Mention
Network

Agarwal et al.

Enron Corpus
Related work

Mention Network
Intuition
Name resolution
Types of network
Results

Thank You

References

- Agarwal et al. [2012]
  - Much larger gold standard (1,518 employees, 13,724 dominance pairs)
  - A simple social network analysis technique
    - Outperforms Gilbert's techniques by 5.2%

$$1 > 2 > \{3 = 4\} > 5$$

Motivation

- ▶ Enron email corpus is a very biased subset of all email exchange — no emails that are not *to* or *from* the 158 people
- ▶ Goal: to extract hidden links from the content of emails

Intuition

```
From:   Mary <mary@example.com>
  To:   Bill <bill@example.com>

Hey!  I went out to dinner with John today.
```

Intuition

Enron Mention
Network

Agarwal et al.

Enron Corpus
Related work
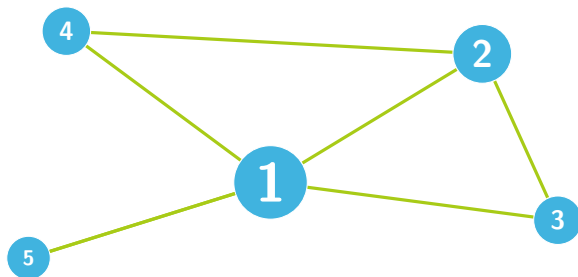Mention Network
Intuition
Name resolution
Types of network
Results

Thank You

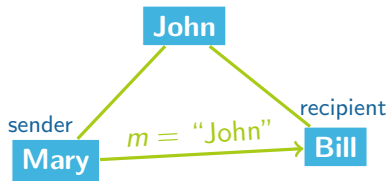References

```
From:   Mary <mary@example.com>
  To:   Bill <bill@example.com>

Hey!  I went out to dinner with John today.
```

▶ Add links between sender, recipients and each person
mentioned

Creating mention networks

Tasks:

- ▶ Detect entity mentions in body of emails
- ▶ Resolve entity mentions to canonical person entities, i.e. nodes in the network

Creating mention networks

Tasks:

- ▶ Detect entity mentions in body of emails (Java Extraction Toolkit [Grishman et al., 2005])
- ▶ Resolve entity mentions to canonical person entities, i.e. nodes in the network

Creating mention networks

Tasks:

- ▶ Detect entity mentions in body of emails (Java Extraction Toolkit [Grishman et al., 2005])
- ▶ Resolve entity mentions to canonical person entities, i.e. nodes in the network (with our algorithm)

Our name disambiguation algorithm

Enron Mention
Network

Agarwal et al.

Enron Corpus
Related work

Mention Network
Intuition
Name resolution
Types of network
Results

Thank You

References

```
From:   Mary <mary@example.com>
  To:   Bill <bill@example.com>

Hey!  I went out to dinner with John today.
```

Our name disambiguation algorithm

Enron Mention
Network

Agarwal et al.

Enron Corpus
Related work

Mention Network
Intuition
Name resolution
Types of network
Results

Thank You

References

```
From:   Mary <mary@example.com>
  To:   Bill <bill@example.com>

Hey!  I went out to dinner with John today.
```

$$min_{\{p_k | f(p_k, m) = 1\}}[d(p_s, p_k) + d(p_r, p_k)]$$

Our name disambiguation algorithm

Enron Mention
Network

Agarwal et al.

Enron Corpus
Related work

Mention Network
Intuition
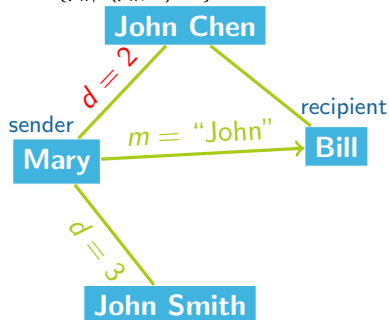Name resolution
Types of network
Results

Thank You

References

```
From:   Mary <mary@example.com>
  To:   Bill <bill@example.com>

Hey!  I went out to dinner with John today.
```

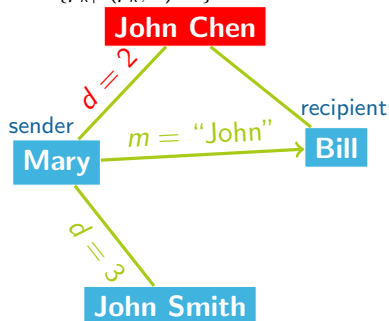$$min_{\{p_k | f(p_k, m) = 1\}}[d(p_s, p_k) + d(p_r, p_k)]$$

Minimize	% acc.
$d(\text{sender}, \text{mention})$	60.4
$d(\text{recipient}, \text{mention})$	55.5
$d(\text{sender}, \text{mention}) +$ $d(\text{recipient}, \text{mention})$	**69.7**
Minkov et al. [2006]	*62.3*

Variations we experimented with

Enron Mention
Network

Agarwal et al.

Enron Corpus
Related work

Mention Network
Intuition
Name resolution
Types of network
Results

Thank You

References

- ▶ Weighted or unweighted
- ▶ Email only (E), mention only (M), or combined
- ▶ In-degree (In), out-degree (Out), or degree centrality (Deg)
- ▶ Link mentioned person to sender (S), recipients (R), or both (RS)

Link to sender, recipients, or both

▶ Worst system which adds links from mention to recipients (In-MRD): 73.6%

▶ Best system which adds links between mention and sender (Deg-MSD): 73.4%

Link to sender, recipients, or both

- Worst system which adds links from mention to recipients (In-MRD): 73.6%
- Best system which adds links between mention and sender (Deg-MSD): 73.4%
- Therefore it's better to add links from mentioned person to *recipients*

Results

Network type	% accuracy
Email only (from Agarwal et al. [2012])	85.2 (Deg-EU)
Mention only	**87.3** (Deg-MRU)
Combined	86.8 (Out-EDMSRU)

Results

Employee	Title	Email	Mention
John Lavorato*	COO	992	452
Phillip K Allen	Trader	1771	248
David W Delainey*	COO	1093	298
Phillip K Allen	Trader	1771	248

Thank You!

Enron Mention
Network

Agarwal et al.

Enron Corpus
Related work

Mention Network
Intuition
Name resolution
Types of network
Results

Thank You

References

For details, please refer to the paper.

Please direct questions to the first author, Apoorv Agarwal:
`apoorv@cs.columbia.edu`

Enron Mention
Network

Agarwal et al.

Enron Corpus
Related work

Mention Network
Intuition
Name resolution
Types of network
Results

Thank You

References

Apoorv Agarwal, Adinoyi Omuya, Aaron Harnly, and Owen Rambow. A comprehensive gold standard for the enron organizational hierarchy. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 161–165, Jeju Island, Korea, July 2012. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/P12-2032.

Philip Bramsen, Martha Escobar-Molano, Ami Patel, and Rafael Alonso. Extracting social power relationships from natural language. In *ACL*, pages 773–782, 2011.

Eric Gilbert. Phrases that signal workplace hierarchy. In *CSCW*, pages 1037–1046, 2012.

Ralph Grishman, David Westbrook, and Adam Meyers Proc. NYU's English ACE 2005 system description. In *ACE Evaluation Workshop*, 2005.

Bryan Klimt and Yiming Yang. Introducing the Enron Corpus. In *CEAS*, 2004.

Einat Minkov, William W. Cohen, and Andrew Y. Ng. Contextual search and name disambiguation in email using graphs. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 27–34, New York, NY, USA, August 2006. ACM Press. ISBN 1595933697. doi: 10.1145/1148170.1148179. URL http://dx.doi.org/10.1145/1148170.1148179.

Ryan Rowe, German Creamer, Shlomo Hershkop, and Salvatore J Stolfo. Automated social hierarchy detection through email network analysis. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, WebKDD/SNA-KDD '07, pages 109–117, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-848-0. doi: 10.1145/1348549.1348562. URL http://doi.acm.org/10.1145/1348549.1348562.