# Social Information Extraction

**Li Dehao**
**Shao Qi**

**2013/05/31**

# Questions

- **Are you using any social network?**

- **Have you ever posted any personal information onto the Internet?**

# Questions (Cont'd)

- **Do you think the information you post is valuable?**

每回离家回京前，都会莫名与爸妈发火争吵，或许这也是我们之间表达爱的一种方式吧。直到送别完看不见彼此，心底深处那个幼时的我，便会蜷缩成一团，尽情嚎啕大哭起来，一遍遍叫喊着：爸爸妈妈，你们不要老...

2月15日 10:13 来自iPhone客户端 (5499) | 转发(37810) | 收藏 | 评论(13356)

# Contents

- Introduction
- Applications
- Methodologies
- Conclusion
  - Novelty
  - Limitation

# Introduction

- **What are social networks?**

  Social networks are mostly referred to as a combination of Information Technologies to make communication and conversations into an interactive format.

- **What is the role that social networks play?**

  The interactivity of social networks makes users become self-media, where they can voluntarily show themselves.

- **What is the motivation for the extraction?**

  Social information contains rich details about the author, so it is potentially useful in observing the characteristics and actual needs of a social network user.

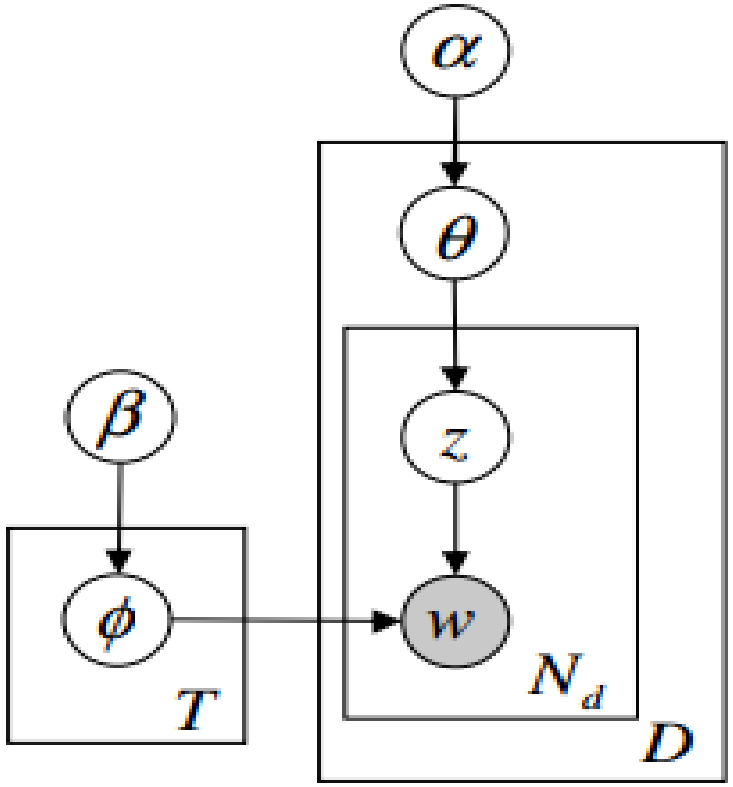
# Applications

- **Classify users**
- **Cluster users**
- **Precisely advertise**

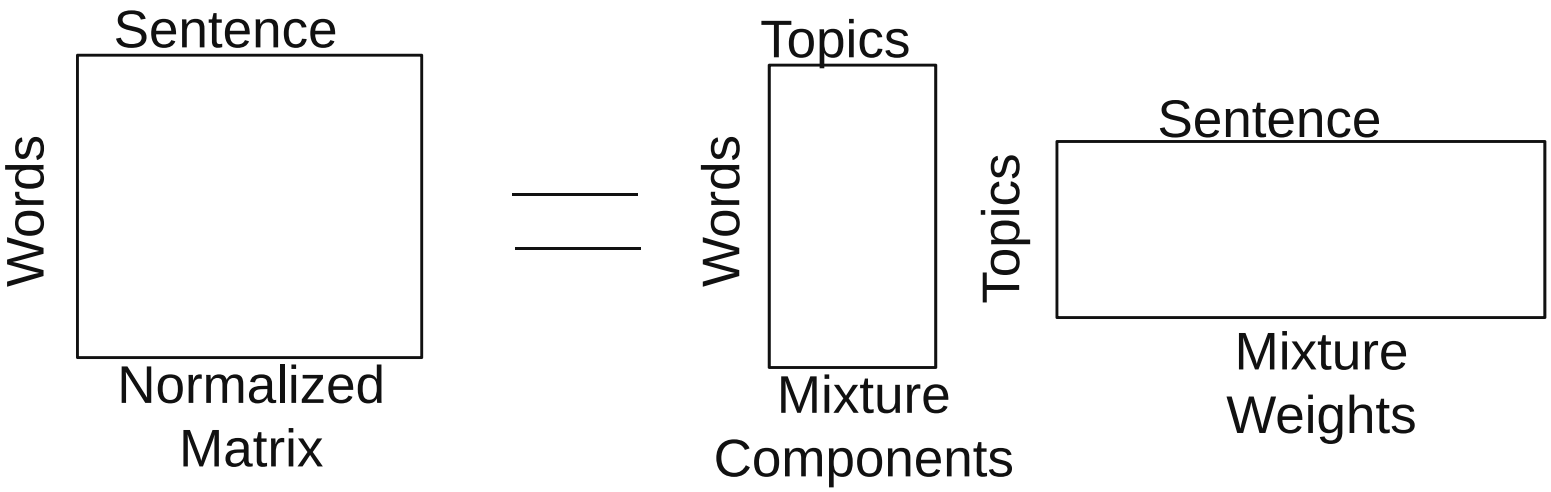
# Methodologies

- **Probabilistic Topic Model**
- **Vector Space Model**
- **Correlation Calculation**

# Standard probabilistic topic model



# Vector Space Model



# Correlation Calculation

Example:

| Matrix | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ |
|---|---|---|---|---|---|---|
| basketball | 1 | 0 | 1 | 0 | 0 | 0 |
| soccer | 0 | 1 | 0 | 0 | 0 | 0 |
| athlete | 1 | 1 | 0 | 0 | 0 | 0 |
| medal | 1 | 0 | 0 | 1 | 1 | 0 |
| man | 0 | 0 | 0 | 1 | 0 | 1 |

Using linear algebra method:

$$\mathrm{sim}(\overrightarrow{S_2}, \overrightarrow{S_3}) = \frac{0 \times 1 + 1 \times 0 + 1 \times 0 + 0 \times 0 + 0 \times 0}{|\overrightarrow{S_2}| \times |\overrightarrow{S_3}|} = 0$$

# Novelty

- **User’s historical information**
- **Matrix decomposition**

# Limitation

- **Synonym**
- **Polysemy**

# Thank you!

# References


1. Chemudugunta, Chaitanya, & Steyvers, Padhraic Smyth Mark. (2007). *Modeling General and Specific Aspects of Documents with a Probabilistic Topic Model.* Paper presented at the Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference.

2. Feilmayr, Christina. (2011). *Text Mining-Supported Information Extraction: An Extended Methodology for Developing Information Extraction Systems.* Paper presented at the Database and Expert Systems Applications (DEXA), 2011 22nd International Workshop on.

3. Jayagopi, Dinesh Babu, & Gatica-Perez, Daniel. (2010). Mining group nonverbal conversational patterns using probabilistic topic models. *Multimedia, IEEE Transactions on, 12*(8), 790-802.

4. Steyvers, Mark, & Griffiths, Tom. (2007). Probabilistic topic models. *Handbook of latent semantic analysis, 427*(7), 424-440.

5. Wassell, Alexander, Rubin, Stuart H, & Frost, Eric G. (2011). *Integrated social information engineering.* Paper presented at the Information Reuse and Integration (IRI), 2011 IEEE International Conference on.