Modeling Topic-level Academic Influence in Scientific Literatures

Jiaming Shen, Zhenyu Song, Shitao Li, Zhaowei Tan, Yuning Mao, Luoyi Fu, Li Song, Xinbing Wang

Shanghai Jiao Tong University

Feb 13, 2016



Jiaming Shen, Zhenyu Song, Shitao Li, Zhaov<mark>Modeling Topic-level Academic Influence in S</mark>

Feb 13, 2016 1 / 37

Outline

1 Motivation

- **2** J-Index Framework
- Seference Topic Model (RefTM) Generative Model Parameter Estimation

4 Experiments

Datasets Evaluation Aspects Evaluation Results

Outline

1 Motivation

- **2** J-Index Framework
- 8 Reference Topic Model (RefTM) Generative Model Parameter Estimation

Output State St

Datasets Evaluation Aspects Evaluation Results

When a beginner starts to explore a new field ...

Jiaming Shen, Zhenyu Song, Shitao Li, Zhaov<mark>Modeling Topic-level Academic Influence in S</mark>

Feb 13, 2016 4 / 37

Google	Statistical Machine Learning		
Scholar	About 2,160,000 results (0.05 sec)		
Articles Case law My library	Machine learning, neural and statistical classification D Michie, <u>DJ Spiegehalter</u> , CC Taylor - 1994 - Classer Abstract The aim of this book is to provide an up-to-date review of different approaches to classification, compare their performance on a wide range of challenging data-sets, and draw conclusions on their applicability to realistic industrial problems. Cited by 2686 Related articles all 9 versions C cite. Save More		
Any time Since 2016 Since 2015 Since 2012 Custom range	[BOOK] Statistical learning theory <u>YN Vapnik</u> . V Vapnik - 1988 - ai.ato.ms The theory provides a sound statistical hasis for assessing model adequacy under these circumstances, which are precisely the circumstances encountered in MACHINE LEARNING, PATTERN RECOGNITION, and exploratory data analysis Cited by 2572 Related articles. All Y versions C to: Save More		
Sort by relevance Sort by date	[PDF] Statistical machine learning makes automatic control practical for internet datacenters <u>PBodk</u> , R Griffith, <u>CSution</u> , A Fox, <u>M Jordan</u> Proceedings of the, 2009 - usenix.org Abstract Horizontally-scalable Internet services on clusters of commodity computers appear to be a center if the expensition experiment there is a treat center of commodity computers appear		
 ✓ include patents ✓ include citations 	observed output (actual latency), and gain controller (adjusting the number of servers). Yet Cited by 137 Related articles All 15 versions Cite Save More		
Create alert	Distributed optimization and statistical earning via the alternating direction method of multipliers <u>S Boyd, N Parish, E Chu, B Peleato</u> © In Machine Learning, 2011 - diacm.org Abstract Mary problems of recent interest in statistics and machine learning can be posed in the framework of convex optimization. Due to the explosion in size and complexity of modern datasets, it is increasingly important to be able to solve problems with wery large number Cited by 2964 Related anticles All 40 versions Cite Save More		

Figure 1 : Result of Google Scholar

Jiaming Shen, Zhenyu Song, Shitao Li, Zhaov<mark>Modeling Topic-level Academic Influence in S</mark>

 Image: wide black
 Image: wide black

<ロト (四) (三) (三) (三)

Google	Statistical Machine Learning - Q				
Scholar	About 2,160,000 results (0.05 sec How to rank these papers?				
Articles	Machine learning, neural and statistical classification				
Case law	Abstract The aim of this book is to provide an up-to-date review of different approaches to				
My library	classification, compare their performance on a wide range of challenging data-sets, and draw conclusions on their applicability to realistic industrial problems. Cited by 2688 Related articles All 9 versions Cite Save More				
Any time Since 2016 Since 2015 Since 2012 Custom range	(BOOK) Statistical learning theory (YN Vapnik, - Vlapnik, - 1996 - a lato, ms The theory provides a sound statistical basis for assessing model adequacy under these oricrumstances, which are precisely the circrumstances encountered in MACHINE LEARNING, -PATTERN RECOGNITION, and exploratory data analysis Cited by 25723 Related articles All 7 versions Cite Save More (POF) Statistical machine learning makes automatic control practical for internet datacenters Badytis, R offittin, C Suiton, A For M. Jordan, 2009 - usenix.org Abstract Horizontally-sealable Internet services on clusters of commodily computers appear the as one of for automatic control them is a target of using the advected internet services on clusters of commodily computers.				
Sort by relevance Sort by date					
 ✓ include patents ✓ include citations 	observed output (actual latency), and gain controller (adjusting the number of servers). Yet Cited by 137 Related articles All 15 versions Cite Save More				
Create alert	Instruction organization and statistical rearming via the alternating direction method of multipliers Sext, N.Parkh, E. Chu, B.Peleato,, @ In Machine Learning, 2011 - diacm.org Abstract Mary problems of recent inferest in statistical and machine learning can be posed in the framework of convex optimization. Due to the explosion in size and complexity of modern datasets, it is increasingly important to be able to solve problems with a very large number Cited by 2964 Related articles All 40 versions Cite Save More				

Figure 2 : Defects of Google Scholar

Jiaming Shen, Zhenyu Song, Shitao Li, Zhaov<mark>Modeling Topic-level Academic Influence in S</mark>

 ▲ ≣
 ● ○ < ○</th>

 Feb 13, 2016
 6 / 37

<ロト (四) (三) (三) (三)

Google	Statistical Machine Learning - Q					
Scholar	About 2,160,000 results (0.05 sec					
Articles Case law My library Any time Since 2016 Since 2016 Since 2015 Since 2015 Custom range Sort by relevance Sort by date	Machine learning, neural and statistical classification D Michie, DJ Spiegelhalitz, CC Taylor - 1994 - Citeseer Abstract. The ain of this book is to provide an up-to-date review of different approaches to classification, compare their performance on a wide range of challenging data-sets, and draw conclusions on their applicability to realistic industrial problems. Cited by 2868 Related articles All 9 versions Cite Save More Suppose we intend to follow this direction Which related works. Suppose the performance on a substract The ain statistic classification compare their performance on a wide range of challenging data-sets, and the performance on a wide range of challenging data-sets, and stave contains on their applicability to realistic industrial problems. Suppose we intend to follow this direction Which related works. The theory provides a sound statistic bases into motion of motion and provides and the sound statistic bases into motion of the ain theory provides a sound statistic bases and the statistic on the performance on contrader in Machine Elated works. Cited by 25723 Related articles All 7 versions Cite Save More PPDF Statistical machine learning makes automatic control practical for internet datacenters P. Paddis, R. Griffith, C. Sutido, R. For M. Judon "Droceedings of the, 2009 - users, and Abstract herizontally-scalable internet services on clusters of commotify computers ansear to be a great fifth. C. Sutido, R. For A. Maddis, and and controller (adjusting the number of servers), Yet Cleabely 13 for automatic control fifth 5 versions. Cite Save More					
Create alert	Distributed optimization and statistical learning via the alternating direction method of multipliers Says, N. Parkb, E. Chu, B. Pelesto,, @ in Machine Learning, 2011 - di.acm.org Abstract Many problems of recent interest in statistics and machine learning can be posed in the framework of orows optimization. Due to the explosion in size and complexity of modern datasets, it is increasingly important to be able to solve problems with a very large number Cited by 2864. Related articles all 40 versions. Cite. Says Wore.					

Figure 3 : Defects of Google Scholar

Jiaming Shen, Zhenyu Song, Shitao Li, Zhao<mark>y</mark>Modeling Topic-level Academic Influence in S

▲ ≣ ► ≣ ∽ ९ ०
 Feb 13, 2016 7 / 37

<ロト (四) (三) (三) (三)

Stand on the shoulders of giants

- Isaac Newton

Jiaming Shen, Zhenyu Song, Shitao Li, Zhaov<mark>Modeling Topic-level Academic Influence in S</mark>

Feb 13, 2016 8 / 37

Motivation

Find those giants !!!

Jiaming Shen, Zhenyu Song, Shitao Li, Zhaov<mark>Modeling Topic-level Academic Influence in S</mark>

Feb 13, 2016 9 / 37

< 47 ▶

Outline

Motivation

2 J-Index Framework

8 Reference Topic Model (RefTM) Generative Model Parameter Estimation

4 Experiments

Datasets Evaluation Aspects Evaluation Results



Figure 4 : Factors of one paper

<ロト (四) (三) (三) (三)

э



Figure 5 : Factors of one paper

э



Figure 6 : Factors of one paper

3



Figure 7 : Factors of one paper

Feb 13, 2016 14 / 37

3

・ロト ・ 日 ・ ・ ヨ ・ ・ ヨ ト



Figure 8 : Factors of one paper

3



Figure 9 : Factors of one paper

• Three assumptions of J-Index:

A B A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A

< E.

- Three assumptions of J-Index:
 - 1 A paper's academic influence increases as it gains more citations.

- Three assumptions of J-Index:
 - 1 A paper's academic influence increases as it gains more citations.
 - **2** A paper with stronger citations intends to be more influential.

- Three assumptions of J-Index:
 - 1 A paper's academic influence increases as it gains more citations.
 - **2** A paper with stronger citations intends to be more influential.
 - **3** A paper cited by more innovative papers is more influential.

- Three assumptions of J-Index:
 - 1 A paper's academic influence increases as it gains more citations.
 - 2 A paper with stronger citations intends to be more influential.
 - 3 A paper cited by more innovative papers is more influential.
- We define the *J-Index* as follows:

$$\mathsf{J-Index-Score}(u) = \sum_{c \in C(u)} \lambda(c) \times \delta(c, u)$$

- C(u): the set of paper u's citations, obtained from input dataset.
- $\lambda(c)$: the innovativeness of paper c.
- $\delta(c, u)$: the citation strength between paper c and paper u.
- Both $\lambda(c)$ and $\delta(c, u)$ are obtained from subsequent model.

Outline

Motivation

2 J-Index Framework

Seference Topic Model (RefTM) Generative Model Parameter Estimation

4 Experiments

Datasets Evaluation Aspects Evaluation Results

4 A N

20 / 37

Generative Model

• Reference Topic Model is one way to obtain $\lambda(c)$ and $\delta(c, u)$.

- Reference Topic Model is one way to obtain $\lambda(c)$ and $\delta(c, u)$.
- The intuition: a researcher may write a word based on his/her own idea or "inherits" some thoughts from one of its references.

- Reference Topic Model is one way to obtain $\lambda(c)$ and $\delta(c, u)$.
- The intuition: a researcher may write a word based on his/her own idea or "inherits" some thoughts from one of its references.
- Topic Innovation: come from one's own idea.

- Reference Topic Model is one way to obtain $\lambda(c)$ and $\delta(c, u)$.
- The intuition: a researcher may write a word based on his/her own idea or "inherits" some thoughts from one of its references.
- Topic Innovation: come from one's own idea.
- Topic Inheritance: come from one of cited papers.

- Reference Topic Model is one way to obtain $\lambda(c)$ and $\delta(c, u)$.
- The intuition: a researcher may write a word based on his/her own idea or "inherits" some thoughts from one of its references.
- Topic Innovation: come from one's own idea.
- Topic Inheritance: come from one of cited papers.
- Citation Strength: determine which reference is selected

1. For each topic index $k \in \{1, \ldots, K\}$ (a) Draw a word distribution $\varphi_k \sim \text{Dir}(\beta)$ 2. For each document index $m \in \{1, \ldots, M\}$ (a) Draw a topic distribution $\theta_m \sim \text{Dir}(\alpha)$ (b) Draw a reference distribution $\delta_m \sim \text{Dir}(\eta | L_m)$ (c) Draw an inheritance index $\lambda_m \sim \text{Beta}(\alpha_{\lambda_m}, \alpha_{\lambda_n})$ (d) For each word $n \in \{1, ..., N_m\}$ in document m: (i) Flip a coin $s_{m,n} \sim \text{Bern}(\lambda_m)$ (ii) if $s_{m,n} = 0$: Draw a topic $z_{m,n} \sim \text{Multi}(\boldsymbol{\theta}_m)$ Draw a word $w_{m,n} \sim \text{Multi}(\varphi_{z_m})$ (iii) else $(s_{m,n} = 1)$: Draw a reference $c_{m,n} \sim \text{Multi}(\boldsymbol{\delta}_m)$ Draw a topic $z_{m,n} \sim \text{Multi}(\theta_{c_{m,n}})$ Draw a word $w_{m,n} \sim \text{Multi}(\varphi_{z_m})$

Figure 10 : Generative Model of RefTM

Feb 13, 2016 21 / 37

Generative Model

1. For each topic index $k \in \{1, \ldots, K\}$ (a) Draw a word distribution $\varphi_k \sim \text{Dir}(\beta)$ 2. For each document index $m \in \{1, \ldots, M\}$ (a) Draw a topic distribution $\theta_m \sim \text{Dir}(\alpha)$ (b) Draw a reference distribution $\delta_m \sim \text{Dir}(\eta | L_m)$ (c) Draw an inheritance index $\lambda_m \sim \text{Beta}(\alpha_{\lambda_m}, \alpha_{\lambda_n})$ (d) For each word $n \in \{1, \ldots, N_m\}$ in document m: Topic (i) Flip a coin $s_{m,n} \sim \text{Bern}(\overline{\lambda_m})$ Innovativeness Innovation (ii) if $s_{m.n} = 0$: Draw a topic $z_{m,n} \sim \text{Multi}(\boldsymbol{\theta}_m)$ Draw a word $w_{m,n} \sim \text{Multi}(\varphi_{z_m})$ (iii) else $(s_{m,n} = 1)$: Topic Citation Draw a reference $c_{m,n} \sim \text{Multi}(\delta_m)$ Inheritence Strength Draw a topic $z_{m,n} \sim \text{Multi}(\boldsymbol{\theta}_{c_m})$ Draw a word $w_{m,n} \sim \text{Multi}(\varphi_{z_m})$

Figure 11 : Generative Model of RefTM

Feb 13, 2016

22 / 37

Outline

Motivation

2 J-Index Framework

Seference Topic Model (RefTM) Generative Model Parameter Estimation

4 Experiments

Datasets Evaluation Aspects Evaluation Results

5 Conclusions & Future works

Jiaming Shen, Zhenyu Song, Shitao Li, Zhaov<mark>Modeling Topic-level Academic Influence in S</mark>

Feb 13, 2016 23 / 37

• Roll a dice N times, side i shows n_i times.

24 / 37

- Roll a dice N times, side i shows n_i times.
- What's the best estimation of each side's probability.

24 / 37

- Roll a dice N times, side i shows n_i times.
- What's the best estimation of each side's probability.
- For each side *i*, probability $P_i = \frac{n_i}{N} = \frac{n_i}{\sum_{i=1}^6 n_j}$

- Roll a dice N times, side i shows n_i times.
- What's the best estimation of each side's probability.
- For each side *i*, probability $P_i = \frac{n_i}{N} = \frac{n_i}{\sum_{i=1}^6 n_j}$
- Smooth Effect: Suppose we have already view each side $i \lambda_i$ times.

- Roll a dice N times, side i shows n_i times.
- What's the best estimation of each side's probability.
- For each side *i*, probability $P_i = \frac{n_i}{N} = \frac{n_i}{\sum_{i=1}^6 n_j}$
- Smooth Effect: Suppose we have already view each side $i \lambda_i$ times.
- Update: For each side i, probability $P_i = \frac{n_i + \lambda_i}{\sum_{j=1}^6 (n_j + \lambda_j)}$

- Roll a dice N times, side i shows n_i times.
- What's the best estimation of each side's probability.
- For each side *i*, probability $P_i = \frac{n_i}{N} = \frac{n_i}{\sum_{j=1}^6 n_j}$
- Smooth Effect: Suppose we have already view each side $i \lambda_i$ times.
- Update: For each side *i*, probability $P_i = \frac{n_i + \lambda_i}{\sum_{i=1}^{6} (n_i + \lambda_i)}$
- Inference: Observation \rightarrow Parameters

- Roll a dice N times, side i shows n_i times.
- What's the best estimation of each side's probability.
- For each side i, probability $P_i = \frac{n_i}{N} = \frac{n_i}{\sum_{j=1}^6 n_j}$
- Smooth Effect: Suppose we have already view each side $i \lambda_i$ times.
- Update: For each side *i*, probability $P_i = \frac{n_i + \lambda_i}{\sum_{i=1}^{6} (n_j + \lambda_i)}$
- Inference: Observation >>> Parameters
- In RefTM, observations: words & citations; parameters (we mainly concerned): λ and δ

RefTM Inference: Gibbs Sampling

$$\begin{split} p(s_i = 0 | \mathbf{s}_{\neg i}, \mathbf{w}, \mathbf{z}, \cdot) &\propto \\ & \frac{(n_m^{s_i^{(0)}} - 1) + n_m^{s_i^{(1)}} + \alpha}{n_m^{(\gamma)(0)} + n_m^{(\gamma)(1)} + K\alpha - 1} \cdot \frac{N_m^{(0)} - 1 + \alpha_{\lambda_n}}{N_m^{(1)} + (N_m^{(0)} - 1) + \alpha_{\lambda_n} + \alpha_{\lambda_c}} \\ p(s_i = 1 | \mathbf{s}_{\neg i}, \mathbf{w}, \mathbf{z}, \mathbf{c}_i, \cdot) &\propto \\ & \frac{n_{c_i}^{z_i^{(0)}} + (n_{c_i}^{s_i^{(1)}} - 1) + \alpha}{n_{c_i}^{(\gamma)(0)} + n_{c_i}^{(\gamma)(1)} + K\alpha - 1} \cdot \frac{N_m^{(1)} - 1 + \alpha_{\lambda_c}}{(N_m^{(1)} - 1) + N_m^{(0)} + \alpha_{\lambda_n} + \alpha_{\lambda_c}} \\ p(c_i | \mathbf{c}_{\neg i}, \mathbf{w}, \mathbf{z}, \mathbf{s}_i = 1, \cdot) &\propto \\ & \frac{n_{c_i}^{z_i^{(0)}} + (n_{c_i}^{z_i^{(1)}} - 1) + \alpha}{n_{c_i}^{\gamma(0)} + n_{c_i}^{\gamma(1)} + K\alpha - 1} \cdot \frac{R_m^{c_i} - 1 + \eta}{R_m^{(-)} + L_m \eta - 1} \\ p(z_i | \mathbf{z}_{\neg i}, \mathbf{w}, \mathbf{s}_i = 0, \cdot) &\propto \\ & \frac{n_{s_i}^{s_i} + \beta - 1}{n_{s_i}^{\gamma(1)} + V\beta - 1} \cdot \frac{(n_m^{s_i^{(0)}} - 1) + n_m^{s_i^{(1)}} + \alpha}{n_{c_i}^{\gamma(0)} + n_m^{\gamma(1)} + K\alpha - 1} \\ p(z_i | \mathbf{z}_{\neg i}, \mathbf{w}, \mathbf{s}_i = 1, c_i, \cdot) &\propto \\ & \frac{n_{s_i}^{s_i} + \beta - 1}{n_{s_i}^{\gamma(1)} + V\beta - 1} \cdot \frac{n_{c_i}^{z_i^{(0)}} + (n_{c_i}^{z_i^{(1)}} - 1) + \alpha}{n_{c_i}^{\gamma(0)} + n_{c_i}^{\gamma(1)} + K\alpha - 1} \end{split}$$

Algorithm 1 Gibbs Sampling Algorithm for RefTM **Input:** $K, w, \alpha, \beta, \eta, \lambda_c, \lambda_n$ **Output:** Parameter sets $\{\theta, \varphi, \delta, \lambda\}$ Read in data and zero out all count caches Randomly initialize z_i, c_i, s_i for iter = 1 to N_{iter} do for all documents $m \in [1, M]$ do for all words $n \in [1, N_m]$ in document m do if $s_{m,n} = 0$ then Update the counts $n_m^{(k)(0)}, n_m^{(0)}$ else Update the counts $n_c^{(k)(1)}, n_c^{(1)}, R_m^c, R_m$ Draw a new \tilde{s} from Eqs.(2-3) if $\tilde{s} = 0$ then Update the counts $n_{k}^{w_{m,n}}, n_{k}$ Draw a new topic \tilde{k} from Eq.(5) Update the counts $n_m^{(\tilde{k})(0)}, n_m^{(0)}, n_{\tilde{k}}^{w_{m,n}}, n_{\tilde{k}}$ else Draw a new reference \tilde{c} from Eq.(4) Update the counts $R_m^{\tilde{c}}, R_m, n_k^{w_{m,n}^*}, n_k$ Draw a new topic \tilde{k} from Eq.(6) Update the counts $n_{\tilde{z}}^{(\tilde{k})(1)}, n_{\tilde{c}}, n_{\tilde{z}}^{w_{m,n}}, n_{\tilde{k}}$ Read out parameters set $\theta, \varphi, \lambda, \delta$ by Eqs.(7-10)

Figure 12 : Gibbs sampling equations & Algorithm for RefTM

Jiaming Shen, Zhenyu Song, Shitao Li, Zhaov<mark>Modeling Topic-level Academic Influence in S</mark>

Visualization of RefTM's output



Figure 13 : Right hand side is an illustrative citation graph in which the thickness of edge represents the citation strength and the vertex size indicates one papers academic influence. Left hand side presents each paper's *J-Index*.

Feb 13, 2016 26 / 37

Outline

Motivation

- **2** J-Index Framework
- 8 Reference Topic Model (RefTM) Generative Model Parameter Estimation

4 Experiments

Datasets Evaluation Aspects Evaluation Results

5 Conclusions & Future works

Jiaming Shen, Zhenyu Song, Shitao Li, Zhaov<mark>Modeling Topic-level Academic Influence in S</mark>

Feb 13, 2016

27 / 37

Datasets

- Dataset 1: a large unsupervised collection of 426728 articles with over 209 million citations.
- Dataset 2: a small supervised collection of 799 papers obtained from (Liu et al. 2010).
- The average paper length of two corpora are 83 and 98 words.

< 47 ►

3

29 / 37

Evaluation Aspects

• Topic Coherence

- Topic Coherence
 - Metrics: PMI-Score (Newman et al. 2010) and topic coherence-Score (Mimno et al. 2011).

29 / 37

- Topic Coherence
 - Metrics: PMI-Score (Newman et al. 2010) and topic coherence-Score (Mimno et al. 2011).
 - **2** Using dataset 1 and an external dataset of 3.34 million papers when calculating PMI-Score.

- Topic Coherence
 - Metrics: PMI-Score (Newman et al. 2010) and topic coherence-Score (Mimno et al. 2011).
 - **2** Using dataset 1 and an external dataset of 3.34 million papers when calculating PMI-Score.
- Citation Strength Prediction

- Topic Coherence
 - Metrics: PMI-Score (Newman et al. 2010) and topic coherence-Score (Mimno et al. 2011).
 - **2** Using dataset 1 and an external dataset of 3.34 million papers when calculating PMI-Score.
- Citation Strength Prediction
 - Using dataset 2, in which the strength of each citation is classified into three levels - "1, 2, 3".

- Topic Coherence
 - Metrics: PMI-Score (Newman et al. 2010) and topic coherence-Score (Mimno et al. 2011).
 - 2 Using dataset 1 and an external dataset of 3.34 million papers when calculating PMI-Score.
- Citation Strength Prediction
 - Using dataset 2, in which the strength of each citation is classified into three levels - "1, 2, 3".
 - Metrics: averaged AUC value for decision boundaries "1 vs. 2, 3" and "1, 2 vs. 3".

- Topic Coherence
 - Metrics: PMI-Score (Newman et al. 2010) and topic coherence-Score (Mimno et al. 2011).
 - 2 Using dataset 1 and an external dataset of 3.34 million papers when calculating PMI-Score.
- Citation Strength Prediction
 - Using dataset 2, in which the strength of each citation is classified into three levels - "1, 2, 3".
 - Metrics: averaged AUC value for decision boundaries "1 vs. 2, 3" and "1, 2 vs. 3".
- Case Study: Rank INFOCOM

Experiments

Evaluation Results

Topic Coherence





- PMI-Score: RefTM outperforms LDA by 12% when K = 50.
- Topic Coherence-Score: RefTM outperforms LDA slightly.

Citation Strength Prediction



Figure 15 : Citation Strength Prediction measured by averaged AUC

- Reduce the normalization constraint of δ in RefTM.
- RefTM clearly outperforms two baseline methods.

Case Study: Rank INFOCOM

Title	J-Index	citation counts			
Top 5 Articles in INFOCOM 2003 ranked by <i>J-Index</i>					
Ad hoc positioning system (APS) using AOA		115			
Performance anomaly of 802.11b		127			
Packet leashes: a defense against wormhole attacks in wireless networks		74			
Unreliable sensor grids: coverage, connectivity and diameter	4.00	82			
Sensor deployment and target localization based on virtual forces		60			
Top 5 Articles in INFOCOM 2003 ranked by citation number					
Performance anomaly of 802.11b	5.17	127			
Ad hoc positioning system (APS) using AOA		115			
Optimal routing, link scheduling and power control in multihop wireless networks		109			
Sprite: a simple, cheat-proof, credit-based system for mobile ad-hoc networks		88			
Unreliable sensor grids: coverage, connectivity and diameter		82			

Table 2: Top 5 Articles in INFOCOM 2003 randked by J-Index & citaions

Figure 16 : Citation Strength Prediction measured by averaged AUC

- Rankings by *J-Index* and citations number are correlated.
- J-Index favors those paper that propose novel "ideas".

Outline

Motivation

- **2** J-Index Framework
- 8 Reference Topic Model (RefTM) Generative Model Parameter Estimation

4 Experiments

Datasets Evaluation Aspects Evaluation Results

G Conclusions & Future works

Feb 13, 2016

33 / 37

• Conclusions:

A B A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A

< E.

Conclusions:

1 Model academic influence – facilitate ranking and recommendation.

- Conclusions:
 - **1** Model academic influence facilitate ranking and recommendation.
 - 2 J-Index framework consider citation strength and paper's novelty.

- Conclusions:
 - **1** Model academic influence facilitate ranking and recommendation.
 - 2 J-Index framework consider citation strength and paper's novelty.
 - **3** Reference Topic Model combine citation network into topic model.

- Conclusions:
 - 1 Model academic influence facilitate ranking and recommendation.
 - 2 J-Index framework consider citation strength and paper's novelty.
 - **3** Reference Topic Model combine citation network into topic model.
- Future works:

- Conclusions:
 - 1 Model academic influence facilitate ranking and recommendation.
 - 2 J-Index framework consider citation strength and paper's novelty.
 - **3** Reference Topic Model combine citation network into topic model.
- Future works:

1 RefTM in the incremental citation network.

- Conclusions:
 - **1** Model academic influence facilitate ranking and recommendation.
 - 2 J-Index framework consider citation strength and paper's novelty.
 - **3** Reference Topic Model combine citation network into topic model.
- Future works:
 - **1** RefTM in the incremental citation network.
 - **2** Consider multiple factors, especially the temporal information.

References

- Dietz, L.; Bickel, S.; and Scheffer, T. 2007. Unsupervised prediction of citation influences. In *Proceedings of the 24th international conference on Machine learning*, 233 – 240. ACM.
- Liu, L.; Tang, J.; Han, J.; Jiang, M.; and Yang, S. 2010. Mining topic-level influence in heterogeneous networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, 199 208. ACM.
- Mimno, D.; Wallach, H. M.; Talley, E.; Leenders, M.; and McCallum, A. 2011. Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262 – 272. ACM.
- Newman, D.; Lau, J. H.; Grieser, K.; and Baldwin, T. 2010. Automatic evaluation of topic coherence. In *Human Language Technologies: The* 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 100 – 108. ACM.

Thank you!

Jiaming Shen, Zhenyu Song, Shitao Li, Zhaov<mark>Modeling Topic-level Academic Influence in S</mark>

Feb 13, 2016 36 / 37

Q & A

Jiaming Shen, Zhenyu Song, Shitao Li, Zhaov<mark>Modeling Topic-level Academic Influence in S</mark>

æ Feb 13, 2016 37 / 37

(日) (四) (三) (三) (三)