Towards Interactive Construction of Topical Hierarchy

Chi Wang
Microsoft Research, Redmond

Xueqing Liu, Yanglei Song, Jiawei Han
University of Illinois at Urbana-Champaign
Top 10 researchers in data mining?  
- And their specializations?

Important research areas in SIGIR conference?
Construct A Topical Hierarchy: Manual or Automated?

MANUAL APPROACH

MANUAL APPROACH

AUTOMATED APPROACH (TOPIC MODELING)

<table>
<thead>
<tr>
<th>Manual Approach</th>
<th>Automated Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality; Labor-intensive</td>
<td>Low human effort; Low quality</td>
</tr>
<tr>
<td><strong>Art</strong></td>
<td><strong>Business</strong></td>
</tr>
<tr>
<td>Movies, Television, Music</td>
<td>Jobs, Real Estate, Investing</td>
</tr>
<tr>
<td>4,249,724 sites</td>
<td></td>
</tr>
<tr>
<td>89,312 editors</td>
<td></td>
</tr>
<tr>
<td>over 1,020,274 categories</td>
<td></td>
</tr>
<tr>
<td>1998-2014</td>
<td></td>
</tr>
<tr>
<td>hierarchical Pachinko Allocation [Mimno 07]</td>
<td>splitLDA [Pujara &amp; Skomoroch 12]</td>
</tr>
</tbody>
</table>

*DMOZ* open directory project
Interactive Approach: Arm Human Curators with Automated Operators

- A human curator runs an operator
- The operator returns an initial hierarchy

Create a hierarchy

- Computer Science
  - Information
    - Database
    - Theory
  - Information System
  - Information Theory
Interactive Approach: Arm Human Curators with Automated Operators (cont’d)

- A human curator runs an operator
- The operator returns a modified hierarchy

Modify the hierarchy

Modify the hierarchy

Information System

Information Theory

Database

Computer Science

Information technology & system

Theory of computation

Information theory
Interactive Approach: Arm Human Curators with Automated Operators (cont’d)

- A human curator runs an operator
- The operator returns a modified hierarchy

Expand the hierarchy

- Computer Science
  - Information technology & system
  - Theory
    - Theory of computation
    - Information theory
  - Database
    - Information retrieval

- Computer Science
  - Information technology & system
  - Theory
    - Theory of computation
    - Information theory
Challenge: Consistency & Efficiency

- **Single-run consistency**
  The generative models before and after an operator should be equivalent except for a few affected nodes.

- **Multi-run consistency**
  Returned hierarchy should be (nearly) identical with identical input.

- **Efficiency**
  Small # data scans

- Traditional inference methods produce large variance across runs
- Traditional inference methods require hundreds to thousands of iterations to converge (no guarantee)
Our Solution:
Scalable Recursive Tensor Decomposition

- **Single-run consistency**
  The generative models before and after an operator should be equivalent except for a few affected nodes

- **Multi-run consistency**
  Returned hierarchy should be (nearly) identical with identical input

- **Efficiency**
  Small # data scans

- A new hierarchical topic model that supports consistent manipulation operators

- A moment-based inference method with theoretical bound of output variance

- Only requires 3 scans of data
Latent Dirichlet Allocation with Topic Tree

\[
\begin{align*}
\theta & \rightarrow z_1 \rightarrow \ldots \rightarrow z_n \\
\alpha & \rightarrow \theta \\
\phi & \rightarrow w \\
\end{align*}
\]

Topic distributions

Word distributions

Dirichlet prior

#words in \(d\)

#docs
To generate a token in document $d$:

1. **Sample a topic path** $z_1 \rightarrow \cdots z_h$ according to $\theta_d$
2. Sample a word $w$ according to $\phi_{z_h}$
Atomic Operators

- **EXP\((t, k)\)**
  Discover \(k\) subtopics of a leaf topic \(t\)

- **MER\((t_1, t_2)\)**
  Merge two topics

- **MOV\((t_1, t_2)\)**
  Move the subtree rooted at \(t_1\) to be under \(t_2\)

These 3 atomic operators are sufficient

**Consistency condition**
For each unaffected leaf node, \(\alpha\) and \(\phi\) remain unchanged
For each internal node \(t\),

\[
\alpha = \sum_{c \text{ is } t' \text{'s child}} \alpha_c
\]
Implementation of Operators

- Decompose *moments* (expectation of patterns) for EXP
  
<table>
<thead>
<tr>
<th></th>
<th>AB: 0.001</th>
<th>ABC: 0.001</th>
<th>DEF: 0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: 0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B: 0.01</td>
<td>BC: 0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C: 0.04</td>
<td>AC: 0.003</td>
<td>GHI: 0.004</td>
<td></td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td></td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td></td>
</tr>
</tbody>
</table>

- Leverage the special property of the moments (*sparse, low rank, and decoupled decomposition*) for scale up

- Manipulate moments efficiently for MER and MOV
Tensor Orthogonal Decomposition for \(\text{EXP}\) operator

**Theorem.** The patterns up to length 3 are sufficient for \(\text{EXP}(t, k)\)

\[
M_2(t) = \sum_{j=1}^{k} \lambda_j \phi_{t/j} \otimes \phi_{t/j}, \quad M_3 = \sum_{j=1}^{k} \lambda_j \phi_{t/j} \otimes \phi_{t/j} \otimes \phi_{t/j},
\]

\(V\): vocabulary size; \(k\): subtopic number

- **length 1**
  - computing: 0.03
  - machinery: 0.01
  - intelligence: 0.04

- **length 2** (pair)
  - computing machinery: 0.001
  - computing intelligence: 0.002
  - machinery intelligence: 0.003

- **length 3** (triple)
  - computing machinery intelligence: 0.001
  - support vector machines: 0.005
  - conditional random fields: 0.004
Tensor Orthogonal Decomposition for $\text{EXP}$ operator

- **Input corpus**: Topic $t$

  Normalized pattern counts:
  
<table>
<thead>
<tr>
<th>A</th>
<th>AB</th>
<th>ABC</th>
<th>B</th>
<th>BC</th>
<th>ABD</th>
<th>C</th>
<th>AC</th>
<th>BCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.03</td>
<td>0.001</td>
<td>0.001</td>
<td>0.01</td>
<td>0.002</td>
<td>0.005</td>
<td>0.04</td>
<td>0.003</td>
<td>0.004</td>
</tr>
</tbody>
</table>

  $V$: vocabulary size
  $k$: subtopic number

- **Step 1**: eigen-decomposition

- **Step 2**: tensor product

- **Step 3**: power iteration

**Information retrieval system**

- 0.3
- 0.2
- 0.1
- 0.05

**Database**

- 0.1
- 0.05
Tensor Orthogonal Decomposition for $EXP$ operator – Not Scalable

Normalized pattern counts

<table>
<thead>
<tr>
<th></th>
<th>A: 0.03</th>
<th>AB: 0.001</th>
<th>ABC: 0.001</th>
<th>B: 0.01</th>
<th>BC: 0.002</th>
<th>ABD: 0.005</th>
<th>C: 0.04</th>
<th>AC: 0.003</th>
<th>BCD: 0.004</th>
</tr>
</thead>
</table>

$V$: vocabulary size; $k$: subtopic number
$L$: # tokens; $l$: average doc length

Prohibitive to compute

Time: $O(V^3 k + Ll^2)$
Space: $O(V^3)$
Scalable Tensor Orthogonal Decomposition

Normalized pattern counts

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>AB</th>
<th>ABC</th>
<th>B</th>
<th>BC</th>
<th>ABD</th>
<th>C</th>
<th>AC</th>
<th>BCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.03</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.01</td>
<td>0.002</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.04</td>
<td>0.003</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sparse & low rank

# nonzero \( m \ll V^2 \)

Decouple

Speed-up construction

Time: \( O(Lk^2 + km) \)

Space: \( O(m) \)
Find Eigen- Decomposition of $M_2$

Step 1: Eigen-Decomp. of A Sparse Matrix

$$M_2 = E_2 - c_1 E_1 \otimes E_1 = U_1 \tilde{M}_2 U_1^T \in \mathbb{R}^{V \times V}$$

$E_2$ (Sparse)

<table>
<thead>
<tr>
<th>AB: 0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC: 0.002</td>
</tr>
<tr>
<td>AC: 0.003</td>
</tr>
<tr>
<td>:</td>
</tr>
</tbody>
</table>

Time: $O(km)$

Space: $O(m)$
Find Eigen- Decomposition of $M_2$

Step 1. Eigen-decomposition of $E_2$

$\Rightarrow (M_2 = U_1 \tilde{M}_2 U_1^T)$

Step 2. Eigen-decomposition of $\tilde{M}_2$ (small)

$M_2 = (U_1 U_2) \Sigma (U_1 U_2)^T = M \Sigma M^T$
Construction of Small Tensor Via Decoupling of Third-Order Moment

\[ \tilde{T} = M_3(W, W, W) \]

\[ W = \Sigma^{-\frac{1}{2}}, W^T M_2 W = I \]

\[
(v \otimes^3)(W, W, W) = (W^T v) \otimes^3 \\
(v \otimes E_2)(W, W, W) = W^T v \otimes W^T E_2 W
\]

Time: \(O(Lk^2)\) 
Space: \(O(Vk)\)
Efficiency

- Several orders of magnitude faster
- Three scans vs. thousands of scans

<table>
<thead>
<tr>
<th>#Tokens</th>
<th>#Docs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBLP title</td>
<td>11M</td>
</tr>
<tr>
<td>CS abstract</td>
<td>39M</td>
</tr>
<tr>
<td>TREC AP news</td>
<td>19M</td>
</tr>
<tr>
<td>PubMed abstract</td>
<td>169M</td>
</tr>
</tbody>
</table>

STROD – Scalable tensor orthogonal decomposition
CATHY – EM algorithm for network-based clustering
splitLDA – Recursively apply LDA
hPAM – hierarchical Pachinko Allocation Model
nCRP – nested Chinese Restaurant Process
Consistency & Quality

- Variance is almost 0
- Convergence is fast
- Good performance in topic intrusion study

(a) Variance (avg KL-divergence) across 10 runs

<table>
<thead>
<tr>
<th>Method</th>
<th>DBLP title</th>
<th>CS abstract</th>
<th>TREC AP news</th>
<th>Pubmed abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>hPAM</td>
<td>5.58</td>
<td>5.72</td>
<td>5.89</td>
<td>Too slow</td>
</tr>
<tr>
<td>splitLDA</td>
<td>3.39</td>
<td>1.60</td>
<td>1.58</td>
<td>Too slow</td>
</tr>
<tr>
<td>CATHY</td>
<td>17.3</td>
<td>1.96</td>
<td>1.42</td>
<td>3.12</td>
</tr>
<tr>
<td>STROD</td>
<td><strong>0.611</strong></td>
<td><strong>0.000138</strong></td>
<td><strong>0.00452</strong></td>
<td><strong>0.000527</strong></td>
</tr>
</tbody>
</table>

(b) Variance w.r.t # iterations

(c) Quality

![Graph showing variance and runtime](image)
Phrase-Represented Hierarchy Sample
Conclusion

1. Interactive operators help topic hierarchy curators; the challenge is consistency and efficiency

2. A solution based on scalable tensor recursive orthogonal decomposition:
   • A new hierarchical topic model that supports consistent operators
   • One operator requires at most three scans of the whole corpus
   • Fast runtime, Low variance, and high quality
References


2. [Li & McCallum 06] W. Li, A. McCallum. Pachinko allocation: Dag-structured mixture models of topic correlations, ICML’06.


Code and data are available at http://illimine.cs.uiuc.edu/software/strod/