Latent Semantic Indexing Of Online Content

Nayan Sharma
Basic Concept

- SVD (Singular Value Decomposition)
  - Identifies Patterns in Relationships

- Construct TDM
- Perform SVD (Too mathematical)
- Use matrix to identify concepts
Understanding LSI

- Definition:
  - Latent Semantic Indexing
  - The underlying meanings of words sorted into things.

- Focus is on relation and not density of words
- Analyses all the words and makes a decision
- Results: Best and most informative content retrieved
### Old IR Problem

**Query keyword:** vehicle  
**Document Retrieved:** A

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>...bought a new <strong>vehicle</strong>...</td>
<td>...bought a new <strong>automobile</strong>...</td>
</tr>
<tr>
<td>....drove her <strong>vehicle</strong>....</td>
<td>...drove her <strong>automobile</strong>...</td>
</tr>
<tr>
<td>...<strong>vehicle</strong> number....</td>
<td>...<strong>automobile</strong> number</td>
</tr>
</tbody>
</table>
LSI has a solution
Why to utilize LSI?

- Provides defense against ‘Keyword Stuffing’

- Targets Synonymy and Polysemy

- Better results and best ranked pages
Polysemy and Context

- Document similarity on single word level: polysemy and context

"... Cell phone ...

... Apple ...

meaning 1

meaning 2

 contribution to similarity, if used in 1st meaning, but not if in 2nd

iPhone
iOS
... itunes

Apple tart
Pie
Shakes
Evaluating Web Content

- Search engines use “spiders” to crawl the contents of your site’s pages.
- Algorithms look for the words and phrases most representative of the content.
- The page title sets the tone for the content.
- The meta description, headline, subheads and content should match the established topic.
LSI for Ranking

- Better ranking with LSI Keywords
  - Search engines look for LSI Keywords
  - Determines quality of content
  - Moderate keywords and more LSI variations
Future Work

- Better efficiency is required
- Need to focus on ‘Word Order’
- Cross Media Image Retrieval
- 3D Human Document Retrieval using LSI
  - Categorize Motion types into frames
  - Extract Motion Characteristics
References


Questions?