The Expected Determinant of the Random Gram Matrix and its Application to Information Retrieval Systems

Jacob Martin

October 22, 2007

Introduction

- **Latent Semantic Indexing** is a method of using **Singular Value Decomposition** to approximate high dimensional spaces for information retrieval.
- **Selective Latent Semantic Indexing** is an improvement to Latent Semantic Indexing based on random gram determinants that Rod Canfield and I developed.

Introduction

- **Latent Semantic Indexing** is a method of using **Singular Value Decomposition** to approximate high dimensional spaces for information retrieval.
- **Selective Latent Semantic Indexing** is an improvement to Latent Semantic Indexing based on random gram determinants that Rod Canfield and I developed.

Selective Latent Semantic Indexing

Efficiently estimates the singular values based on

- probability models, and
- matrix representations

in order to

- compute singular values faster, and
- improve information retrieval.

Outline

- Background
- Theoretical Results
- Computational Results

Outline

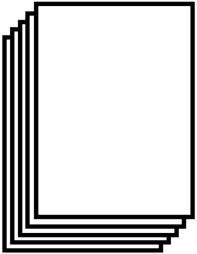
- Background
- Theoretical Results
- Computational Results

Outline

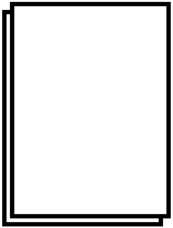
- Background
- Theoretical Results
- Computational Results

Topics

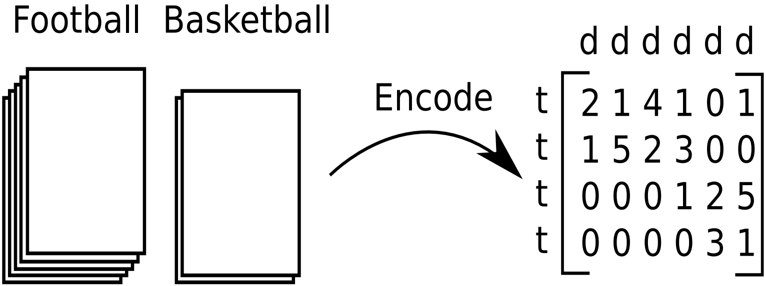
Football



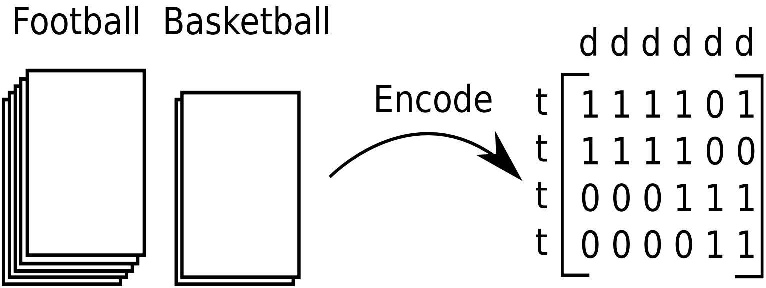
Basketball



Counting Representation



Indicator Representation



Singular Value Decomposition

$$\begin{bmatrix} 2&1&4&1&0&1 \\ 1&5&2&3&0&0 \\ 0&0&0&1&2&5 \\ 0&0&0&0&3&1 \end{bmatrix} = \begin{bmatrix} \quad & \quad \end{bmatrix} \begin{bmatrix} \quad & \quad \end{bmatrix} \begin{bmatrix} \quad & \quad \end{bmatrix}$$

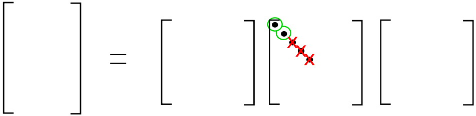
$$A = U \Sigma V^T$$

Singular Values

$$\begin{bmatrix} \quad & \quad \end{bmatrix} = \begin{bmatrix} \quad & \quad \end{bmatrix}\begin{bmatrix} \ddots \end{bmatrix}\begin{bmatrix} \quad & \quad \end{bmatrix}$$

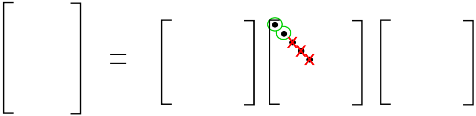
$$A = U(\Sigma)V^T$$

Latent Semantic Indexing



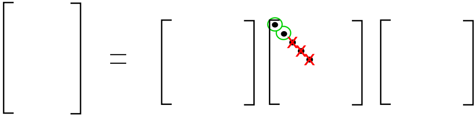
- $A_2 = U_2 \Sigma_2 V_2^T$ is the best approximation among all rank 2 matrices
- Compresses
- Increases Information Retrieval performance (empirically)

Latent Semantic Indexing



- $A_2 = U_2 \Sigma_2 V_2^T$ is the best approximation among all rank 2 matrices
- Compresses
- Increases Information Retrieval performance (empirically)

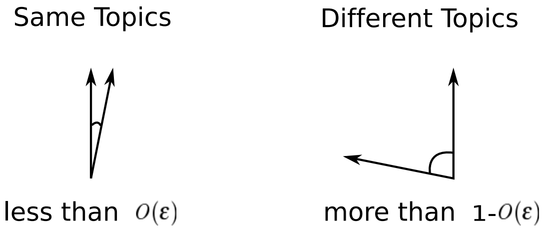
Latent Semantic Indexing



- $A_2 = U_2 \Sigma_2 V_2^T$ is the best approximation among all rank 2 matrices
- Compresses
- Increases Information Retrieval performance (empirically)

Defining Success

Measure all *intratopic* (same) and *intertopic* (different) angles in the reduced space to calculate $O(\varepsilon)$**-skewness**



Reduced Rank SVD

$$A_k = U_k \Sigma_k V_k^T$$

- **Question**: How can a rank reduction increase Information Retrieval performance (in terms of skewness)?
- *Use probability models to investigate.*

Reduced Rank SVD

$$A_k = U_k \Sigma_k V_k^T$$

- **Question**: How can a rank reduction increase Information Retrieval performance (in terms of skewness)?
- *Use probability models to investigate.*

Document Generation Model

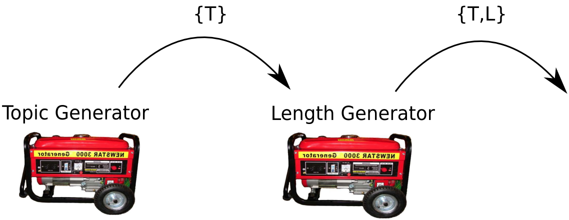
Creating a Document

{T}

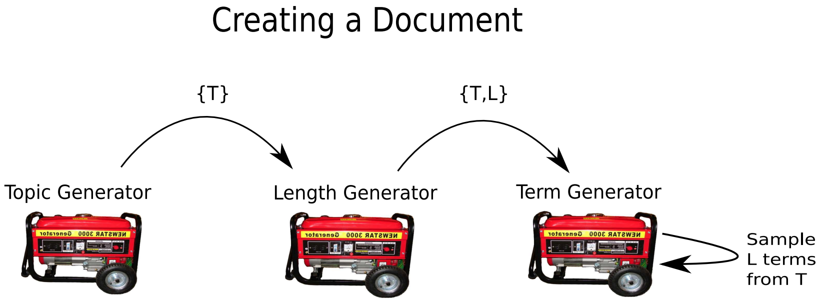
Topic Generator

Document Generation Model

Creating a Document

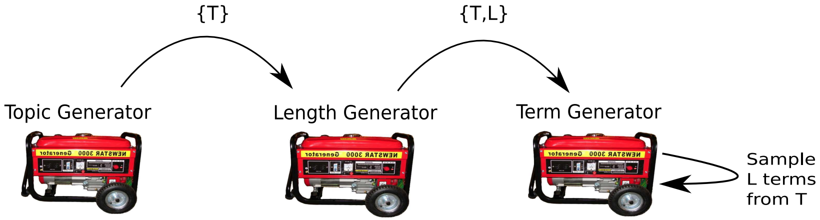


Document Generation Model



Document Generation Model

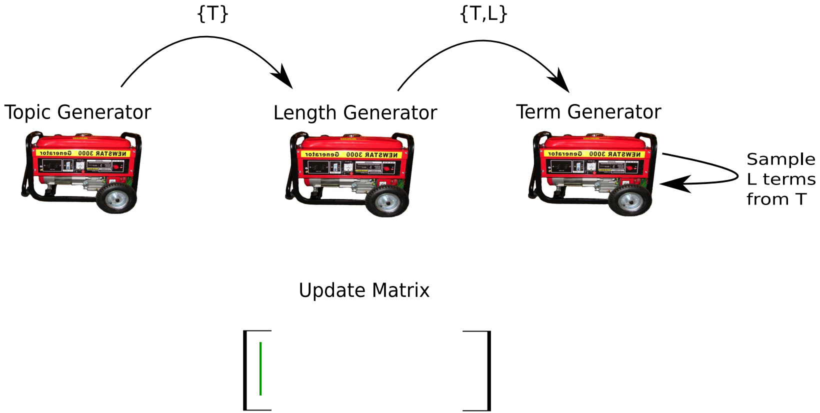
Creating a Document



Update Matrix

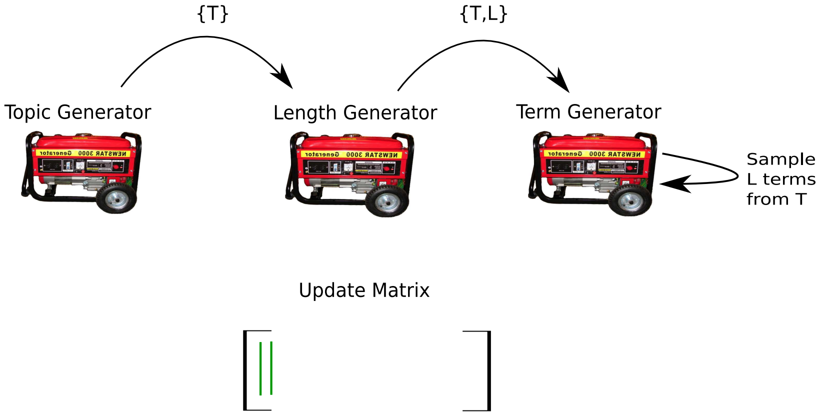
Document Generation Model

Creating a Document

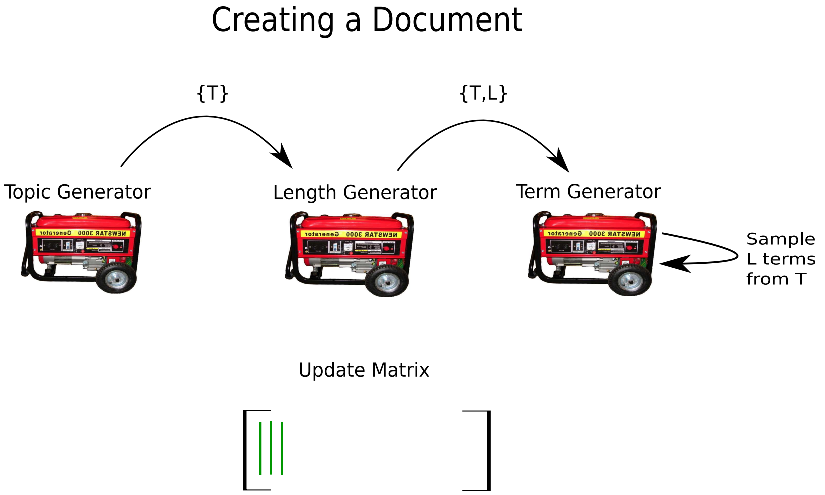


Document Generation Model

Creating a Document



Document Generation Model



Pure Document Generation Models

A model is **pure** when each document is on a single topic

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
t_1	1	1	1	0	0	0	0	0	0
t_2	1	1	1	0	0	0	0	0	0
t_3	1	1	1	0	0	0	0	0	0
t_4	0	0	0	1	1	1	0	0	0
t_5	0	0	0	1	1	1	0	0	0
t_6	0	0	0	1	1	1	0	0	0
t_7	0	0	0	0	0	0	1	1	1
t_8	0	0	0	0	0	0	1	1	1
t_9	0	0	0	0	0	0	1	1	1

$$T_1 = \{t_1, t_2, t_3\}$$

$$T_2 = \{t_4, t_5, t_6\}$$

$$T_3 = \{t_7, t_8, t_9\}$$

Pure Document Generation Models

A model is **pure** when each document is on a single topic

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
t_1	1	1	1	0	0	0	0	0	0
t_2	1	1	1	0	0	0	0	0	0
t_3	1	1	1	0	0	0	0	0	0
t_4	0	0	0	1	1	1	0	0	0
t_5	0	0	0	1	1	1	0	0	0
t_6	0	0	0	1	1	1	0	0	0
t_7	0	0	0	0	0	0	1	1	1
t_8	0	0	0	0	0	0	1	1	1
t_9	0	0	0	0	0	0	1	1	1

Theorem (Papadimitriou *et al.*)

The rank-k SVD is 0-skewed with probability $1 - O(\frac{1}{m})$ when performed on any $t \times m$ matrix drawn from a pure document generation model with k topics.

Pure Document Generation Models

Perturbing the matrix A with $\|E\|_F = O(\varepsilon)$

$$
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1
\end{bmatrix}
+
\begin{bmatrix}
0 & 0 & 0 & 0 & \underline{\mathbf{1}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \underline{\mathbf{1}} & 0 \\
0 & 0 & 0 & \underline{\mathbf{1}} & \underline{\mathbf{1}} & 0 & 0 & 0 & 0 \\
0 & 0 & \underline{\mathbf{1}} & 0 & 0 & 0 & 0 & 0 & \underline{\mathbf{1}} \\
0 & \underline{\mathbf{1}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \underline{\mathbf{1}} & 0 & 0 \\
\underline{\mathbf{1}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \underline{\mathbf{1}} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \underline{\mathbf{1}} & 0 & 0 & 0 & 0
\end{bmatrix}
$$

Pure Document Generation Models

$$\begin{bmatrix}
1 & 1 & 1 & 0 & \underline{\mathbf{1}} & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & \underline{\mathbf{1}} & 0 \\
1 & 1 & 1 & \underline{\mathbf{1}} & \underline{\mathbf{1}} & 0 & 0 & 0 & 0 \\
0 & 0 & \underline{\mathbf{1}} & 1 & 1 & 1 & 0 & 0 & \underline{\mathbf{1}} \\
0 & \underline{\mathbf{1}} & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & \underline{\mathbf{1}} & 0 & 0 \\
\underline{\mathbf{1}} & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & \underline{\mathbf{1}} & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & \underline{\mathbf{1}} & 0 & 1 & 1 & 1
\end{bmatrix}$$

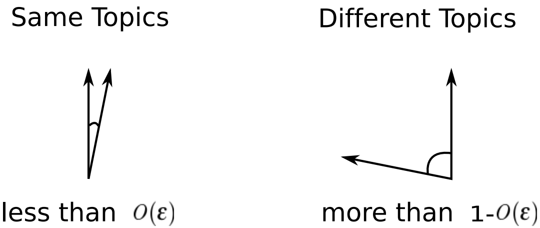
Theorem (Papadimitriou *et al.*)

If A is created from a pure model and $\|E\|_F = O(\varepsilon)$ then the rank-k SVD on $A+E$ is $O(\varepsilon)$-skewed

Papadimitriou *et al.*

Theorem (Papadimitriou *et al.*)

If A is created from a pure model and $\|E\|_F = O(\varepsilon)$ then the rank-k SVD on $A+E$ is $O(\varepsilon)$-skewed



However, things can go wrong...

$$\left[\begin{array}{cc|cccccccc} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 12 & 0 & 3 & 22 & 11 & 21 & 14 & 16 \\ 0 & 0 & 0 & 5 & 60 & 8 & 7 & 0 & 20 & 41 \end{array}\right]$$

$$T_2's \text{ eigenvalues} = \{\mathbf{6228.43}, \mathbf{1241.57}\}$$

$$T_1's \text{ eigenvalues} = \{2.62, 0.38\}$$

For a rank$-\mathbf{2}$ decomposition, all singular values of T_1 are outweighed by the largest two singular values of T_2!

However, things can go wrong, even when we normalize the matrix!

$$\left[\begin{array}{cc|cccccccc} 0.71 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.71 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ \hline 0.00 & 0.00 & 1.00 & 0.00 & 0.05 & 0.94 & 0.84 & 1.00 & 0.57 & 0.36 \\ 0.00 & 0.00 & 0.00 & 1.00 & 1.00 & 0.34 & 0.54 & 0.00 & 0.82 & 0.93 \end{array}\right]$$

$$T_2's \text{ eigenvalues} = \{\mathbf{2.4}, \mathbf{1.5}\}$$

$$T_1's \text{ eigenvalues} = \{1.3, 0.5\}$$

For a rank$-\mathbf{2}$ decomposition, all singular values of T_1 are outweighed by the largest two singular values of T_2!

Counterexamples to the theorem

Singular values depend on:

- topic probabilities,
- document lengths,
- topic sizes,
- representation!

Counterexamples to the theorem

Singular values depend on:

- topic probabilities,
- document lengths,
- topic sizes,
- representation!

Counterexamples to the theorem

Singular values depend on:

- topic probabilities,
- document lengths,
- topic sizes,
- representation!

Counterexamples to the theorem

Singular values depend on:

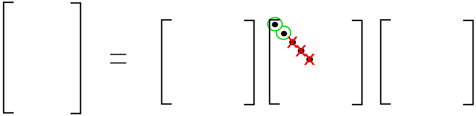
- topic probabilities,
- document lengths,
- topic sizes,
- representation!

How to fix this problem?

- Calculate the expected singular values according to the document generation model's topics

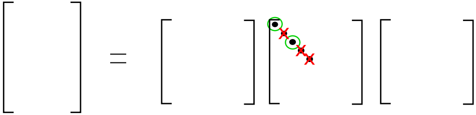
How to fix this problem?

So, instead of retaining the largest



How to fix this problem?

Predict and **select** the singular values *according to each topic*.



Expectation[1]

Predict singular values using the expectation operator E (which is linear).

$$E(X+Y) = E(X) + E(Y)$$

$$E(cX) = cE(X)$$

[1] In probability theory the expected value of a discrete random variable is the sum of the probability of each possible outcome of the experiment multiplied by the outcome value.

Singular Values

- $$A = U\Sigma V^T$$
- $$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

Singular Values

- $$A = U\Sigma V^T$$
- $$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

Singular Values

- $$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$
- $$\begin{aligned} \{\sigma_i\} &= \sqrt{\text{eigenvalues}(AA^T)} \\ &= \sqrt{\text{eigenvalues}(A^TA)} \\ &= \sqrt{\text{rootsof}\left(\det(A^TA - \lambda I)\right)} \end{aligned} \quad (1)$$

Singular Values

- $$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

- $$\begin{aligned} \{\sigma_i\} &= \sqrt{\text{eigenvalues}(AA^T)} \\ &= \sqrt{\text{eigenvalues}(A^TA)} \\ &= \sqrt{\text{rootsof}\left(\det(A^TA - \lambda I)\right)} \end{aligned} \qquad (1)$$

Singular Values

- $$\{\sigma_i\} = \sqrt{\text{rootsof}\left(\det(A^T A - \lambda I)\right)}$$
- $$\det(A^T A - \lambda I) = \begin{vmatrix} (d_1 \cdot d_1 - \lambda) & & \\ & \ddots & \\ & & (d_m \cdot d_m - \lambda) \end{vmatrix}$$
- $$(-\lambda)^m + (-\lambda)^{m-1} \sum_{i=1}^{m} A^T A_{ii} \cdots + (-\lambda)^{m-t} \sum_{t \times t \text{ psm } \alpha} \det(\alpha)$$

Singular Values

- $$\{\sigma_i\} = \sqrt{\text{rootsof}\left(\det(A^TA - \lambda I)\right)}$$
- $$\det(A^TA - \lambda I) = \begin{vmatrix} (d_1 \cdot d_1 - \lambda) & & \\ & \ddots & \\ & & (d_m \cdot d_m - \lambda) \end{vmatrix}$$
- $$(-\lambda)^m + (-\lambda)^{m-1} \sum_{i=1}^{m} A^TA_{ii} \cdots + (-\lambda)^{m-t} \sum_{t\times t \text{ psm } \alpha} \det(\alpha)$$

Singular Values

- $$\{\sigma_i\} = \sqrt{\text{rootsof}\left(\det(A^T A - \lambda I)\right)}$$

- $$\det(A^T A - \lambda I) = \begin{vmatrix} (d_1 \cdot d_1 - \lambda) & & \\ & \ddots & \\ & & (d_m \cdot d_m - \lambda) \end{vmatrix}$$

- $$(-\lambda)^m + (-\lambda)^{m-1} \sum_{i=1}^{m} A^T A_{ii} \cdots + (-\lambda)^{m-t} \sum_{t \times t \text{ psm } \alpha} \det(\alpha)$$

Random Gram[2] Matrices

Let w be a t-tall vector whose components w_i are random variables (not necessarily independent)

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_t \end{bmatrix} .$$

Next, sample n independent vectors $w^{(1)}, \ldots, w^{(n)}$; creating a $t \times n$ matrix B. Then the random gram matrix of size n is

$$G_n = B^T B \, ,$$

and has a distribution that depends on the underlying distribution of the random vector w.

[2] Jørgen Pedersen Gram (1850-1916) was a Danish Mathematician

Singular Values

- $$\det(A^TA-\lambda I)=(-\lambda)^m+(-\lambda)^{m-1}\sum_{i=1}^{m}A^TA_{ii}\cdots+(-\lambda)^{m-t}\sum_{t\times t\ \text{psm}\ \alpha}\det(\alpha)$$

- $$a_i=E(\det(G_i))$$

- $$E(\det(A^TA-\lambda I))=$$
  $$a_0\binom{m}{0}(-\lambda)^m+a_1\binom{m}{1}(-\lambda)^{m-1}+a_2\binom{m}{2}(-\lambda)^{m-2}+\cdots+a_t\binom{m}{t}(-\lambda)^{m-t}$$

Singular Values

- $$\det(A^TA-\lambda I)=(-\lambda)^m+(-\lambda)^{m-1}\sum_{i=1}^{m}A^TA_{ii}\cdots+(-\lambda)^{m-t}\sum_{t\times t \text{ psm } \alpha}\det(\alpha)$$
- $$a_i=E(\det(G_i))$$
- $$E(\det(A^TA-\lambda I))=$$
$$a_0\binom{m}{0}(-\lambda)^m+a_1\binom{m}{1}(-\lambda)^{m-1}+a_2\binom{m}{2}(-\lambda)^{m-2}+\cdots+a_t\binom{m}{t}(-\lambda)^{m-t}$$

Singular Values

- $$\det(A^TA-\lambda I)=(-\lambda)^m+(-\lambda)^{m-1}\sum_{i=1}^{m}A^TA_{ii}\cdots+(-\lambda)^{m-t}\sum_{t\times t\text{ psm }\alpha}\det(\alpha)$$
- $$a_i=E(\det(G_i))$$
- $$E(\det(A^TA-\lambda I))=$$
$$\mathrm{a}_0\binom{m}{0}(-\lambda)^m+a_1\binom{m}{1}(-\lambda)^{m-1}+a_2\binom{m}{2}(-\lambda)^{m-2}+\cdots+a_t\binom{m}{t}(-\lambda)^{m-t}$$

E(det(G_n)) Recursion

$$a_0\binom{m}{0}(-\lambda)^m + a_1\binom{m}{1}(-\lambda)^{m-1} + a_2\binom{m}{2}(-\lambda)^{m-2} + \cdots + a_t\binom{m}{t}(-\lambda)^{m-t}$$

GOAL: efficient recursive calculation for the expected characteristic coefficients $a_n = E(det(G_n))$

Characteristic coefficients:

- are used in computational chemistry
- give an algebraic solution of the inverse metric problem in general relativity.

E$(\det(G_n))$ Recursion

$$a_0\binom{m}{0}(-\lambda)^m + a_1\binom{m}{1}(-\lambda)^{m-1} + a_2\binom{m}{2}(-\lambda)^{m-2} + \cdots + a_t\binom{m}{t}(-\lambda)^{m-t}$$

GOAL: efficient recursive calculation for the expected characteristic coefficients $a_n = E(det(G_n))$
Characteristic coefficients:

- are used in computational chemistry
- give an algebraic solution of the inverse metric problem in general relativity.

Generating Functions

A generating function is a clothesline on which we hang up a sequence of numbers for display. [3]

What is the generating function for a sequence of n ones?

$$1, 1, 1, 1, \ldots$$

$$\frac{1-x^n}{1-x} = 1 + x^1 + x^2 + x^3 + \cdots + x^{n-1}$$

$$\begin{aligned} 1 - x^n &= (1-x)(1 + x^1 + x^2 + x^3 + \cdots + x^{n-1}) \\ &= (1 + x^1 + x^2 + \cdots + x^{n-1}) - (x^1 + x^2 + x^3 + \cdots + x^n) \\ &= 1 - x^n \end{aligned}$$

So,

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \qquad (|x| < 1)$$

[3] Herbert Wilf. Generatingfunctionology (2004)

$E(\det(G_n))$ Exponential Generating Function

Based on Euler's[4] infinite expansion

$$\exp\{x\} = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad ,$$

and the Leibniz[5] formula for the determinant,

$$\det(G_n) = \sum_{\sigma \in S_n} (-1)^{\text{sgn}(\sigma)} \prod_{i=1}^{n} (G_n)_{i,\sigma(i)} \quad ,$$

we have proven the following exponential generating function equation for $a_n = E(det(G_n))$

$$\sum_{n=0}^{\infty} a_n \frac{x^n}{n!} = \exp\left\{ \frac{c_1 x}{1} - \frac{c_2 x^2}{2} + \frac{c_3 x^3}{3} - \dots \right\}. \qquad (2)$$

(the c_i's will be defined shortly)

[4] Leonhard Euler (1707-1783). A Swiss Mathematician and Physicist

[5] Gottfried Leibniz: (1646-1716) A German Mathematician and Natural Philosopher

$E(\det(G_n))$ Recursion

$$\sum_{n=0}^{\infty} a_n \frac{x^n}{n!} = \exp\left\{ \frac{c_1 x}{1} - \frac{c_2 x^2}{2} + \frac{c_3 x^3}{3} - \dots \right\}. \quad (3)$$

Differentiate both sides to help discover a recursion...

$$\sum_{n=0}^{\infty} a_{n+1} \frac{x^n}{n!} = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!} \left(c_1 - c_2 x + c_3 x^2 - \dots \right) . \quad (4)$$

By equating the coefficients of $\left[\frac{x^n}{n!}\right]$ on the left and right hand side, we find:

$$a_0 = 1$$

$$a_{n+1} = \sum_j \binom{n}{j} (-1)^j j! a_{n-j} c_{j+1}$$

What about the c_i???

Define,

$$M_{ij} = E(w_i w_j) \tag{5}$$

Then,

$$\begin{aligned} c_j &= \text{trace}(M^j) \\ &= \lambda_1^j + \cdots + \lambda_t^j \end{aligned}$$

Representations

$$\begin{array}{c|cccccc|} & d & d & d & d & d & d \\ t & 1 & 1 & 1 & 1 & 0 & 1 \\ t & 1 & 1 & 1 & 1 & 0 & 0 \\ t & 0 & 0 & 0 & 1 & 1 & 1 \\ t & 0 & 0 & 0 & 0 & 1 & 1 \end{array} \quad \text{vs} \quad \begin{array}{c|cccccc|} & d & d & d & d & d & d \\ t & 2 & 1 & 4 & 1 & 0 & 1 \\ t & 1 & 5 & 2 & 3 & 0 & 0 \\ t & 0 & 0 & 0 & 1 & 2 & 5 \\ t & 0 & 0 & 0 & 0 & 3 & 1 \end{array}$$

$$c_j = \text{trace}(M^j)$$

$$a_0 = 1$$

$$a_{n+1} = \sum_j \binom{n}{j} (-1)^j j! a_{n-j} c_{j+1}$$

To get the expected characteristic coefficients, we just need to create $M_{ij} = E(w_i w_j)$ according to the representation used.

$0, 1$ indicator function

$$A_{ij} = \begin{cases} 0, & \text{if } term_i \text{ is not present in } doc_j \\ 1, & \text{if } term_i \text{ is present in } doc_j \end{cases}$$

$$\begin{aligned} M_{ii} &= \sum_{t_i \geq 1} \dots \\ &= 1 - (1 - p_i)^l \\ M_{ij} &= \sum_{\substack{t_i \geq 1 \\ t_j \geq 1}} \\ &= \sum_{\substack{t_i \geq 0 \\ t_j \geq 0}} - \sum_{\substack{t_i = 0 \\ t_j \geq 0}} - \sum_{\substack{t_i \geq 0 \\ t_j = 0}} - \sum_{\substack{t_i = 0 \\ t_j = 0}} \\ &= 1 - (1 - p_i)^l - (1 - p_j)^l - (1 - p_i - p_j)^l \end{aligned}$$

$-1,1$ indicator function

$$A_{ij} = \begin{cases} -1, & \text{if } \mathit{term}_i \text{ is not present in } \mathit{doc}_j \\ 1, & \text{if } \mathit{term}_i \text{ is present in } \mathit{doc}_j \end{cases}$$

$$M_{ij} = \begin{cases} 1 & \text{if } i = j \\ 1 - 4\left(1 - p_i - p_j\right)^l - 2\left(1 - p_i\right)^l - 2\left(1 - p_j\right)^l & \text{if } i \neq j \end{cases}$$

Counting function

$A_{ij} =$ number of times $term_i$ appears in doc_j

$$M_{ij} = \begin{cases} \sum_l \left(l(l-1)p_i^2 + lp_i\right) prob(l), & i = j \\ \sum_l \left(l(l-1)p_i p_j\right) prob(l), & i \neq j \end{cases}$$

Complexity Sketch

$$a_0\binom{m}{0}(-\lambda)^m + a_1\binom{m}{1}(-\lambda)^{m-1} + a_2\binom{m}{2}(-\lambda)^{m-2} + \cdots + a_t\binom{m}{t}(-\lambda)^{m-t}$$

$$a_0 = 1$$

$$a_{n+1} = \sum_j \binom{n}{j}(-1)^j j! a_{n-j} c_{j+1}$$

- t coefficients to compute
- The $c_j = \text{trace}(M^j)$ can be parallelized (hypercube) down to $O(t)$

Complexity Sketch

Nice properties of our algorithm:

- Simple and easily parallelized
- Doesn't depend on the number of documents (full SVD is $O(m^2n + n^2m + n^3)$ with Golub-Reinsch algorithm)
- Reduced computation of characteristic coefficients to a recursion involving symmetric matrix powers.
- Accuracy increases as more documents are included

Singular Value Calculation Times

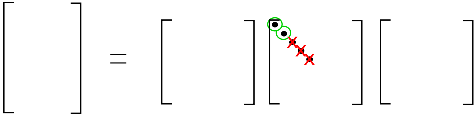
Read matrix, infer model, calculate expected coeffs, get roots

	32 x 500,000	32 x 1,000,000
Maple	580.63 secs	1211.59 secs
Selective	4.10 secs	10.47 secs

Table: Singular value calculation times of two random matrices for both Maple and Selective (not parallelized)

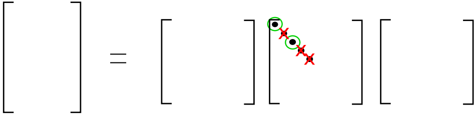
Information Retrieval Comparison

Compare retaining the largest



Information Retrieval Comparison

To **selecting** the singular values *according to each topic*.



Information Retrieval: (Real world sports articles)

	Algorithm	Intratopic Cosine	Intertopic Cosine
Rank 2	**Selective**	**0.998**	**0.037**
	LSI	0.767	0.620
Rank 4	**Selective**	**0.821**	**0.040**
	LSI	0.625	0.433
Rank 8	**Selective**	**0.653**	**0.038**
	LSI	0.673	0.007

Table: Comparison of our algorithm to traditional latent semantic indexing.

Information Retrieval: (Probabilistic documents)

	Algorithm	Intratopic cosine	Intertopic cosine
Rank 2	**Selective**	**1.0**	**0.0**
	LSI	0.46	0.04

Table: Average cosine over 100 different groups of 1000 random documents generated from a probabilistic document generation model containing two topics with probability 99% and 1%.

Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors

Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors

Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors

Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors

Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors

Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors

Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors

Conclusion

- New theoretical results concerning SVD
- More speed
- Better quality reduced rank approximation
- Might help with the “choice–of–rank conundrum”

Conclusion

- New theoretical results concerning SVD
- More speed
- Better quality reduced rank approximation
- Might help with the “choice–of–rank conundrum”

Conclusion

- New theoretical results concerning SVD
- More speed
- Better quality reduced rank approximation
- Might help with the “choice–of–rank conundrum”

Conclusion

- New theoretical results concerning SVD
- More speed
- Better quality reduced rank approximation
- Might help with the “choice–of–rank conundrum”

Contact Information

Jacob Martin
jacobgmartin@gmail.com
`http://www.cs.tcd.ie/Jacob.Martin`