Computing Word Senses by Semantic Mirroring and Spectral Graph Partitioning

Masters Thesis by Martin Fagerlund

Martin Fagerlund, Magnus Merkel, Lars Eldén, Lars Ahrenberg
Departments of Mathematics and Computer Science, Linköping University

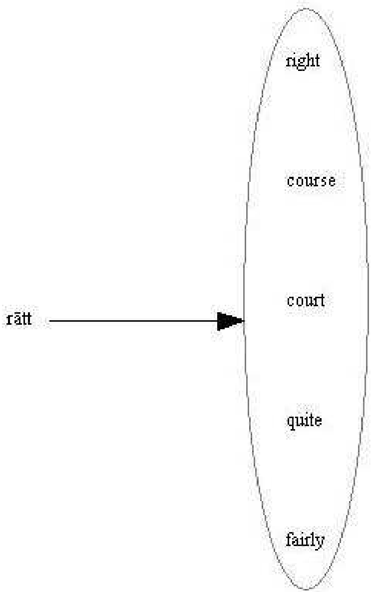
ACL 2010

Outline

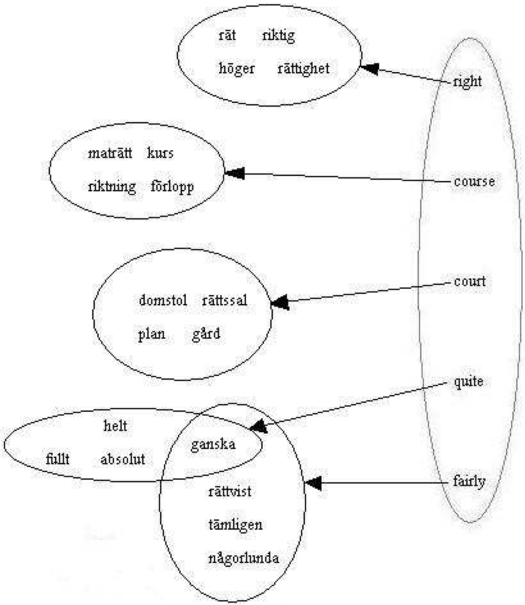
Semantic Mirrors (Dyvik)

- Use translations to extract information from a bilingual lexicon.
- Semantically closely related words tend to have strongly overlapping sets of translations.
- Words with a wide meaning tend to have a higher number of translations, than words with a more narrow meaning.

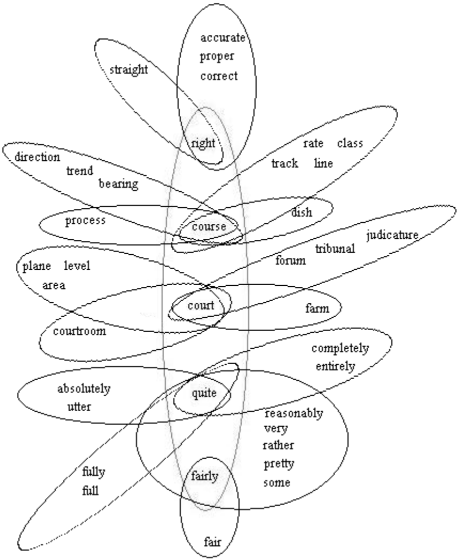
Semantic Mirrors



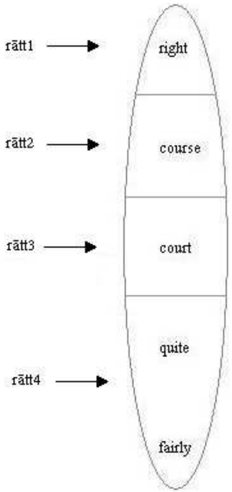
Semantic Mirrors



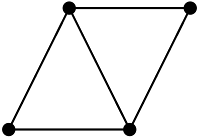
Semantic Mirrors



Semantic Mirrors

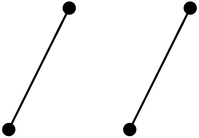


Graph theory



- Connected graph with vertex set V and edge set E.

Graph theory



- Disconnected graph.

Graph theory

- The weight function.
- $w(v_i, v_j) = w(v_j, v_i)$.
- $w(v_i, v_j) \geq 0$.

- The adjacency matrix.

$$A(i,j) = \begin{cases} w(v_i, v_j), & \text{if } v_i \text{ and } v_j \text{ are adjacent,} \\ 0, & \text{otherwise.} \end{cases}$$

Graph theory - Graphs and Matrices

- The Laplacian.

$$d(v_i) = \sum_j w(v_i, v_j).$$

$$D(i,j) = \begin{cases} d(v_i), & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

- $L = D - A$.

Graph theory - Graphs and Matrices

- The Laplacian.

$$d(v_i) = \sum_j w(v_i, v_j).$$

$$D(i,j) = \begin{cases} d(v_i), & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

- $L = D - A$.

Graph theory - Graphs and Matrices

- The normalised Laplacian.

$$\mathcal{L} = D^{-1/2} L D^{-1/2}.$$

$$\mathcal{L} = \begin{cases} 1 - \frac{w(v_i,v_i)}{d(v_i)}, & \text{if } i = j \text{ and } d(v_i) \neq 0, \\ -\frac{w(v_i,v_j)}{\sqrt{d(v_i)d(v_j)}}, & \text{if } v_i \text{ and } v_j \text{ are adjacent,} \\ 0, & \text{otherwise.} \end{cases}$$

Spectral graph partitioning

Compute the eigenvalues

$$0 = \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{n-1},$$

and the eigenvectors

$$\bar{u}_0, \ldots, \bar{u}_{n-1}$$

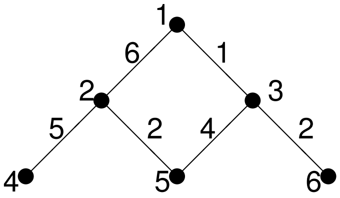
of $\mathcal{L}$.
λ_1 is called the Fiedler value, and $\bar{u}_1$ is called the Fiedler vector.

Spectral graph partitioning

Let $\bar{u} = (u_1, \ldots, u_n)$ be the Fiedler vector.
Divide the vertices into S and $\bar{S}$ using a splitting value s:
$v_i \in S$ if $u_i > s$
$v_i \in \bar{S}$ if $u_i \leq s$.
A few ways of choosing s:

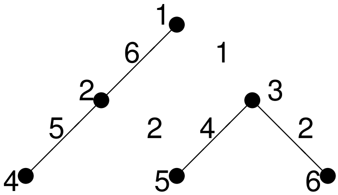
- bisection: s is the median of $\{u_1, \ldots, u_n\}$.
- sign cut: s =0.
- gap cut: s is a value in the largest gap in the sorted Fiedler vector.

Spectral graph partitioning - An example



$$\begin{pmatrix} -0.2523 \\ -0.4183 \\ 0.5701 \\ -0.3603 \\ 0.3571 \\ 0.4232 \end{pmatrix}$$

Spectral graph partitioning - An example



$$\begin{pmatrix} -0.2523 \\ -0.4183 \\ 0.5701 \\ -0.3603 \\ 0.3571 \\ 0.4232 \end{pmatrix}$$

Spectral graph partitioning

Measures of the partitioning. Let

$$d(S) = \sum_{v_i \in S} d(v_i),$$

and

$$|\,E(S, \bar{S})\,| = \sum_{v_i \in S, v_j \in \bar{S}} w(v_i, v_j).$$

Conductance,

$$\phi(S) = d(V)\frac{|\,E(S, \bar{S})\,|}{d(S)d(\bar{S})},$$

Spectral graph partitioning

Measures of the partitioning. Let

$$d(S) = \sum_{v_i \in S} d(v_i),$$

and

$$|E(S, \bar{S})| = \sum_{v_i \in S, v_j \in \bar{S}} w(v_i, v_j).$$

Sparsity,

$$sp(S) = \frac{|E(S, \bar{S})|}{\min(d(S), d(\bar{S}))}$$

Spectral graph partitioning

The conductance of a graph

$$\phi_G = \min_S \phi(S)$$

The sparsity of a graph

$$sp_G = \min_S sp(S)$$

Spectral graph partitioning

The Cheeger inequalities

$$2\phi_G > \lambda_1 \geq \frac{\phi_G^2}{8}.$$

$$2sp_G \geq \lambda_1 \geq \frac{sp_G^2}{2}.$$

Separating word senses - Translation

- English-Swedish lexicon of adjectives.
  Words into two lists.
- Translation matrix **B**.
  Rows correspond to English words.
  Columns correspond to Swedish words.

Separating word senses - Translation

- English word j defines a vector $\bar{e}_j$ by

$$\bar{e}_j(i) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

- $\mathbf{B}^T \bar{e}_j$ gives translations from English to Swedish.

Separating word senses - Computation

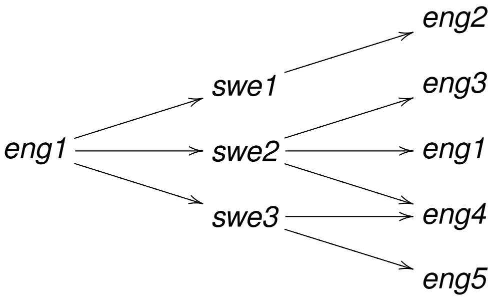
Start with an English word, called *eng1*.

$$\mathbf{B}\mathbf{B}^T \bar{e}_{eng1}.$$

Separating word senses - Computation

Start with an English word, called *eng1*.

$$\mathbf{BB}^T\bar{e}_{eng1}.$$



Separating word senses - Computation

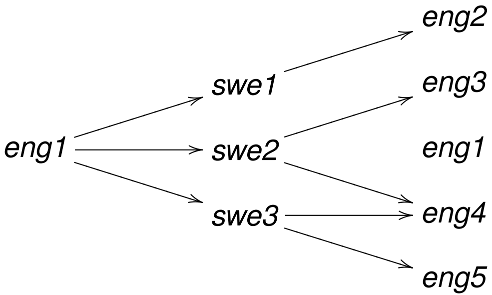
Replace the row in **B** corresponding to *eng*1, with an all-zero row. Call this new matrix $\mathbf{B}_{mod1}$.

$$\mathbf{B}_{mod1}\mathbf{B}^T\bar{e}_{eng1}.$$

Separating word senses - Computation

Replace the row in **B** corresponding to *eng*1, with an all-zero row. Call this new matrix $\mathbf{B}_{mod1}$.

$$\mathbf{B}_{mod1}\mathbf{B}^T\bar{e}_{eng1}.$$

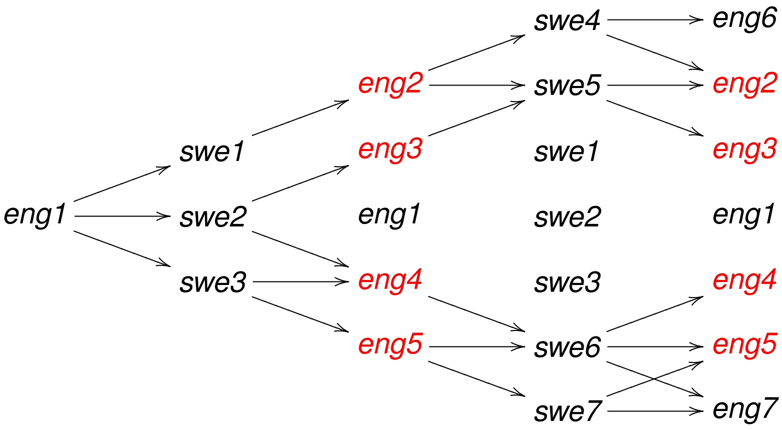


Separating word senses - Computation

Let $\mathbf{B}_{mod2}$ be the matrix **B**, with columns corresponding to the Swedish words *swe1,...,swe3*, replaced with all-zero columns.
Then

$$\mathbf{B}_{mod2}\mathbf{B}_{mod2}^{T}\mathbf{E}$$

Separating word senses - Computation



Separating word senses - Computation

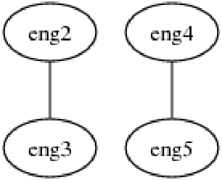
If in

$$\mathbf{B}_{mod2}\mathbf{B}_{mod2}^{T}\mathbf{E}$$

keeping only the rows corresponding to *eng2,...eng5*, we get a symmetric matrix **A.**

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

Separating word senses - Computation



Separating word senses - Computation

- Create $\mathcal{L}$, and compute the Fiedler vector.
- Sort the Fiedler vector, and make $n - 1$ cuts.
- For each cut, compute the conductance and choose the the cut that produce the lowest value.
- Partition the graph.

Separating word senses - Computation

- Continue and partition the largest of the subgraphs obtained.
- End the iteration when a stopping criterion is fulfilled.

Separating word senses - Example 1

Global

absolute	full-scale	round
aggregate	intact	teetotal
all-out	integral	total
clear	integrate	unbroken
complete	international	universal
entire	mondial	utter
full	one-piece	whole
full-length	outright	worldwide

Separating word senses - Example 1

Global

intact whole full-length one-piece clear round full	absolute utter complete all-out teetotal outright entire total	integrate unbroken integral	aggregate full-scale
worldwide international	mondial	universal	

Separating word senses - Example 2

Visiting

adventitious
alien
exotic
foreign
outlandish
strange
unaccustomed
uncouth
unfamiliar
ungenial

Separating word senses - Example 2

Visiting

alien foreign outlandish exotic	unaccustomed unfamiliar strange
adventitious	uncouth
ungenial	

Evaluation and Results

Evaluation

- 10 random words with at least 8 vertices in the graph were evaluated.
- Both conductance and sparsity were used as a measure.

Results

- The sense groups are not always completely synonymous with the seed word.
- The words are often consistent within the sense groups.

Evaluation and Results

Evaluation

- 10 random words with at least 8 vertices in the graph were evaluated.
- Both conductance and sparsity were used as a measure.

Results

- The sense groups are not always completely synonymous with the seed word.
- The words are often consistent within the sense groups.

Conclusions and future work

Conclusion

Our preliminary results indicate that graph partitioning-based semantic mirroring can be developed into an automatic method for computing word senses.

Future work

- automatic treatment of homonyms
- Tests on other dictionaries, languages
- Systematic evaluation of the method

To get a copy of Martin Fagerlund's Masters thesis, send me or him a message.