Merging the Meaning: An N-gram Graph Approach

George Giannakopoulos[1,2]

[1]NCSR Demokritos, Greece
ggianna@disi.unitn.it

[2]University of Trento, Italy

16/11/2010 - Text Analysis Conference 2010

Outline

The problem

Given *a set of model summaries*, determine the *quality* (Responsiveness, Pyramid Score) of a given *peer* summary text.

The problem

Given *a set of model summaries*, determine the *quality* (Responsiveness, Pyramid Score) of a given *peer* summary text.

Proposals using N-Gram Graphs

The problem

Given *a set of model summaries*, determine the *quality* (Responsiveness, Pyramid Score) of a given *peer* summary text.

Proposals using N-Gram Graphs

AutoSummENG Average similarity between models and peer.

The problem

Given *a set of model summaries*, determine the *quality* (Responsiveness, Pyramid Score) of a given *peer* summary text.

Proposals using N-Gram Graphs

AutoSummENG Average similarity between models and peer.

MeMoG Similarity between merged model and peer.

The problem

Given *a set of model summaries*, determine the *quality* (Responsiveness, Pyramid Score) of a given *peer* summary text.

Proposals using N-Gram Graphs

- **AutoSummENG** Average similarity between models and peer.
- **MeMoG** Similarity between merged model and peer.
- **HPG** Average similarity between a model hierarchy of graphs to peer.

Outline

Overview of AutoSummENG

- Represents peer and model summaries as n-gram graphs
- Compare each peer graph to model graphs
- Grade is the average similarity

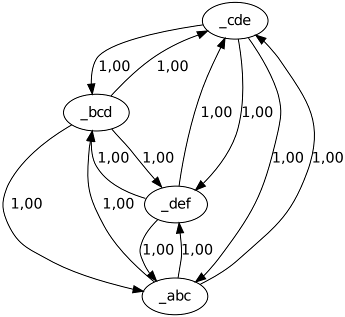
AutoSummENG method [Giannakopoulos et al., 2008]: DUC 2005-2007, TAC 2008-2009

Overview of AutoSummENG

- Represents peer and model summaries as n-gram graphs
- Compare each peer graph to model graphs
- Grade is the average similarity
- **No Preprocessing, Language Neutrality, Effective**

AutoSummENG method [Giannakopoulos et al., 2008]: DUC 2005-2007, TAC 2008-2009

An N-gram Graph



What does an n-gram graph do?

- Indicates neighborhood
- **Edges** are important

Extraction Process

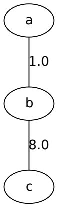
- Extract n-grams of ranks $[L_{\text{min}}, L_{\text{MAX}}]$. One graph per rank.
- Determine neighborhood (window size D_{win}).
- Assign weights to edges.

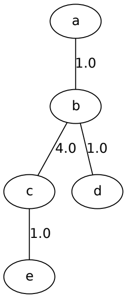
Example

String:	*abcde*
Character N-grams (Rank 3):	*abc, bcd, cde*
Edges (Window Size 1):	*abc-bcd, bcd-cde*
Weights (Occurrences):	*abc-bcd (1.0) , bcd-cde (1.0)*

N-gram Graph – Value Similarity

Example





Result: $\frac{\frac{1.0}{1.0}}{4} + \frac{\frac{4.0}{8.0}}{4} = \frac{1}{4} + \frac{1}{8} = 0.375$

Outline

Overview: Merged Model Graphs

- Merge models into one, common representation
- ...representative of the set of models
- ...easily comparable to peer summaries.

Representing Sets of Graphs

Merge operator: $G_1 \cup G_2$

- $E = E_1 \cup E_2$, E_i edgeset of G_i
- $w(e) = \frac{w_1(e)+w_2(e)}{2}$, if $e \in E_1 \cap E_2$
- else $w(e) = w_1(e)$, if $e \in E_1$, $e = w_2(e)$, if $e \in E_2$

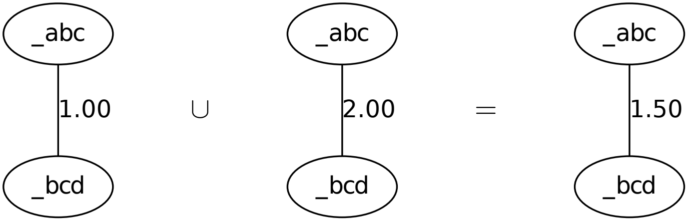
Non-commutative for > 2 graphs

Update operator: $U(G_1, G_2, l)$

- $E = E_1 \cup E_2$
- $w(e) = w_1(e) + l \times (w_2(e) - w_1(e))$

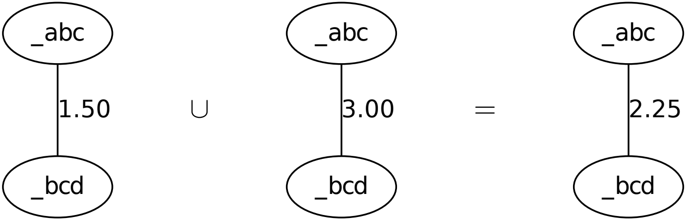
Parametric (learning factor); can be made commutative

Merge vs. Update (1)



$$\frac{1+2}{2} = 1.5$$

Merge vs. Update (2)



$$\frac{1.5+3}{2} = 2.25$$

But what if we want $\frac{1+2+3}{3} = 2$?

Merge vs. Update (3)

$$\text{updatedValue} = \text{oldValue} + l \times (\text{newValue} - \text{oldValue}) \qquad (1)$$

where $0 \leq l \leq 1$ is the learning factor

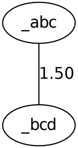
Representative (or *Centroid*) Graph

Use update operator with learning factor: $\frac{1}{\text{instanceCount}}$, where *instanceCount* is the number of instances that will be described by the graph *after* the update.

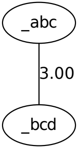
Merge vs. Update (4)

$$l = \frac{1}{3}$$

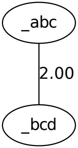
update(



,



$, l) =$



$$1.5 + \frac{1}{3} \times (3 - 1.5) = \frac{3}{2} + \frac{1}{3} \times \frac{3}{2} = \frac{4}{2} = 2$$

AutoSummENG vs MeMoG

AutoSummENG

ModelGraph1	PeerGraph	VS1
ModelGraph2	PeerGraph	VS2
ModelGraph3	PeerGraph	VS3
AutoSummENG Grade:		Avg(VSi)

MeMoG

U(ModelGraph1, ModelGraph2, ModelGraph3)	PeerGraph	VS
	MeMoG Grade:	VS

Outline

Overview: Hierarchical Proximity Graph

- Use multiple granularity levels in one representation
- Keep statistical, language-independent analysis
- Minimize parameters

Proximity Graph

- Generalization of n-gram graphs
- Represents neighborhood in any applicable domain
- Allows representation of neighborhood between graphs

From text to hierarchy — (1)

The original text

abcdabde

The n-grams

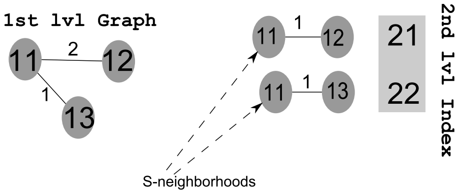
ab cd ab de

The 1st lvl text

11 12 11 13

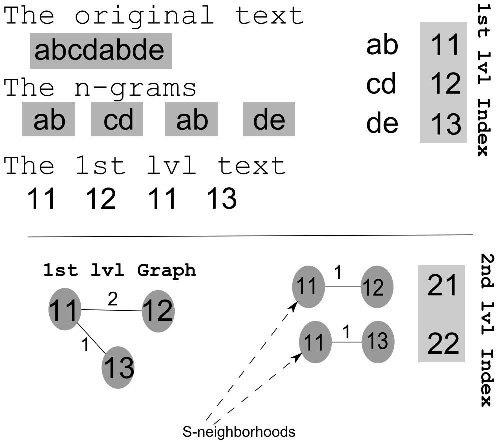
1st lvl Index

ab	11
cd	12
de	13



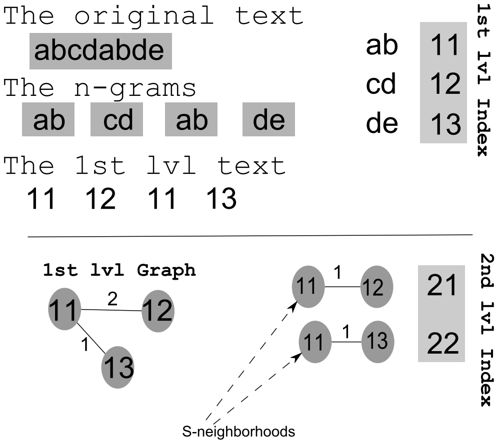
- Text into distinct n-grams

From text to hierarchy — (1)



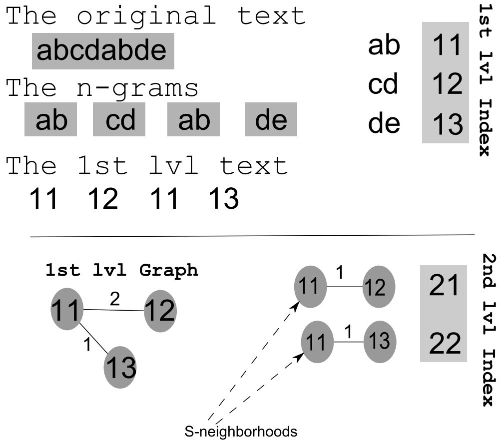
- Text into distinct n-grams
- Index I_1 of symbols to n-grams

From text to hierarchy — (1)



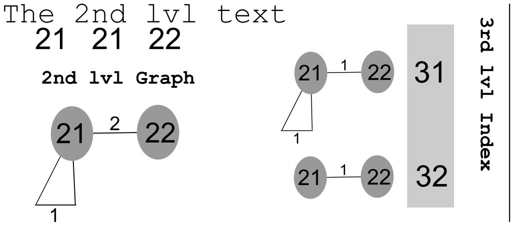
- Text into distinct n-grams
- Index I_1 of symbols to n-grams
- Recreate text as symbol sequence

From text to hierarchy — (1)



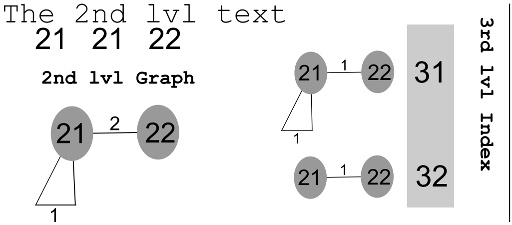
- Text into distinct n-grams
- Index I_1 of symbols to n-grams
- Recreate text as symbol sequence
- Extract *s-neighborhoods* and level graph

From text to hierarchy — (2)



- Neighborhood size increases over levels

From text to hierarchy — (3)



- Neighborhood size increases over levels
- Similarity: weighted average over all levels

Outline

1. Introduction
   - Summary Evaluation
   - The story so far...
2. Innovation in TAC2010
   - Merged Model Graphs: MeMoG
   - Hierarchical Proximity Graph: HPG
3. Experiments
   - Ranking Correlation
   - Discriminative Power
4. Closing
   - Summary and Future Work

Correlation to Pyramid — All Peers

Variant (ID)	P (Rank)	S (Rank)	K (Rank)
Group A - All Peers			
AutoSummENG (25)	0.897 (10)	0.945 (8)	0.824 (7)
MeMoG (16)	**0.956 (3)**	**0.956 (4)**	**0.834 (4)**
HPG (12)	0.880 (15)	0.941 (10)	0.818 (9)
Group B - All Peers			
AutoSummENG (25)	0.807 (12)	0.920 (9)	0.779 (8)
MeMoG (16)	**0.968 (1)**	**0.935 (4)**	**0.799 (4)**
HPG (12)	0.726 (16)	0.931 (6)	0.788 (6)

Legend P: Pearson, S: Spearman, K: Kendall

Correlation to Pyramid — No Models

Variant (ID)	P (Rank)	S (Rank)	K (Rank)
Group A - No Models			
AutoSummENG (25)	0.950 (11)	0.913 (9)	0.778 (10)
MeMoG (16)	**0.970 (4)**	**0.934 (5)**	**0.798 (6)**
HPG (12)	0.951 (10)	0.908 (10)	0.774 (11)
Group B - No Models			
AutoSummENG (25)	0.899 (13)	0.870 (11)	0.711 (11)
MeMoG (16)	**0.956 (5)**	**0.901 (7)**	**0.758 (6)**
HPG (12)	0.898 (14)	0.892 (9)	0.729 (9)

Correlation to Responsiveness — All Peers

Variant (ID)	P (Rank)	S (Rank)	K (Rank)
Group A - All Peers			
AutoSummENG (25)	0.909 (11)	0.942 (9)	**0.836 (4)**
MeMoG (16)	**0.944 (3)**	**0.949 (6)**	0.818 (10)
HPG (12)	0.892 (14)	0.945 (8)	0.820 (8)
Group B - All Peers			
AutoSummENG (25)	0.797 (11)	0.892 (9)	0.747 (9)
MeMoG (16)	**0.977 (1)**	0.915 (7)	0.772 (5)
HPG (12)	0.724 (15)	**0.925 (2)**	**0.777 (4)**

Correlation to Responsiveness — No Models

Variant (ID)	P (Rank)	S (Rank)	K (Rank)
Group A - No Models			
AutoSummENG (25)	0.923 (11)	0.911 (8)	**0.799 (5)**
MeMoG (16)	**0.949 (7)**	**0.920 (6)**	0.772 (8)
HPG (12)	0.932 (9)	0.916 (7)	0.779 (7)
Group B - No Models			
AutoSummENG (25)	0.876 (13)	0.829 (11)	0.686 (10)
MeMoG (16)	**0.933 (5)**	0.866 (9)	0.709 (8)
HPG (12)	0.888 (10)	**0.885 (5)**	**0.733 (6)**

Outline

1. Introduction
   - Summary Evaluation
   - The story so far...
2. Innovation in TAC2010
   - Merged Model Graphs: MeMoG
   - Hierarchical Proximity Graph: HPG
3. Experiments
   - Ranking Correlation
   - Discriminative Power
4. Closing
   - Summary and Future Work

Discriminative Power — All Peers

Group A - All Peers

	Pyramid		Responsiveness	
	Agree	Disagree	Agree	Disagree
AutoSummENG (25)	280	64	280	64
MeMoG (16)	**344**	**0**	**344**	**0**
HPG (12)	200	144	200	144

Group B - All Peers

	Pyramid		Responsiveness	
	Agree	Disagree	Agree	Disagree
AutoSummENG (25)	284	60	284	60
MeMoG (16)	**344**	**0**	**344**	**0**
HPG (12)	157	187	157	187

Discriminative Power — No Models

Group A - No Models	Pyramid		Responsiveness	
	Agree	Disagree	Agree	Disagree
AutoSummENG (25)	**813**	**90**	**793**	**110**
MeMoG (16)	798	105	788	115
HPG (12)	798	105	788	115
Group B - No Models	**Pyramid**		**Responsiveness**	
	Agree	Disagree	Agree	Disagree
AutoSummENG (25)	742	161	723	180
MeMoG (16)	764	139	753	150
HPG (12)	**792**	**111**	**787**	**114**

Outline

Summary

- Merged Model Graphs: Non-linear merging of models
- Hierarchical Proximity Graphs: Multiple granularity unification
- Both MeMoG and HPG methods offer improvement over AutoSummENG
- HPG is still in an early stage
- N-gram graph methods provide a complete evaluation toolbox
- AutoSummENG still highly competitive (Responsiveness)

Future Work

- HPGs: Study HPGs parameters and effect
- N-gram Graphs as a Constraint Satisfaction Problem
- Combine evaluators into single method, based on non-linear optimization
- Get to the summary level

Thank you

Do not forget: it is open source! **JInsect@Sourceforge**
Question answering time...

AutoSummENG – Evaluation Over DUC & TAC

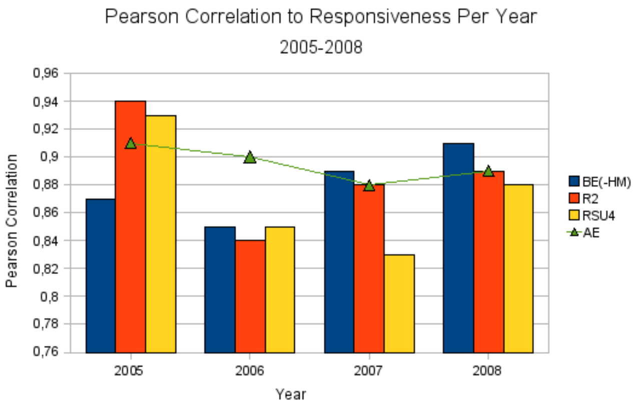


Figure: Pearson Correlation: Measures to (Content) Responsiveness for peers only

Giannakopoulos, G., Karkaletsis, V., Vouros, G., and Stamatopoulos, P. (2008).
Summarization system evaluation revisited: N-gram graphs.
ACM Trans. Speech Lang. Process., 5(3):1–39.