Entity-Grid Model

Youngsam Kim
Cog-Sci program

Coherence

Coherent text

George Bush was looking for a cup of water to drink.
His bodyguard saw what he was doing.
The bodyguard went to the president with a bottle of water.

Incoherent text

The bodyguard went to the president with a bottle of water.
His bodyguard saw what he was doing.
George Bush was looking for a cup of water to drink.

- Local coherence is necessary for global coherence.
- The distribution of entities in locally coherent texts shows certain regularities.

Entity-based approaches to local coherence

- Discourse coherence is achieved in view of the way discourse entities are introduced and discussed.
- The salience status of an entity is often reflected in its grammatical function and the linguistic form of its subsequent mentions.
- Salient entities are more likely to appear in prominent syntactic positions, and to be introduced in a main clause.
- Entity-based theories capture coherence by characterizing the distribution of entities across discourse utterances, distinguishing between salient entities and the rest.

Entity-grid representation

TEXT	Syntactic role
George Bush was looking for a cup of water to drink.	Subject
His bodyguard saw what **he** was doing.	Object
The bodyguard went to **the president** with a bottle of water.	Object
Johnson stood by the agent to give some help.	None

Grid Representation

	S 1	S 2	S 3	S 4
Bush	Subj	Obj	Obj	–
Bodyguard	–	Subj	Subj	X
Water	Obj	–	X	–
Johnson	–	–	–	Subj

Entity grids as feature vectors

- Grids of coherent texts are likely to have dense occurrence.
- Also, the dense sequences are more often subjects or objects.
- Local entity transition captures the patterns of entity distributions.
- A local entity transition is a sequence $\{s, o, x, -\}^n$ that represents entity occurrences and their syntactic roles in n adjacent sentences.
- Each transitions will have a certain probability in a given grid.
- Each text can thus be viewed as a distribution defined over transition types.

Using feature vector notation

- Each grid rendering j of a document d_i corresponds to a feature vector,

$$\Phi(x_{ij}) = (p_1(x_{ij}), p_2(x_{ij}), \ldots, p_m(x_{ij}))$$

A simple example

	S 1	S 2	S 3	S 4
Bush	Subj	Obj	Obj	–
Bodyguard	–	Subj	Subj	X
Water	Obj	–	X	–
Johnson	–	–	–	Subj

What is the probability of the transition
(Subject to Object) in grid x_{ij}?

P_t(SO) = 1/12 = 0.08

- We can go further and get a feature space with transitions of length two as illustrated below:

Table 3
Example of a feature-vector document representation using all transitions of length two given syntactic categories S, O, X, and –.

	S S	S O	S X	S –	O S	O O	O X	O –	X S	X O	X X	X –	– S	– O	– X	– –
d_1	.01	.01	0	.08	.01	0	0	.09	0	0	0	.03	.05	.07	.03	.59
d_2	.02	.01	.01	.02	0	.07	0	.02	.14	.14	.06	.04	.03	.07	0.1	.36
d_3	.02	0	0	.03	.09	0	.09	.06	0	0	0	.05	.03	.07	.17	.39

- Each row indicates a document containing all transitions each.
- This feature vector is useful for machine learning algorithms.

Three considerations for modeling

- The linguistic importance of a parameter.
- The accuracy of its automatic computation.
- The size of the resulting feature space.

Three factors in linguistic dimensions

- Entity extraction
  - Employ an automatic coreference resolution (Ng & Cardie, 2002)
  - It is trained on the MUC data, yielding state-of-the-art performance.
- Grammatical function
  - Discriminate subjects, objects, other relations, and absent (-)
  - Used Collins parser (1997)
- Salience
  - Identify salient entities based on their frequency
  - Salience classes are set to be binary.

Exp 1: Sentence ordering

- In sentence ordering, a document is viewed as a bag of sentences.
- The algorithm's task is to try find the ordering which maximizes coherence according to some criterion.
- In this task, random permutations of a test documents are generated and measured how often a permutation is ranked higher than the original document.
- A non-deficient model should prefer the original text more frequently than its permutations.

Exp 1: Modeling

- Training set consists of ordered pairs of alternative renderings (x_{ij}, x_{ik}) of the same document d_i, where x_{ij} exhibits a higher degree of coherence than x_{ik}.
- The goal of the training procedure is to find a parameter vector **w** that yields a 'ranking score' function which minimizes the number of violations of pairwise rankings.

$$\forall (x_{ij}, x_{ik}) \in r^* : \mathbf{w} \cdot \Phi(x_{ij}) > \mathbf{w} \cdot \Phi(x_{ik})$$

where $\Phi(x_{ij})$ and $\Phi(x_{ik})$ are a mapping onto features correspond to the entity transition probabilities.

- Finding w is solved by Support Vector Machine (Joachims, 2002).

Exp1: Method

- Data
  - News articles and accident reports
  - Used 100 source items with up to 20 randomly generated permutations for training
  - A similar method was used for the test data.
- Features and parameter settings
  - Focused on three sources: syntax, coreference resolution, and salience
- Comparison with state-of-the-art methods
  - Latent Semantic analysis model was used for comparative result. (Foltz et al, 1998)
  - HMM based content model (Barzilay & Lee, 2004)

Exp 1: Results

- Impact of linguistic dimension

Model	Earthquakes	Accidents
Coreference+Syntax+Salience+	**87.2**	**90.4**
Coreference+Syntax+Salience−	88.3	90.1
Coreference+Syntax−Salience+	86.6	88.4**
Coreference−Syntax+Salience+	83.0**	89.9
Coreference+Syntax−Salience−	86.1	89.2
Coreference−Syntax+Salience−	82.3**	88.6*
Coreference−Syntax−Salience+	83.0**	86.5**
Coreference−Syntax−Salience−	81.4**	86.0**
HMM-based Content Models	88.0	75.8**
Latent Semantic Analysis	81.0**	87.3**

Exp 1: Results

- Training requirements

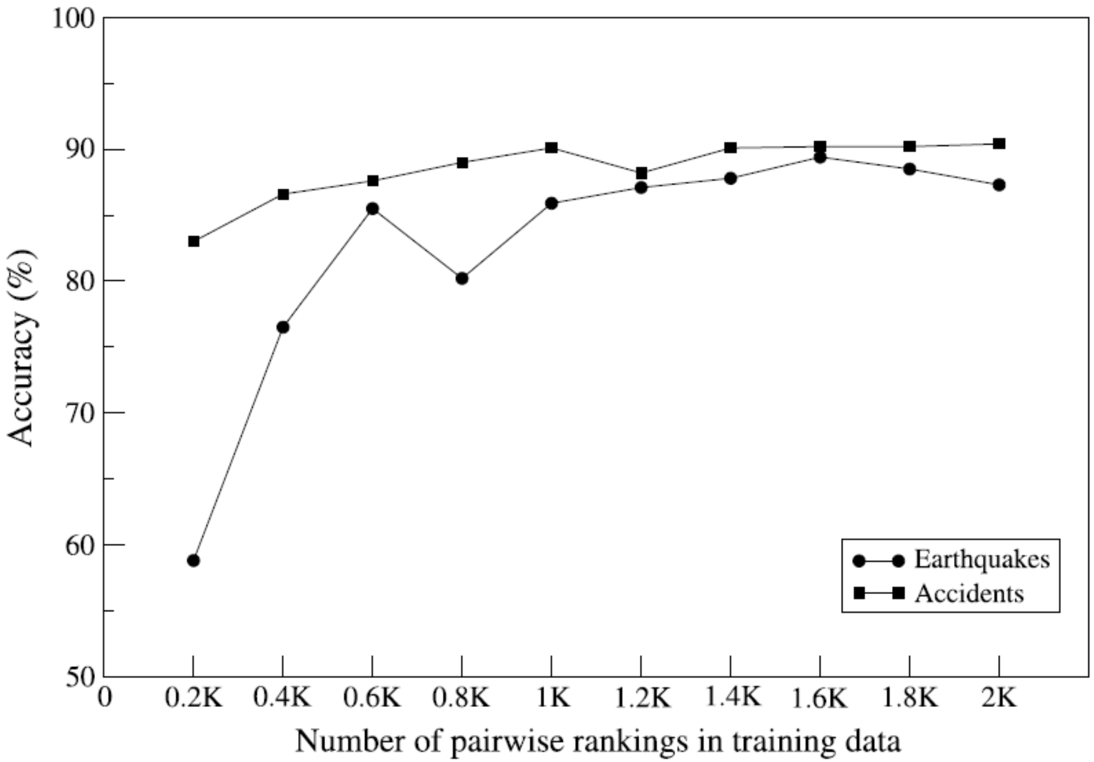


Figure 1
Learning curves for the entity-based model **Coreference**+**Syntax**+**Salience**+ on the Earthquakes and Accidents corpora.

Exp 1: Results

- In/out-domain performance

Table 7
Accuracy of entity-based model (Coreference+Syntax+Salience+) and HHM-based content model on out-of-domain texts. Diacritics ** ($p < .01$) and * ($p < .05$) indicate whether performances on in-domain and out-of-domain data are significantly different using a Fisher Sign Test.

Coreference+Syntax+Salience		
Train \ Test	Earthquakes	Accidents
Earthquakes	**87.3**	67.0**
Accidents	69.7**	**90.4**
EarthAccid	86.7	88.5*

HMM-Based Content Models		
Train \ Test	Earthquakes	Accidents
Earthquakes	**88.0**	31.7**
Accidents	60.3**	**75.8**

Exp 1: Conclusions

- The full model clearly outperforms the LSA and HMM-based models.
- Learning curve shows that 1,000 pairs would be good enough for training.
- The result indicates that coreference resolution can be helpful in the domain which uses pronominalizing.
- The result of domain difference seems to show that accuracy is not only due to coherence properties, but also reflects stylistic and genre-specific discourse properties.

Exp 2: Summary coherence rating

- Summary coherence rating can be a ranking learning task.
- The training data includes pairs of summaries (x_{ij}, x_{ik}) of the same documents di where x_{ij} is more coherent than x_{ik}.
- An optimal learner should return a ranking r* that orders the summaries to their coherence.

Exp 2: Method

- Data
  - Evaluations are based on materials from the DUC, produced by human writers and automatic summarization systems.
  - Obtained judgments for automatically generated summaries from human subjects.
  - Randomly selected 16 input document clusters and five systems produced summaries for these sets.
  - Human subjects also provided reference summaries.
  - 177 volunteers rated the documents on 1~7 point scale.
  - Consequently, 144 summaries for training and 80 pairwise rankings for the test set were obtained.

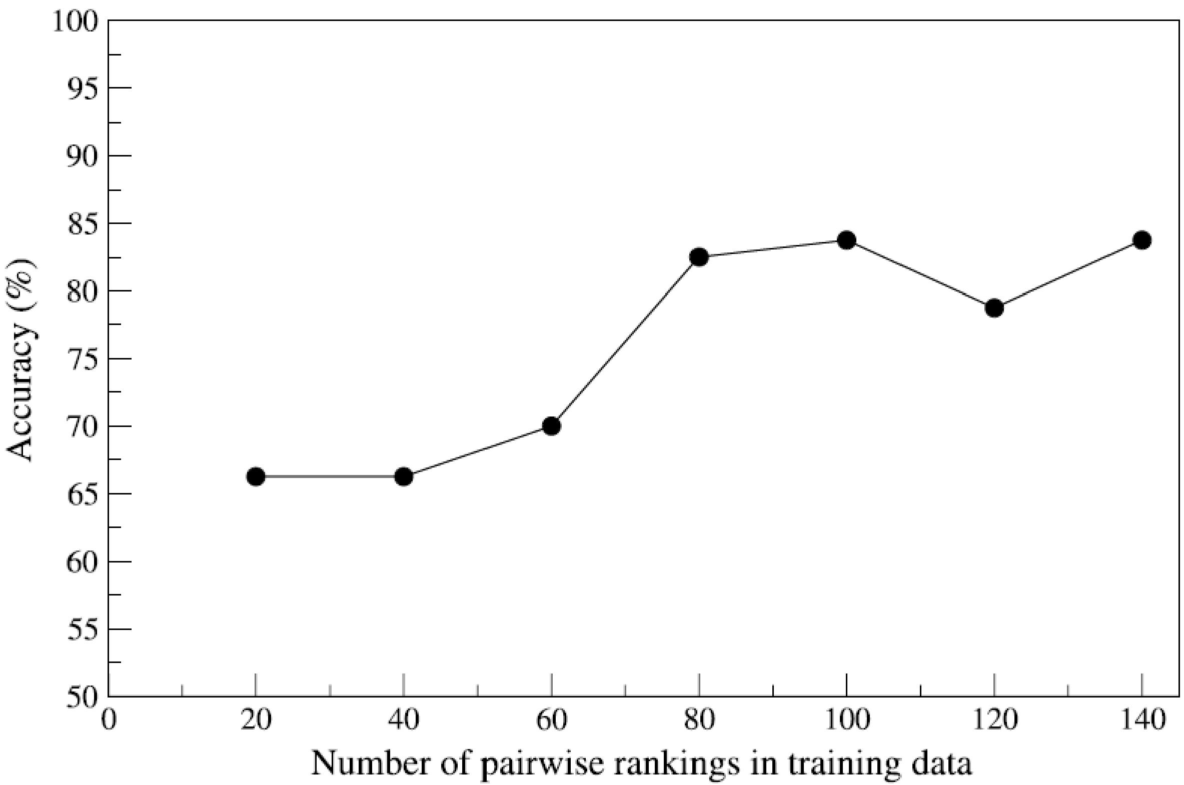
Exp 2: Results

- Impact of linguistic dimension

Model	Accuracy
Coreference+Syntax+Salience+	80.0
Coreference+Syntax+Salience−	75.0
Coreference+Syntax−Salience+	78.8
Coreference−Syntax+Salience+	**83.8**
Coreference+Syntax−Salience−	71.3*
Coreference−Syntax+Salience−	78.8
Coreference−Syntax−Salience+	77.5
Coreference−Syntax−Salience−	73.8*
Latent Semantic Analysis	52.5**

Exp 2: Results

- Training requirement



Exp 3: Readability assessment

- Put the entity-grid representation into a system that assesses document readability.
- Schwarm and Ostendorf(2005) developed a method for assessing readability.
- Following the way of Schwarm & Ostendorf, readability assessment is viewed as a classification task here.
- The unit of classification is a single article and the task is to predict whether it is easy or difficult to read.

Exp 3: Method

- Data
  - Difficult-to-read corpus from Encyclopedia Britannica, easy-to-read corpus from Britannica Elementary.
  - The corpus contains 107 articles from each Britannica versions.

Exp 3: Method

- Features
  - Schwarm & Ostendorf(2005) use three broad classes of features: syntactic, semantic, and their combination.
  - Syntactic features: sentence length, parsing info (mean number of NPs, VPs, etc)
  - Semantic features: number of syllables per word, language model perplexity scores
  - Flesch-Kincaid Grade Level score:

$$0.39\left(\frac{total\ words}{total\ sentences}\right)+11.8\left(\frac{total\ syllables}{total\ words}\right)-15.59$$

  - Feature space was extended with entity-based features.

Exp 3: Results

- Classification accuracy was measured by SVM over the size of the test data.

Model	Accuracy
Schwarm & Ostendorf	78.56
Schwarm & Ostendorf, Coreference+Syntax+Salience+	**88.79***
Schwarm & Ostendorf, Coreference−Syntax+Salience+	79.49
Schwarm & Ostendorf, Latent Semantic Analysis	78.56
Coreference+Syntax+Salience+	50.90**
Coreference−Syntax+Salience+	49.55**
Latent Semantic Analysis	48.58**

Exp 3: Implications of the results

- Easy texts tend to employ less coreference and the use of personal pronouns is relatively small.
- Thus, coreference information is a good indicator of the reading difficulty.
- The entity-based model supplements their feature space with information spanning two or more in discourse.
- Difficult texts tend to have twice as many entities as easy ones.
- LSA scores capture similar aspects of the Flesch Kincaid estimates, therefore it does not contribute to its performance.

Extending entity-grid model

- Elsner and Charniak (2011) added discourse prominence, named entity type and coreference features to distinguish between important and unimportant entries.
- Distinguishing important from unimportant entity types is important in coreference and summarization.
- The standard grid model does not distinguish between different types of entity.
- Thus, the same number of occurrences of items means that they are assigned the same probability, contrary to our intuition.

Entity-specific features

- Features for importance
  - **Proper** Does the entity have a proper mention?
  - **Named entity** The majority OPENNL Morton et al. (2005) named entity label for the coreferential chain.
  - **Modifiers** The total number of modifiers in all mentions in the chain, bucketed by 5s.
  - **Singular** Does the entity have a singular mention?

- Features for unimportance
  - If there are pronouns coreferent with entities found in NANC corpus.
  - Was the head word of the entity marked as coreferring in MUC6 corpus?

- Elsner and Charniak (2011) shows employing extended features about entity-types increased the performance of the model.
- They tested the model on the WSJ news articles using conditional probability model($P(x_{ij}|extended\ features)$).
- For instance,

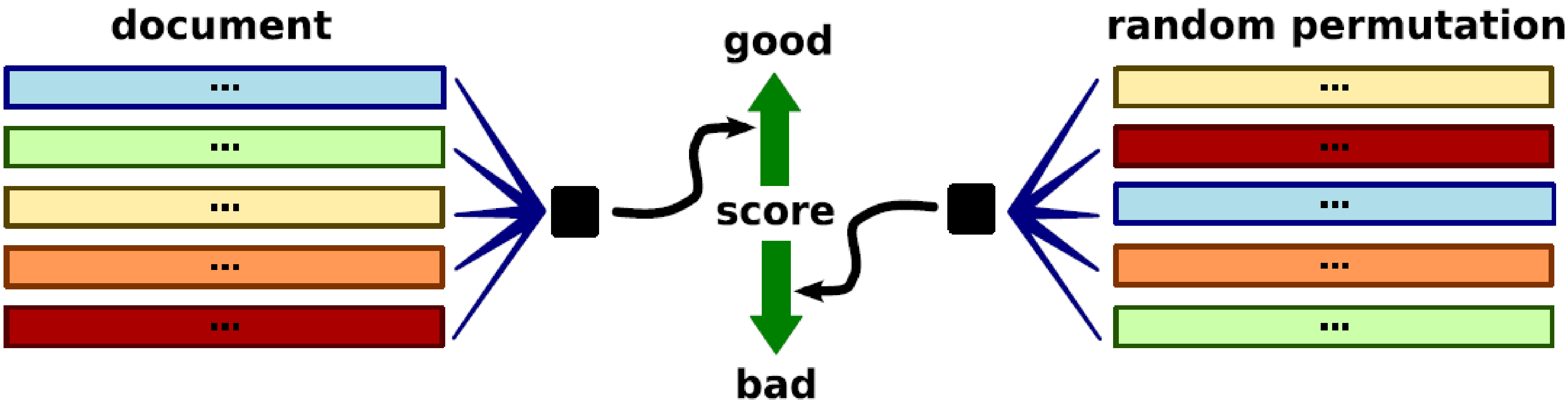
in previous model:

P(6th | -, x, 3s) = 0.34

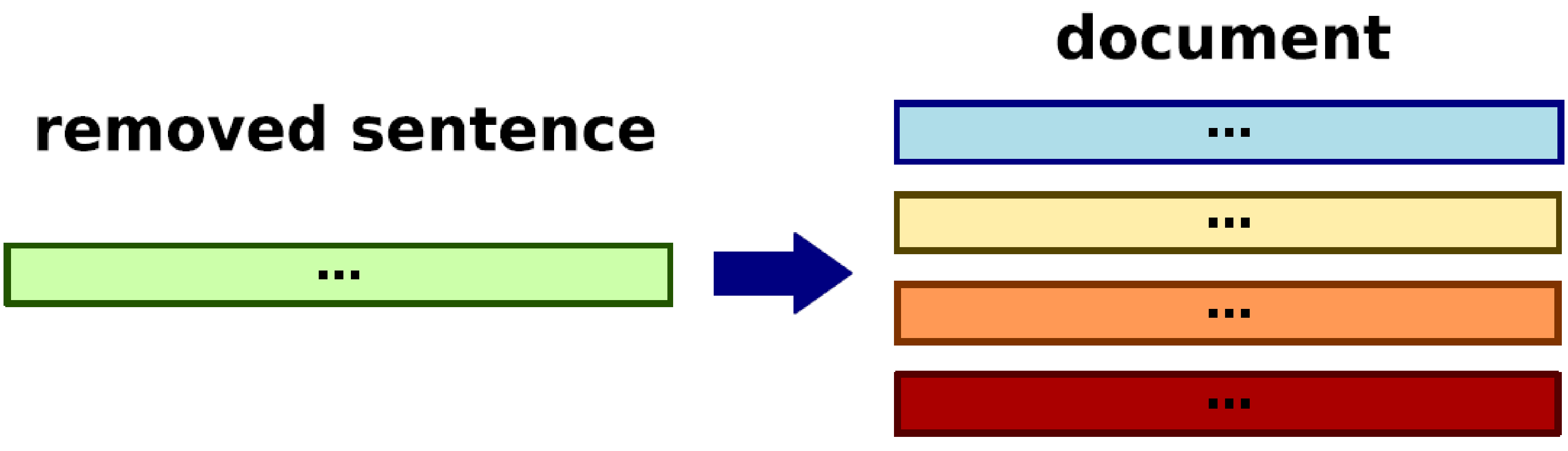
in extended model:

P(Obama | -, x, 3s,
| NE type PERSON
| proper noun
| corefer in MUC6) = 0.12

Discrimination



Insertion



	Accuracy	Insertion
Random	50.0	12.6
Elsner&Charniak, 2008	79.6	23.0
Entity grid model	79.5	21.4
Extended grid model	**84.0**	**24.2**
Grid+combo	82.6	24.3
ExtEGrid+combo	**86.0**	**26.7**

- Combination models incorporate pronoun coreference, discourse-new NP detection, and IBM translation model.
- Although the improvement in ExtGrid+combo model is not perfectly additive, about 3% increase retains.
- These results are the best reported for sentence ordering of English news articles.