Automatic Extraction of Briefing Templates

Dipanjan Das, Mohit Kumar and Alexander I. Rudnicky

Language Technologies Institute
Carnegie Mellon University

{dipanjan, mohitkum, air}@cs.cmu.edu
Outline

• Introduction
• Related Work
• Data
• Approaches
• Results
• Conclusion
Outline

• Introduction
• Related Work
• Data
• Approaches
• Results
• Conclusion
Introduction

Weather reports: examples

A warm front from Iceland to northern Scotland will move SE across the northern North Sea today and tomorrow.

Reiter et al., 2005
Introduction

Weather reports: examples

A warm front from Iceland to northern Scotland will move SE across the northern North Sea today and tomorrow.

A ridge from the British Isles to Iceland will move NE across the North Sea today.

Reiter et al., 2005
Introduction

Weather reports: examples

A warm front from Iceland to northern Scotland will move SE across the northern North Sea today and tomorrow.

A ridge from the British Isles to Iceland will move NE across the North Sea today.

Domain entities labeled

Reiter et al., 2005
Introduction

Weather reports: examples

A warm front from Iceland to northern Scotland will move SE across the northern North Sea today and tomorrow.

A ridge from the British Isles to Iceland will move NE across the North Sea today.

Reiter et al., 2005
Introduction

Weather reports: examples

A warm front from Iceland to northern Scotland will move SE across the northern North Sea today and tomorrow.

A ridge from the British Isles to Iceland will move NE across the North Sea today.

Reiter et al., 2005
Introduction

Weather reports: examples

A warm front from Iceland to northern Scotland will move SE across the northern North Sea today and tomorrow.

A ridge from the British Isles to Iceland will move NE across the North Sea today.

Reiter et al., 2005
Introduction

Weather reports: examples

A warm front from Iceland to northern Scotland will move SE across the northern North Sea today and tomorrow.

A ridge from the British Isles to Iceland will move NE across the North Sea today.

Reiter et al., 2005
Introduction

Weather reports: examples

A warm front from Iceland to northern Scotland will move SE across the northern North Sea today and tomorrow.

A ridge from the British Isles to Iceland will move NE across the North Sea today.

Equivalent sentences have very similar surface forms!

Reiter et al., 2005
Introduction

• Briefing generation from non-textual events
  • medical reports
  • weekly class project reports
  • weather reports
  • traffic reports
  • ....
Introduction

Weather reports: examples

A warm front from Iceland to northern Scotland will move SE across the northern North Sea today and tomorrow.

A ridge from the British Isles to Iceland will move NE across the North Sea today.

[PRESSURE_ENTITY] from [LOCATION] to [LOCATION] will move [DIRECTION] across [LOCATION] [TIME].
Introduction

• The task of closed domain briefing generation
• Aggregation of information from data
• Filling up templates
Introduction

• The task of closed domain briefing generation
• Aggregation of information from data
• Filling up templates

Often manually designed
Introduction

• The task of closed domain briefing generation

• Aggregation of information from data

• Filling up templates

Our focus: automatic creation from human briefings
Outline

• Introduction

• Related Work

• Data

• Approaches

• Results

• Conclusion
Related Work

- In evaluations like MUC and ACE
  - template based summarization systems
  - hand engineered templates
  - focus on slot filling, rather than template creation
Related Work

• Collier (1998) proposed MUC type template extraction
  • relied on identification of statistically significant words
  • discovery of common subject-verb-object patterns
Related Work

• Filatova et al. (2006) improve the paradigm
  • identification of most common verbs in the corpus
  • alignment of similar subtrees for each verb
  • however, long distance dependencies of verbs not examined
Outline

• Introduction
• Related Work
• Data
• Approaches
• Results
• Conclusion
Data

- Weather Forecasts
- Corpus containing human written reports
  - natural language summaries of computer data
- Reiter et al. (2005)
- Regularity in sentence structure
  - generalizable to traffic reports, medical reports, student class reports
Data

- 3262 sentences
  - 3000 for template extraction
  - 262 for testing coverage of templates
Preprocessing

• ASSERT toolkit (Pradhan et al., 2004)

  • tags each sentence with semantic roles for each verb
  
  • produces a phrase structure tree as a byproduct (Charniak, 2001)
Preprocessing

- Semantic role tagging
- BaseNP recognition
- Domain Entity tagging
Preprocessing

A low over the Norwegian Sea will move North and weaken

[ARG0 A low over the Norwegian Sea] [ARGM-MOD will] [TARGET move] [ARGM-DIR North] and weaken

[ARG0 A low over the Norwegian Sea] [ARGM-MOD will] move North and [TARGET weaken]
Preprocessing

A low over the Norwegian Sea will move North and weaken
Preprocessing

ARGO

NP

ARGM-MOD

NP

VP

ARGM-DIR

A low over the Norwegian Sea will move North and weaken

TARGET

Das, Kumar and Rudnicky
A low over the Norwegian Sea will move North and weaken
A low over the Norwegian Sea will move North and weaken.
A low over the Norwegian Sea will move North and weaken
Preprocessing

- Domain entity tagging
  - Manually engineered module
  - < 1000 vocabulary for content words
  - Often, entities form a closed list
    - e.g. DIRECTION – North, South...
  - 13 such entity types
Outline

• Introduction
• Related Work
• Data
• Approaches
• Results
• Conclusion
Overall Approach

• Clustering of sentences in corpus
  • sentences conveying similar information
  • instantiation of the same template

• Extraction of templates from each cluster
Approaches

• Two types of clustering approaches
  • Semantic role labeling (SRL) approach
  • ROUGE based approach
SRL Approach

• Advantages:
  • complicated sentences broken down into set of propositions
  • propositions better generalizable units across a corpus
  • long distance dependencies of constituents with verbs are modeled
  • for each verb, sentences with the same semantic roles convey similar meaning
SRL Approach

- Foreword
  - Each sentence is analyzed by ASSERT to produce a set of propositions
  - Clustering proceeds on the set of all propositions
SRL Approach

• Three stages of clustering
• Verb based clustering
• Semantic role based clustering
• Clustering based on syntax/semantics of roles
SRL Approach

• Verb based clustering
  • First, verbs in the corpus are clustered together using Wordnet
  • Propositions with the same verb type are grouped in the same cluster
  • 82 such clusters produced for 6632 propositions
SRL Approach

• Semantic role based clustering
  • For each verb cluster
    • all propositions with the exact semantic role sequences are further grouped
SRL Approach

Example:

[ARG0 A low over the Norwegian Sea] [ARGM-MOD will] [TARGET move] [ARGM-DIR North] and weaken

[ARG0 A high pressure area] [ARGM-MOD will] [TARGET move] [ARGM-DIR southwestwards] and build on Sunday

• Clusters with only one proposition are eliminated
• 33 verb type clusters with several sub-clusters each are produced
SRL Approach

• Looking inside semantic roles

• For propositions having the same verb and same sequence of semantic roles

• individual roles are examined to find out matching structures
A low over the Norwegian Sea

A frontal trough across Scotland
SRL Approach
SRL Approach

Matching structure

A low over the Norwegian Sea

A frontal trough across Scotland
SRL Approach

• Thus, for each role
  • Reduction of parse structure to sequence of POS tags and domain entity tags
  • Clustering of roles having exact same sequence of tags
  • Concatenation of role sequence tags produce final templates
  • 209 templates produced
ROUGE based Approach

• ROUGE: de facto standard for evaluation of text summaries (Lin, 2004)
  • Comparison of machine summary with a set of reference summaries
  • Token co-occurrence statistics based measure
  • Weighted longest common subsequence (WLCS) a useful variation

• Advantages
  • Robust metric
  • Parsing independent
  • Generalizable across domains
Overview

- Preprocessing
- Abstracted Sentences
- Cluster
- Align
- Longest Common Subsequence based
- ROUGE based
- Templates
ROUGE based Approach

• Preprocessing
  • Each sentence in the corpus was converted to a base NP and domain entity tagged form

  A low over the Norwegian Sea will move North and weaken

• Clustering
  • Distance metric: ROUGE-WLCS score of the abstracted sentences
  • an indication of how related they are
ROUGE based Approach

• Unsupervised Clustering
  • Hierarchical clustering performed using the scores
  • However, resulting clusters were not coherent
  • Classical parameter estimation problem
ROUGE based Approach

- **Non-parametric Unsupervised Clustering**
  - Cross-association based approach (Chakrabarti et al., 2004)
  - Based on boolean similarity matrices
    - thresholding of ROUGE scores
  - poor performance; incoherent clusters
  - MDL principle did not work for this data
ROUGE based Approach

- Deterministic Clustering
  - Underlying intuition
    - Sentences $X_{1..n}$ that are similar to any other sentence $Y_i$ should be in the same cluster
    - $X_j$ and $X_k$ may not be similar to each other
    - Connected components in the similarity matrix are discovered to get individual clusters
ROUGE based Approach

- *Deterministic Clustering*
  - Original similarity metric retrieved incoherent clusters
  - To make the similarity metric more interpretable
  - ROUGE-WLCS was factored into precision and recall components
  - High precision match essential in the domain
  - Recall too becomes important to match similar length sentences
  - However, individual thresholding of the measures gave better clustering
ROUGE based Approach

• The grouping produced 149 clusters
• For each cluster, the longest common subsequence was computed to give the template
• 149 templates were produced
Templates

Examples:

PRESSURE_ENTITY expected over LOCATION by_0.5/on_0.5 DAY

PRESSURE_ENTITY to DIRECTION of LOCATION will drift slowly

WAVE will run_0.5/move_0.5 DIRECTION then DIRECTION

Associated PRESSURE_ENTITY will move DIRECTION across LOCATION TIME
Outline

• Introduction
• Related Work
• Data
• Approaches
• Results
• Conclusion
Results

• Evaluation Scheme
  • no standard evaluation metric
  • adoption of both subjective and automatic measures
Results

\[
\text{precision} = \frac{\text{number of domain relevant templates}}{\text{total number of extracted templates}}
\]

- subjective measure
- three human subjects were asked to mark each template as
  - domain relevant or not
  - grammatical or not
Results

• Recall

  • automatic metric
  
  • Preprocessed test set
    • domain entity and baseNP tagged
    
    • ROUGE-WLCS score between each preprocessed sentence and valid templates

  • At various precision thresholds, recall is measured
Results

• Results: SRL Approach
  • overall inter-rater agreement
    • $\kappa = 0.69$.
  • substantial agreement
• Overall precision: 84.21%
• 47.47% of irrelevant templates marked ungrammatical
• Results: SRL Approach
  • 30% near exact match at 0.9 ROUGE-WLCS score
  • For 0.6 precision, recall is encouraging 81%
Results

• ROUGE based approach

• At $\kappa = 0.79$, precision: 76.3%

• 96.7% of the irrelevant templates were marked ungrammatical

• simple errors in syntax leading to poor precision
Results

- ROUGE based approach
  - Error recovery module that strips leading and trailing prepositions
  - Boost of precision to 80.98%
  - $\kappa = 1$
Results

- **ROUGE** based approach
  - For high precision points, recall is low
    - however, the surface form extracted is richer than the SRL approach
    - exact match is infeasible
Outline

• Introduction
• Related Work
• Data
• Approaches
• Results
• Conclusion
Conclusion

• High precision values for both approaches

• SRL approach identifies more generalizable units containing one predicate

• ROUGE approach finds richer templates

• Future directions

  • Merger of the two methods

  • New domains
Questions?