Empirically Estimating Order Constraints for Content Planning in Generation

Pablo A. Duboue and Kathleen R. McKeown

Computer Science Department
Columbia University
in the city of New York

(ACL'01)

A Natural Language Generation Pipeline

1. Content Planning

   What to say and its ordering.

2. Sentence Planning

   Division into sentences.

3. Surface Realisation

   How to say it.

Content Planning

- Content Selection
  - Arguably the most critical part from the user's perspective
- Ordering
  - conciseness and coherentness goals.
  - Information in context.
  - Take into account communicative goals.
  - **Problem: given n items there are n! possible orderings**

Long-term Scenario

Input		Output
• Raw data • Target documents	$\Rightarrow$	• Content Planner

- Problems:
  - Lack of ontological information.
  - Matching documents to sections in the data.
  - Matching text clauses to particular input.

Current Scenario

Input		Output
• Semantic Input • Tagged transcripts	$\Rightarrow$	• Order Constraints

- Advantages:
  - Domain semantics.
  - Human annotated text.
  - Easier task, although important.

Our Task

- Applying Empirical Methods to Content Planning
  - Content Planning is deeply tied to semantics.

- Learning Backbone Ordering Constraints
  - Important in practice – reducing the search space.
  - Dependent only on the domain semantics.

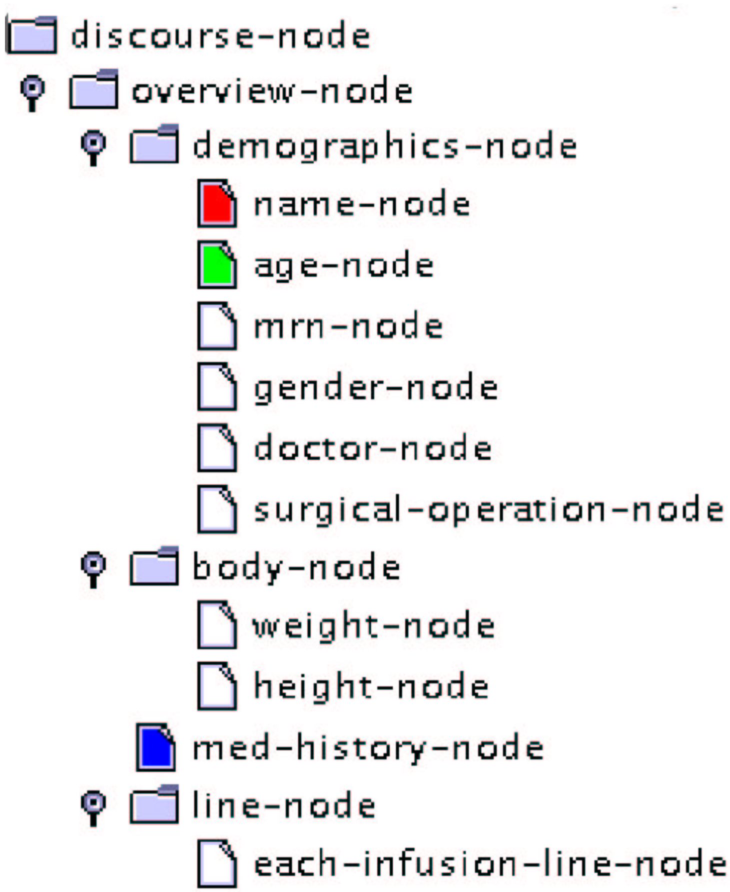
Task Specification

- Input
  - Set of semantically tagged texts.
- Output
  - Elements A, B, C
    - Sequence of semantic tags $A =$ab$?$d
  - Global ordering over elements $A \prec B$
- Methods
  - Apply computational biology over the sequences of tags

Our System: MAGIC

- MAGIC
  - Fully developed.
  - Intelligent multimedia presentation system.
  - Medical domain.
- Task
  - Reporting cardiac surgery patient status.
  - Time critical.

MAGIC: Example



"J. Doe is a seventy-eight year-old male patient of Doctor Smith undergoing aortic valve replacement. His medical history includes allergy to penicillin and congestive heart failure. He is sixty-six kilograms and one hundred sixty centimeters. . . . . . "

The Data

- From the Evaluation Described in McKeown et al., (2000)
  – Annotated transcriptions of physicians briefings.
- Semantic Annotation
  – Assisted by a domain expert.
  – Semantically tagged chunks (clausal level, non-overlapping).
  – Tag-set
    * Over 200 tags
    * 29 categories
- Expensive Task
  – Intensive Care Unit, a busy environment.
  – A total number of 24 transcripts.
  – Average length of around 33 tags.

The Data: Example

"He is 58-year-old (age) male (gender). History is significant for Hodgkin's disease (pmh), treated with . . . to his neck, back and chest. Hyperspadias (pmh), BPH (pmh), hiatal hernia (pmh) and proliferative lymph edema in his right arm (pmh). No IV's or blood pressure down in the left arm. Medications — Inderal (med-preop), Lopid (med-preop), Pepcid (med-preop), nitroglycerine (drip-preop) and heparin (med-preop). EKG has PAC's (ekg-preop). His Echo showed AI, MR of 47 cine amps with hypokinetic basal region (echo-preop). Hematocrit 1.2 (hct-preop), otherwise his labs are unremarkable. Went to OR for what was felt to be 2 vessel CABG off pump both mammaries (procedure) . . . . . ."

Our Algorithm

Sequences

$\Downarrow$

Motif (Pattern) Detection

$\Downarrow$

Patterns

A=ab$?$c

$\Downarrow$

Clustering

$\Downarrow$

Generalized patterns

$\tilde{A}$$\{ab?c,ad?c\}$

$\Downarrow$

Constraints Inference

$\Downarrow$

Order Constraints over Clusters

$$A \prec B$$

Analysis of the Problem

- Focus on the **Sequence** of Semantic Tags:

  age, gender, pmh, pmh, pmh, pmh, med-preop, med-preop, med-preop, drip-preop, med-preop, ekg-preop, echo-preop, hct-preop, procedure, . . .

- Find Regularities in Sequences
- Biological Sequence Analysis Techniques
  - Similar problems.
  - Scalability.

More Regularity: Motif Detection

- Motifs
  - A small subsequence, highly conserved through evolution.
  - A fixed-length pattern.
  - Example: (from `http://motif.stanford.edu/emotif/`)

$$h?[kr]?h\,[st]\,g\,[eq]\,[kqr]\,p\,[fy]?c$$

AEF1_DROME	NFCPKHFRQLSTLAN	**HVKIHTGEKPFEC**	VICKKQFRQSSTLNN	(258–270)
AZF1_YEAST	DYCGKRFTQGGNLRT	**HERLHTGEKPYSC**	DICDKKFSRKGNLAA	(639–651)
BCL6_HUMAN	EICGTRFRHLQTLKS	**HLRIHTGEKPYHC**	EKCNLHFRHKSQLRL	(648–660)
BCL6_MOUSE	EICGTRFRHLQTLKS	**HLRIHTGEKPYHC**	EKCNLHFRHKSQLRL	(649–661)
BTD_DROME	PGCERLYGKASHLKT	**HLRWHTGERPFLC**	LTCGKRFSRSDELQR	(353–365)
BTE1_HUMAN	SGCGKVYGKSSHLKA	**HYRVHTGERPFPC**	TWPDCLKKFSRSDEL	(163–175)

intraop-problems, **intraop-problems**, **?**, **drip**

- Motif Detection Algorithms
  - Different techniques: HMM, Alignment, Combinatorial
  - TEIRESIAS

TEIRESIAS

- Pattern Discovery Algorithm
- Algorithm Sketch
  - Identify basic patterns (“scanning”).
  - Grow patterns (“convolution”).
  - Find patterns with enough **support**.
- Benefits
  - Swapped elements:

    $\texttt{abc}A\texttt{de}B\texttt{fg}C\texttt{hij}$

    $\texttt{xyz}C\texttt{pq}B\texttt{rs}A\texttt{tvw}$
  - Hand-tunable parameters.

More Regularity: Clustering

- Capturing Further Regularities

  intraop-problems, intraop-problems, ?, drip

  intraop-problems, ?, drip, drip

- Solution: Clustering
  - Agglomerative clustering.
  - Approximate matching distance
    - Measures similarity related to the training-set.

A cluster

intraop-problems	**intraop-problems**	**operation** 11.11% **drip** 33.33% **intraop-problems** 33.33% **total-meds-anesthetics** 22.22%	**drip**	
intraop-problems	**operation** 14.29% **drip** 14.29% **intraop-problems** 42.86% **total-meds-anesthetics** 28.58%	**drip**	**drip**	
intraop-problems	**intraop-problems**	**operation** 20.00% **drip** 20.00% **intraop-problems** 20.00% **total-meds-anesthetics** 40.00%	**drip**	**drip**

How to Learn Order Constraints

- Measure the Frequency of Possible Orderings
  - Ordering of elements built over semantic tags.
- Reject Incorrect Orderings
- Build Table of Counts, Compute Probabilities
  - Similar to Shaw and Hatzivassiloglou (1999).
- Suitable Elements:
  - Increase regularity in the input.

Final Algorithm

Sequences

$\Downarrow$

Motif (Pattern) Detection

$\Downarrow$

Patterns

A=ab$?$c

$\Downarrow$

Clustering

$\Downarrow$

Generalized patterns

$\tilde{A}$$\{ab?c,ad?c\}$

$\Downarrow$

Constraints Inference

$\Downarrow$

Order Constraints over Clusters

$$A \prec B$$

Results

- Evaluation Settings:
  - Using the 24 transcripts
  - 3-fold cross validation
  - Hand-tuning of parameters
- **Constraint Accuracy:** 89.45%

Qualitative Evaluation

- Evaluation Setting
  - Using all available data (at one time).
  - Same parametric settings as quantitative evaluation.
  - 29 constraints, out of 23 clusters.
- Comparison to the Existing Content Planner
  - The existing planner was carefully crafted.
  - All the constraints found were validated.
  - Gained placement constraints for 2 pieces of new information.
  - Learned minor order variations in the placement of 2 rules.

Conclusion

- A Novel Empirical Method for Learning of Content Planning Elements
  - Relating the problem to biological sequence analysis.
- Successful Results
  - Feasibility of the task.
  - High precision and increased variability of the plan.
  - Easily extendable

    diabetic patients and past medical history

Further Work

- Integrate Results
  - Genetic search over the planners space (as on Mellish et al. (1998)).
  - Alignment scores as a measure of similarity.
- Automatic Tagging
- Explore Other Alternatives
  - Pattern Expressibility