Towards Strict Sentence Intersection: Decoding and Evaluation Strategies

Kapil Thadani and Kathleen McKeown

Columbia University
New York, NY
Text to text generation

Popular approaches: sentence compression, fusion

Concern:
1. Salience-based judgments
2. Semantic variation acceptable in output
Examples

Fusion

- After years of pursuing separate and conflicting paths, AT&T and Digital Equipment Corp. agreed in June to settle their computer-to-PBX differences.

- The two will jointly develop an applications interface that can be shared by computers and PBXs of any stripe.

Daumé III and Marcu (2004)
Examples

Fusion

- After years of pursuing separate and conflicting paths, AT&T and Digital Equipment Corp. agreed in June to settle their computer-to-PBX differences.

- The two will jointly develop an applications interface that can be shared by computers and PBXs of any stripe.

AT&T and Digital Equipment Corp. agreed in June to settle their computer-to-PBX differences and develop an applications interface that can be shared by any computer or PBX.

Daumé III and Marcu (2004)
Examples
Fusion

- After years of pursuing separate and conflicting paths, AT&T and Digital Equipment Corp. agreed in June to settle their computer-to-PBX differences.

- The two will jointly develop an applications interface that can be shared by computers and PBXs of any stripe.

AT&T and Digital Equipment Corp. agreed in June to settle their computer-to-PBX differences and develop an applications interface that can be shared by any computer or PBX.

After years of pursuing different paths, AT&T and Digital agreed to jointly develop an applications interface that can be shared by computers and PBXs of any stripe.

Daumé III and Marcu (2004)
After years of pursuing separate and conflicting paths, AT&T and Digital Equipment Corp. agreed in June to settle their computer-to-PBX differences.

The two will jointly develop an applications interface that can be shared by computers and PBXs of any stripe.

AT&T and Digital Equipment Corp. agreed in June to settle their computer-to-PBX differences and develop an applications interface that can be shared by any computer or PBX.

After years of pursuing different paths, AT&T and Digital agreed to jointly develop an applications interface that can be shared by computers and PBXs of any stripe.
Examples

Compression

- TapeWare, which supports DOS and NetWare 286, is a value-added process that lets you directly connect the QA150-EXAT to a file server and issue a command from any workstation to back up the server.

McDonald (2006)
Examples
Compression

- **TapeWare**, which supports DOS and NetWare 286, is a value-added process that lets you directly connect the QA150-EXAT to a file server and issue a command from any workstation to back up the server.

**TapeWare supports DOS and NetWare 286.**

McDonald (2006)
TapeWare, which supports DOS and NetWare 286, is a value-added process that lets you directly connect the QA150-EXAT to a file server and issue a command from any workstation to back up the server.

TapeWare is a value-added process.
Examples

Compression

- **TapeWare**, which supports DOS and NetWare 286, is a value-added process that lets you directly connect the QA150-EXAT to a file server and issue a command from any workstation to back up the server.

TapeWare lets you directly connect the QA150-EXAT to a file server.
Examples

Compression

▶ **TapeWare**, which supports DOS and NetWare 286, is a value-added process that lets you directly connect the QA150-EXAT to a file server and issue a command from any workstation to back up the server.

TapeWare lets you issue a command from any workstation to back up a file server.
Set-theoretic perspective

Text contains varied elements of information
- Basic Elements (Zhou & Hovy, 2007)
- Pyramids (Nenkova, 2004)
Assume each sentence is a set of such elements

Proposal: Can view text-to-text tasks as set operations
- Marsi & Krahmer (2005), Krahmer et al. (2008)
Strict sentence intersection

Variant of sentence fusion

**Input:** sentence pair

**Output:** only the common information
Strict sentence intersection

An example

- John has a ball.
- Someone has a green ball.
Strict sentence intersection

An example

- John has a ball.
- Someone has a green ball.

Someone has a ball.
Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.

McKeown et al. (2010)
Reduction to entailment

- Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

  ↓ entails

Prosecutors allege that the person was molested after the program aired.

↑ entails

- Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.
Reduction to entailment

Sentence intersection $\Rightarrow$ Mutual entailment generation
\[ A \vdash A \cap B \]
\[ B \vdash A \cap B \]

Sentence union $\Rightarrow$ Mutual entailer generation
\[ A \cup B \vdash A \]
\[ A \cup B \vdash B \]
Outline

Approach
1. Alignment
2. Generalization
3. Abstraction
4. Decoding

Evaluation
Corpus
Metrics
Results

Reflections and future directions
Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.

Monolingual phrase-based alignment

- Thadani & McKeown (2011)
Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.

Monolingual phrase-based alignment
- Thadani & McKeown (2011)
Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.

Monolingual phrase-based alignment

- Thadani & McKeown (2011)
Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.

Ontology-based identification of entailing phrases

Retain more general words

**Future work:** incorporate entailment resources
Generalization

Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.

Ontology-based identification of entailing phrases

Retain more general words

**Future work:** incorporate entailment resources
Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.

Overgenerate potential paths
Drop PPs, dependents
Replace content-bearing NPs, VPs with generic terms
Lattice structure

Aligned

Unaligned

Sentence A

Either A or B

Sentence B
Decoding

1. Beam search
2. Segmented decoding
Decoding
Beam search

LM-based search
- Barzilay & McKeown (2005), Soricut & Marcu (2006)
- Approximate solutions
- Treats all words equally
Decoding

Segmented decoding

New objective:
  All aligned words must be retained
  Maximize LM score of non-aligned “segments”

ILP-based decoding

Exact solutions

**Constraints:**
  - Preserve lattice structure
  - Choose one phrase per alignment
  - Linear ordering via single-commodity flow
Evaluation

Corpus

McKeown et al. (2010)

- 300 overlapping sentence pairs from news clusters
- Five human-generated intersections and unions per instance

Human performance judged for correctness

- Unions: 95%
- Intersections: 54%
Fluency: Is it grammatical?

Validity: Is it mutually entailed?

Coverage: Does it capture all common information?
Ideas:

Separately evaluate whether $A \models A \cap B$ and $B \models A \cap B$

Combine with harmonic mean

Entailment-style tasks on Mechanical Turk
## Metrics

Validity & Fluency

<table>
<thead>
<tr>
<th>Given A</th>
<th>Validity</th>
<th>Fluency</th>
</tr>
</thead>
</table>

## Metrics

### Validity & Fluency

<table>
<thead>
<tr>
<th>Given A</th>
<th>Validity</th>
<th>Fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.188</td>
<td>0.945</td>
</tr>
<tr>
<td>A (aligned words)</td>
<td>0.863</td>
<td>0.563</td>
</tr>
</tbody>
</table>
## Metrics

### Validity & Fluency

<table>
<thead>
<tr>
<th>Given A</th>
<th>Validity</th>
<th>Fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.188</td>
<td>0.945</td>
</tr>
<tr>
<td>A (aligned words)</td>
<td>0.863</td>
<td>0.563</td>
</tr>
<tr>
<td>$A \cap B$ (beam search)</td>
<td>0.729</td>
<td>0.450</td>
</tr>
<tr>
<td>$A \cap B$ (segmented decoder)</td>
<td>0.812</td>
<td>0.504</td>
</tr>
</tbody>
</table>
## Results

### Validity & Fluency

<table>
<thead>
<tr>
<th>Given A</th>
<th>Validity</th>
<th>Fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.188</td>
<td>0.945</td>
</tr>
<tr>
<td>A (aligned words)</td>
<td>0.863</td>
<td>0.563†</td>
</tr>
<tr>
<td>A ∩ B (beam search)</td>
<td>0.729</td>
<td>0.450</td>
</tr>
<tr>
<td>A ∩ B (segmented decoder)</td>
<td>0.812†</td>
<td>0.504</td>
</tr>
<tr>
<td>A ∩ B (oracle combination)</td>
<td>0.813†</td>
<td>0.575†</td>
</tr>
</tbody>
</table>

† not significantly distinct at $p \leq 0.05$
Idea:

Absorption law: $A \cap (A \cup B) = A$

Use unions from McKeown et al. (2010)

Score with MT metrics
## Results

### Coverage

<table>
<thead>
<tr>
<th>Method</th>
<th>BLEU</th>
<th>NIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>aligned words</td>
<td>0.682</td>
<td>11.10</td>
</tr>
<tr>
<td>beam search</td>
<td>0.726</td>
<td>10.53</td>
</tr>
<tr>
<td>segmented decoder</td>
<td>0.818</td>
<td>11.56</td>
</tr>
</tbody>
</table>
Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.
Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.

**Aligned words 1:**
Prosecutors allege that the accuser the program was molested after aired.
Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.

Aligned words 1:
Prosecutors allege that the accuser the program was molested after aired.

Aligned words 2:
Prosecutors allege that the boy was molested after the program aired.
Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.

**Beam search:**
Prosecutors allege that the being, who did something in the program, was molested after something about aired.
Prosecutors allege that the accuser, who appeared in the program, was molested after the show aired.

Prosecutors allege that the boy, a cancer survivor, was molested twice after the program aired.

Beam search:
Prosecutors allege that the being, who did something in the program, was molested after something about aired.

Segmented decoder:
Prosecutors allege that the organism, who did something, was molested after the program aired.
Home Secretary John Reid said Sunday the inquiry would go wherever “the police take it.”

It comes as Home Secretary John Reid said the inquiry into Mr Litvinenko’s poisoning would expand beyond Britain.

**Beam search:**
Home Secretary John Reid said something about the inquiry would move wherever “the something take it”.

**Segmented decoder:**
Home Secretary John Reid said the inquiry would change.
Summary

Sentence intersection
- Connections with entailment
- Segmented decoding approach
- Evaluation strategies for validity and coverage

Future work
- Evaluate existing fusion systems
- Improve fluency with a joint approach
- A principled approach to syntactic decoding
Reflections

Data is paramount!
Simple English Wikipedia seems promising

Separate content selection from generation?
Treat generation as a structured problem