Sentence Compression with Joint Structural Inference

Kapil Thadani & Kathy McKeown
Columbia University

sentence compression

- minimize/bound lexical footprint of a sentence while keeping the most salient information

In 1967 Chapman , who had cultivated a conventional image with his ubiquitous tweed jacket and pipe , by his own later admission stunned a party attended by his friends and future Python colleagues by coming out as a homosexual .

sentence compression

- minimize/bound lexical footprint of a sentence while keeping the most salient information

In 1967 Chapman , who had cultivated a conventional image with his ubiquitous tweed jacket and pipe , by his own later admission stunned a party attended by his friends and future Python colleagues by coming out as a homosexual .

this work

new framework for sentence compression

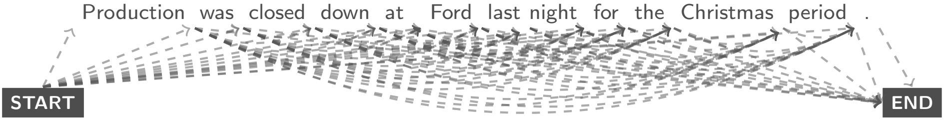
- joint inference over sequential and syntactic structure
- can exploit rich high-order linguistic features
- permits novel dependencies, reordering, etc

structural factorizations

Production was closed down at Ford last night for the Christmas period .

structural factorizations

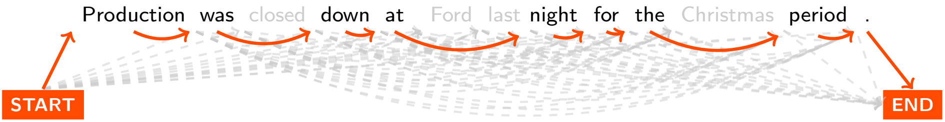
n-grams



⟨ START, Production ⟩	⟨ START, was ⟩	⟨ START, closed ⟩	...
⟨ Production, was ⟩	⟨ Production, closed ⟩	⟨ Production, down ⟩	...
⟨ was, closed ⟩	⟨ was, down ⟩	⟨ was, at ⟩	...
⟨ closed, down ⟩	⟨ closed, at ⟩	⟨ closed, Ford ⟩	...
⟨ down, at ⟩	⟨ down, Ford ⟩	⟨ down, last ⟩	...
⟨ at, Ford ⟩	⟨ at, last ⟩	⟨ at, night ⟩	...
⟨ Ford, last ⟩	⟨ Ford, night ⟩	⟨ Ford, for ⟩	...
⟨ last, night ⟩	⟨ last, for ⟩	⟨ last, the ⟩	...
⋮	⋮	⋮	

structural factorizations

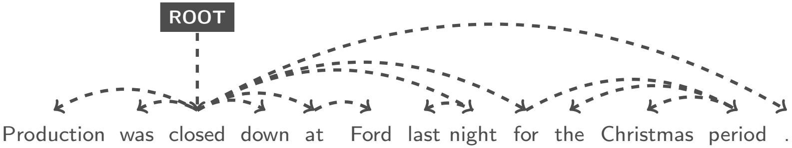
n-grams



⟨ START, Production ⟩	⟨ START, was ⟩	⟨ START, closed ⟩	. . .
⟨ Production, was ⟩	⟨ Production, closed ⟩	⟨ Production, down ⟩	. . .
⟨ was, closed ⟩	⟨ was, down ⟩	⟨ was, at ⟩	. . .
⟨ closed, down ⟩	⟨ closed, at ⟩	⟨ closed, Ford ⟩	. . .
⟨ down, at ⟩	⟨ down, Ford ⟩	⟨ down, last ⟩	. . .
⟨ at, Ford ⟩	⟨ at, last ⟩	⟨ at, night ⟩	. . .
⟨ Ford, last ⟩	⟨ Ford, night ⟩	⟨ Ford, for ⟩	. . .
⟨ last, night ⟩	⟨ last, for ⟩	⟨ last, the ⟩	. . .
⋮	⋮	⋮	

structural factorizations

dependency trees

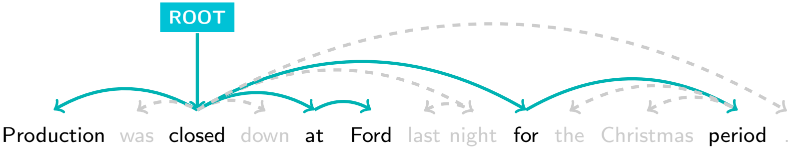


$\langle$ Production $\longleftarrow$ closed $\rangle$
$\langle$ was $\longleftarrow$ closed $\rangle$
$\langle$ closed $\longleftarrow$ ROOT $\rangle$
$\langle$ down $\longleftarrow$ closed $\rangle$
$\langle$ at $\longleftarrow$ closed $\rangle$
$\langle$ Ford $\longleftarrow$ at $\rangle$
$\langle$ last $\longleftarrow$ night $\rangle$
$\langle$ night $\longleftarrow$ closed $\rangle$

$\vdots$

structural factorizations

dependency trees

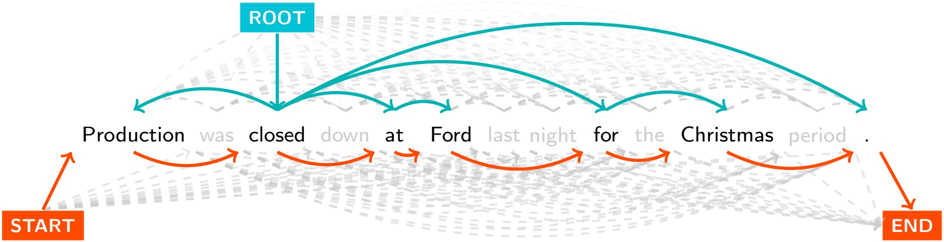


⟨ Production ⟵ closed ⟩
⟨ was ⟵ closed ⟩
⟨ closed ⟵ ROOT ⟩
⟨ down ⟵ closed ⟩
⟨ at ⟵ closed ⟩
⟨ Ford ⟵ at ⟩
⟨ last ⟵ night ⟩
⟨ night ⟵ closed ⟩

⋮

structural factorizations

this work



Goal: recover tokens $\mathbf{x}$, n-gram sequence $\mathbf{y}$ and dependency structure $\mathbf{z}$

joint inference via ILP

objective

$$C = \arg\max_{\mathbf{x},\mathbf{y},\mathbf{z}} \boxed{\sum_i x_i \cdot \mathbf{w}_{tok}^\top \phi(t_i)} \quad \text{token score}$$
$$+ \boxed{\sum_{i,j,k} y_{ijk} \cdot \mathbf{w}_{ngr}^\top \phi(\langle t_i, t_j, t_k \rangle)} \quad \text{ngram score}$$
$$+ \boxed{\sum_{i,j} z_{ij} \cdot \mathbf{w}_{dep}^\top \phi(\langle t_i, t_j \rangle)} \quad \text{dep score}$$

joint inference via ILP

objective

$$
\begin{aligned}
C = \arg\max_{\mathbf{x},\mathbf{y},\mathbf{z}} \quad & \sum_i x_i \cdot \mathbf{w}_{tok}^\top \phi(t_i) && \text{token score} \\
+ \quad & \sum_{i,j,k} y_{ijk} \cdot \mathbf{w}_{ngr}^\top \phi(\langle t_i, t_j, t_k \rangle) && \text{ngram score} \\
+ \quad & \sum_{i,j} z_{ij} \cdot \mathbf{w}_{dep}^\top \phi(\langle t_i, t_j \rangle) && \text{dep score}
\end{aligned}
$$

features

- informativeness
- fluency
- fidelity
- pseudo-normalization

joint inference via ILP

constraints

- tokens $\mathbf{x}$, n-grams $\mathbf{y}$, dependencies $\mathbf{z}$ are consistent
- $\mathbf{x}^\top \mathbf{1} <$ compression rate
- $\mathbf{y}$ forms an acyclic, connected path
- $\mathbf{z}$ specifies a tree

$$\left.\begin{array}{l}\mathbf{y}\text{ forms an acyclic, connected path}\\ \mathbf{z}\text{ specifies a tree}\end{array}\right\}\ ?$$

commodity flow

carried in real-valued variables between pairs of tokens

originate at a single source
$\Rightarrow$ guarantees connectivity

tokens consume commodity

$$\xrightarrow{n} x_i \xrightarrow{n-1}$$

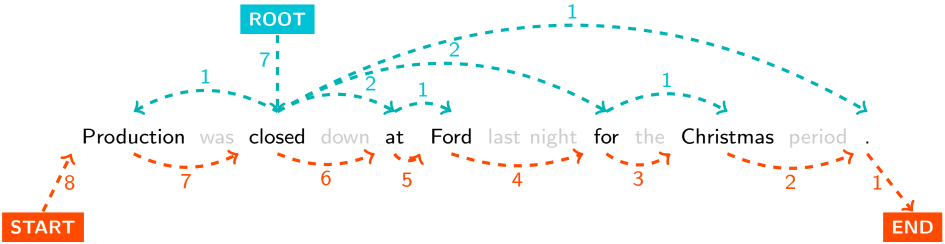
$\Rightarrow$ prevents cycles

ngram y_{ijk} : $x_i \xrightarrow{>0} x_j \xrightarrow{>0} x_k$ adjacency commodity

dependency z_{ij} : $x_i \xrightarrow{>0} x_j$ dependency commodity

commodity flow

backbone



experiments

summary

human-annotated word deletion (Clarke & Lapata, 2008)

- written news, broadcast news

5% absolute gain in {1,2,3,4}-gram F_1 over CL08

- 13-15% relative gain in 4-gram F_1

gains in dependency F_1 against parse of gold compression

content word recall $>$ content word precision

joint model $>$ sequential-only

details at poster

conclusion + future work

holistic: joint production of multiple linguistic structures

expressive: generalizes over previous approaches

permits reordering, multiple input sentences

- new tasks, e.g., sentence fusion

richer dependency structure

- branching fertility, directionality, range

</talk>