Adaptation means ...

- taking *contextual information* into account, for example the domain of a text

Limitations of domain adaptation

- assumption: known, homogenous target domain
- domains defined in terms of corpus boundaries

Context dependence in standard phrase-based models

- translation model: source context within same phrase
- language model context
- no wider sentence or document context
Task: Translation of diverse set of documents

- typical scenario in web translation
- online MT system has to translate diverse documents from different topics & genres
- domain adaptation techniques not easily applicable → adapt dynamically
Old chaps? 3,300-year-old trousers found in China may be world’s oldest

Animal-fur menswear found on the bodies of two mummies in Xinjiang province ‘almost the same shape as today’s trousers’

Two pairs of 3,300-year-old trousers found in China’s far western Xinjiang region may be the world’s oldest, according to state media.

Archaeologists in May found animal-fur menswear on the bodies of two mummies, identified as male shamans in their 40s, the state-run China Daily cited scientists as saying.

An international team is working together to repair and preserve the two pairs, which are the oldest yet discovered with a clear resemblance to modern trousers, the report said.

“They were almost the same shape as today’s trousers,” the report quoted Lu Enguo, a researcher at the Institute of Archaeology in Xinjiang, as saying.

Even older apparel resembling trousers have previously been discovered in the region, but they were made according to a more simple design and lacked a piece of fabric covering the crotch, Lu added.

Archaeologists believe nomads living in the area invented trousers for horse riding. The nomads “at first wore a kind of trousers that only had two legs,” said Xu Dongliang, deputy head of the institute, adding that “crotches were sewed on to the legs, and gradually other styles, such as bloomers, appeared”.

Previously, the oldest pants found with a crotch were just 2,800 years old, the report said.
Two pairs of 3,300-year-old trousers found in China's far western Xinjiang region may be the world's oldest, according to state media.

Archeologists in May found animal-fur menswear on the bodies of two mummies, identified as male shamans in their 40s, the state-run China Daily cited scientists as saying.

An international team is working together to repair and preserve the two pairs, which are the oldest yet discovered with a clear resemblance to modern trousers, the report said.

"They were almost the same shape as today's trousers," the report quoted Lu Enguo, a researcher at the Institute of Archaeology in Xinjiang, as saying.

Even older apparel resembling trousers have previously been discovered in the region, but they were made according to a more simple design and lacked a piece of fabric covering the crotch, Lu added.

Archaeologists believe nomads living in the area invented trousers for horse riding. The nomads "at first wore a kind of trousers that only had two legs," said Xu Dongliang, deputy head of the institute, adding that "crotches were sewed on to the legs, and gradually other styles, such as bloomers, appeared".

Previously, the oldest pants found with a crotch were just 2,800 years old, the report said.
Sentence-level context information

Two pairs of 3,300-year-old trousers found in China’s far western Xinjiang region may be the world’s oldest, according to state media.

Archaeologists in May found animal-fur men’s wear on the bodies of two mummies, identified as male shamans in their 40s, the state-run China Daily cited scientists as saying.

An international team is working together to repair and preserve the two pairs, which are the oldest yet discovered with a clear resemblance to modern trousers, the report said.

“They were almost the same shape as today’s trousers,” the report quoted Lu Enguo, a researcher at the Institute of Archaeology in Xinjiang, as saying.

Even older apparel resembling trousers have previously been discovered in the region, but they were made according to a more simple design and lacked a piece of fabric covering the crotch, Lu added.

Archaeologists believe nomads living in the area invented trousers for horse riding. The nomads “at first wore a kind of trousers that only had two legs,” said Xu Dongliang, deputy head of the institute, adding that “crotches were sewed on to the legs, and gradually other styles, such as bloomers, appeared”.

Previously, the oldest pants found with a crotch were just 2,800 years old, the report said.
Document-level context information

Two pairs of 3,300-year-old trousers found in China’s far western Xinjiang region may be the world's oldest, according to state media.

Archeologists in May found animal-fur menswear on the bodies of two mummies, identified as male shamans in their 40s, the state-run China Daily cited scientists as saying.

An international team is working together to repair and preserve the two pairs, which are the oldest yet discovered with a clear resemblance to modern trousers, the report said.

"They were almost the same shape as today’s trousers," the report quoted Lu Enguo, a researcher at the Institute of Archaeology in Xinjiang, as saying.

Even older apparel resembling trousers have previously been discovered in the region, but they were made according to a more simple design and lacked a piece of fabric covering the crotch, Lu added.

Archaeologists believe nomads living in the area invented trousers for horse riding. The nomads “at first wore a kind of trousers that only had two legs,” said Xu Dongliang, deputy head of the institute, adding that “crotches were sewed on to the legs, and gradually other styles, such as bloomers, appeared.”

Previously, the oldest pants found with a crotch were just 2,800 years old, the report said.
Both types of contextual information

Two pairs of 3,300-year-old trousers found in China’s far western Xinjiang region may be the world’s oldest, according to state media.

Archeologists in May found animal-fur menswear on the bodies of two mummies, identified as male shamans in their 40s, the state-run China Daily cited scientists as saying.

An international team is working together to repair and preserve the two pairs, which are the oldest yet discovered with a clear resemblance to modern trousers, the report said.

“They were almost the same shape as today’s trousers,” the report quoted Lu Enguo, a researcher at the Institute of Archaeology in Xinjiang, as saying.

Even older apparel resembling trousers have previously been discovered in the region, but they were made according to a more simple design and lacked a piece of fabric covering the crotch, Lu added.

Archeologists believe nomads living in the area invented trousers for horse riding. The nomads “at first wore a kind of trousers that only had two legs,” said Xu Dongliang, deputy head of the institute, adding that “crotches were sewed on to the legs, and gradually other styles, such as bloomers, appeared”.

Previously, the oldest pants found with a crotch were just 2,800 years old, the report said.
Related work

Using source side context

• integrate **WSD classifier** with local word and POS features into MT systems [Carpuat and Wu, 2007, Chan et al., 2007]

Using topic models

• learn **document-level topics** to adapt translation features [Eidelman et al., 2012, Hasler et al., 2014]

• learn **sentence-level topics**, compute **similarity** of training and test sentences [Banchs and Costa-jussà, 2011]

Cross-domain adaptation

• define **vector profiles** of phrase pairs over training corpora, adapt to an **in-domain development set** [Chen et al., 2013]
Proposed approach

Capture **semantics** of translation units

- unit: phrase pair
- abstract from lexical forms in training contexts
- encode contextual information independent of corpus boundaries:
  lower-dimensional topic mixture

Test time: compute context **similarity**

- when translating a document, first measure its semantic content
- during decoding, favour semantically similar translation units (dynamic adaptation)
Phrase Pair Topic Model

How to learn semantic representations?

• represent each phrase pair as distributional profile: “document” containing all context words

• collect all source context words in local training contexts of a phrase pair
How to learn semantic representations?

- represent each phrase pair as distributional profile: “document” containing all context words
- collect all source context words in local training contexts of a phrase pair
- learn latent representation $\theta_{pp}$ for each phrase pair
For each of $P$ phrase pairs $pp_i$ in the collection

1. Draw a topic distribution from an asymmetric Dirichlet prior,
   $\theta_p \sim \text{Dirichlet}(\alpha_0, \alpha \ldots \alpha)$.

2. For each position $c$ in the distributional profile of $pp_i$,
   draw a topic from that distribution,
   $z_{p,c} \sim \text{Multinomial}(\theta_p)$.

3. Conditioned on topic $z_{p,c}$, choose a context word
   $w_{p,c} \sim \text{Multinomial}(\psi_{z_{p,c}})$.
For each of \( L \) test sentences (local) in the collection

1. Draw a topic distribution from an asymmetric Dirichlet prior, 
   \( \theta_l \sim \text{Dirichlet}(\alpha_0, \alpha \ldots \alpha) \).

2. For each position \( c \) in the test sentence, draw a topic from that distribution, 
   \( z_{l,c} \sim \text{Multinomial}(\theta_l) \).

3. Conditioned on topic \( z_{l,c} \), choose a context word 
   \( w_{l,c} \sim \text{Multinomial}(\psi_{z_{l,c}}) \).
Learned topic representations

- some ambiguity remains: both *kernel* and *core* occur in *IT* contexts as translations of *noyau*
Applying the model to unseen test documents

noyau → kernel

version
défaut
recompiler

θ_p → cosine(θ_p, θ_c)

noyau → nucleus

cellule
atomique
microscopique

θ_g

θ_l

global test context
supporter
patcher
linux

local test context
erreur
module
annonce
Types of similarity features

- **local**: similarity feature using local context
- **global**: similarity feature using global context
- **⊕**: log-linear combination
- **⊕**: additive combination of topic vectors
- **⊗**: multiplicative combination of topic vectors
- **⊛**: weighted combination that depends on sentence length
Training data (French-English)

<table>
<thead>
<tr>
<th>Data</th>
<th>Mixed</th>
<th>Commoncrawl</th>
<th>NewsCom</th>
<th>Ted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>354K (6450)</td>
<td>110K</td>
<td>103K</td>
<td>140K</td>
</tr>
<tr>
<td>Dev</td>
<td>2453 (39)</td>
<td>818</td>
<td>817</td>
<td>818</td>
</tr>
<tr>
<td>Test</td>
<td>5664 (112)</td>
<td>1892</td>
<td>1878</td>
<td>1894</td>
</tr>
</tbody>
</table>

Baselines (trained with Moses toolkit)

- Concatenation baseline *(no adaptation)*
- Domain adaptation baselines *(train and test domains known)*
  - LIN-TM: linear phrase table interpolation [Sennrich, 2012]
  - FILLUP: phrase table fillup [Bisazza et al., 2011]
  - one model per domain

Topic-adapted model

- add topic-adapted features per sentence-level phrase table
## Combining local and global context

<table>
<thead>
<tr>
<th>Model</th>
<th>Mixed</th>
<th>Cc</th>
<th>Nc</th>
<th>Ted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>26.86</td>
<td>19.61</td>
<td>29.42</td>
<td>31.88</td>
</tr>
<tr>
<td>+ global</td>
<td>27.27</td>
<td>20.12</td>
<td>29.48</td>
<td>32.55</td>
</tr>
<tr>
<td>+ local</td>
<td>*27.43</td>
<td>20.18</td>
<td>29.65</td>
<td>32.79</td>
</tr>
<tr>
<td>⊕ local</td>
<td>*27.49</td>
<td><strong>20.30</strong></td>
<td><strong>29.66</strong></td>
<td><strong>32.76</strong></td>
</tr>
<tr>
<td>○ local</td>
<td>27.34</td>
<td>20.24</td>
<td>29.61</td>
<td>32.50</td>
</tr>
<tr>
<td>⊗ local</td>
<td>*27.45</td>
<td>20.22</td>
<td>29.51</td>
<td>32.79</td>
</tr>
<tr>
<td>+0.63</td>
<td>+0.69</td>
<td>+0.24</td>
<td>+0.88</td>
<td></td>
</tr>
</tbody>
</table>

- model with 50 latent topics
Examples of ambiguous source phrases

**Source:** Le noyau contient de nombreux pilotes, afin de fonctionner chez la plupart des utilisateurs.

**Reference:** The precompiled kernel includes a lot of drivers, in order to work for most users.

**Source:** Il est prudent de consulter les pages de manuel ou lesfaq spécifiques à votre os.

**Reference:** It’s best to consult the man pages or faqs for your os.

**Source:** Nous fournissons nano (un petit éditeur), vim (vi amélioré), qemacs (clone de emacs), elvis, joe .

**Reference:** Nano (a lightweight editor), vim (vi improved), qemacs (emacs clone), elvis and joe.

**Source:** Elle a introduit des politiques [...] à coté des relations de gouvernement à gouvernement traditionnelles.

**Reference:** She has introduced policies [...] alongside traditional government-to-government relations.
Examples of ambiguous source phrases

**Source:** Le *noyau* contient de nombreux pilotes, afin de fonctionner chez la plupart des utilisateurs.

**Reference:** The precompiled *kernel* includes a lot of drivers, in order to work for most users.

**Source:** Il est prudent de consulter les pages de manuel ou les faq spécifiques à votre *os*.

**Reference:** It’s best to consult the man pages or faqs for your *os*.

<table>
<thead>
<tr>
<th>Model</th>
<th>noyau →</th>
<th>os →</th>
<th>elvis →</th>
<th>relations →</th>
</tr>
</thead>
<tbody>
<tr>
<td>global</td>
<td>kernel</td>
<td>os</td>
<td>the king</td>
<td>relationship</td>
</tr>
<tr>
<td>local</td>
<td>nucleus</td>
<td>bones</td>
<td>elvis</td>
<td>relations</td>
</tr>
<tr>
<td>global ⊕ local</td>
<td>kernel</td>
<td>os</td>
<td>the king</td>
<td>relations</td>
</tr>
</tbody>
</table>
Comparison with domain adaptation

<table>
<thead>
<tr>
<th>Adaptation</th>
<th>Model</th>
<th>Mixed</th>
<th>Cc</th>
<th>Nc</th>
<th>Ted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain-adapted</td>
<td>LIN-TM</td>
<td>27.24</td>
<td>19.61</td>
<td>29.87</td>
<td>32.73</td>
</tr>
<tr>
<td></td>
<td>FILLUP</td>
<td>27.12</td>
<td>19.36</td>
<td>29.78</td>
<td>32.71</td>
</tr>
<tr>
<td>Topic-adapted</td>
<td>global⊕local</td>
<td>*27.49</td>
<td>20.30</td>
<td>29.66</td>
<td>32.76</td>
</tr>
<tr>
<td></td>
<td>&gt;LIN-TM</td>
<td>+0.25</td>
<td>+0.69</td>
<td>-0.21</td>
<td>+0.03</td>
</tr>
<tr>
<td></td>
<td>&gt;FILLUP</td>
<td>+0.37</td>
<td>+0.94</td>
<td>-0.12</td>
<td>+0.05</td>
</tr>
</tbody>
</table>

- Commoncrawl documents are the most diverse set
- News Commentary documents are the least diverse set
Combination with a document similarity feature

- similar to [Banchs and Costa-jussà, 2011]
- compute $\text{max sim}$ score for each applicable phrase pair:
  - noyau $\rightarrow$ kernel, noyau $\rightarrow$ nucleus, noyau $\rightarrow$ core
## Combination with a document similarity feature

<table>
<thead>
<tr>
<th>Model</th>
<th>Mixed</th>
<th>Cc</th>
<th>Nc</th>
<th>Ted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>26.86</td>
<td>19.61</td>
<td>29.42</td>
<td>31.88</td>
</tr>
<tr>
<td>+ docSim</td>
<td>27.22</td>
<td>20.11</td>
<td>29.63</td>
<td>32.40</td>
</tr>
<tr>
<td>+ global ⊕ local</td>
<td>*27.58</td>
<td>20.34</td>
<td>29.71</td>
<td>32.96</td>
</tr>
<tr>
<td>+ global ⊙ local</td>
<td>*27.60</td>
<td>20.35</td>
<td>29.70</td>
<td>33.03</td>
</tr>
<tr>
<td></td>
<td>+0.74</td>
<td>+0.74</td>
<td>+0.28</td>
<td>+1.15</td>
</tr>
</tbody>
</table>
Conclusions

• introduced Phrase Pair Topic model:
  • learns semantic representations for translation units
  • provides compact way of storing contextual information
• translation model is dynamically adapted to local/global test context
• adaptation with similarity features → efficient at test time
• combining information from different scopes and topic granularity performs better than each feature separately
Thank you!


