Statistical Machine Translation

Nadir Durrani

21-November-2014
Machine Translation

• Problem: Automatic translation the foreign text:
Open Problems in Machine Translation

• Ambiguity in translation
  – He deposited money in a bank account with a high interest rate
  – Sitting on the bank of the Mississippi, a passing ship piqued his interest
  – How do we find the right meaning and thus translation?
  – Context should be helpful

• Phrase translation problem
  It’s raining cats and dogs
  موسلادهار بارش بو ربي پي
Open Problems in Machine Translation

• Morphological Differences
  
  Collins et. al (2005)
  Koehn and Hoang (2007)
  Fraser et. al (2012)

• Structural Differences
  
  Galley and Manning (2008)
  Green et. al (2010)
  Durrani et al (2011)

uvian احسانة
And be kind with your parents

 photographed والد + والد + ين

Diese Woche ist die grüne Hexe zu Haus

The green witch is at home this week
The Grand Plan
Different Machine Translation Frameworks

- Rule-based

- Empirical
  - Example-based machine translation
  - Statistical machine translation

- Hybrid Machine Translation
Rosetta Stone

- Egyptian language was a mystery for centuries
- The Rosetta stone is written in three scripts
  - Hieroglyphic (used for religious documents)
  - Demotic (common script of Egypt)
  - Greek (language of rulers of Egypt at that time)
Parallel Data
Parallel Data

- UN and European Parliamentary Proceedings
  - German, French, Spanish etc.

- News Corpus and Common Crawl Data

- NIST Data (Arabic, Chinese)
Noisy Channel Model

- Decipherment problem

  Warren Weaver: “When I look at an article in Russian, I say: This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode”

- Bayes Rule: \( p(E \mid F) = p(F \mid E) \times p(E) / p(F) \)
  \( e_{\text{best}} = \arg\max p(E \mid F) = \arg\max p(F \mid E) \times p(E) \)
Statistical Machine Translation

From Koehn 2008. University of Edinburgh
Word-based Models (Brown et. al 1992)

• Word alignments
  – If we had word alignment we can learn translation model
  – If we knew model parameters we can learn word alignments
  – Chicken and Egg problem: EM-algorithm
Word-based Models (Brown et. al 1992)

- Word alignments
  - If we had word alignment we can learn translation model
  - If we knew model parameters we can learn word alignments
  - Chicken and Egg problem: EM-algorithm

- IBM Models
  - Model 1 (Word-to-word translation)
  - Model 2 (+additional distortion model)
  - Model 3 (+fertility: insertions, deletions)
  - Model 4 (+improved distortion model)
  - Model 5 (+non-deficient Model 4)
Phrase-based Model (Och/Koehn et. al 2003)

- State-of-the-art for many language pairs

  Morgen  fliege ich  nach Kanada  zur Konferenz

  Tomorrow  I will fly  to the conference  in Canada

- Translation $p(f|e)$ is estimated through phrases instead of words

From Koehn 2008
Benefits of phrase-based SMT

1. Local reordering

Morgen \rightarrow fliege ich \rightarrow nach Kanada

Tomorrow \rightarrow I will fly \rightarrow to Canada

2. Idioms

in den sauren Apfel beißen

to bite the bullet

er \rightarrow hat ein Buch gelesen

he \rightarrow read a book

lesen Sie mit

read with me

3. Discontinuities in phrases

4. Insertions and deletions
Left-to-Right Stack Decoding
Phrasal Extraction

[Diagram showing a grid with various words and 'x' markers indicating phrasal extraction.]
Reordering Sub-Model (Koehn et. al 2005)

<table>
<thead>
<tr>
<th>Morgan</th>
<th>fleige</th>
<th>ich</th>
<th>nach</th>
<th>Kanada</th>
<th>zur</th>
<th>Konferenz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tomorrow</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>will</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fly</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the conference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Orientation-based model
  - Monotonic (M), Swap (S), Discontinuous (D)
Syntax-based Models

- Phrase-based model can not capture long distance dependencies

- Language is hierarchical and not flat
String-to-Tree Model (Galley et. al 2004, 2006)
Tree-to-tree Model (Zhang et. al 2008)

From Koehn 2010. University of Edinburgh
Chart-based Decoding
Syntax-based Models

- Much progress, but success only for some language pairs

- Many open questions
  - Syntax on source/target/both?
  - Can we learn syntax unsupervised?
  - Phrase structure or dependency structure?
  - What grammar rules should be extracted?
  - Soft or hard constraints?
  - Feature design
Semantic-based Model

- What do existing models don’t capture
  - Who did what to whom
  - Preservation of meaning can be more important than grammaticality/fluency

- ISI (Kevin Knight’s Group)
  - Using semantic role labeling
  - Jones et. al (2012)
Log-linear Model (Och and Ney 2004)

- Typical features in Phrase-based Model
  - 4 Translation model features
  - 6 Reordering model features
  - Length Bonus
  - Phrase Bonus
  - Language Model

\[ e_{\text{best}} = \arg \max p(E | F) = \arg \max p(F | E) \times p(E) \]

- Tuning Algorithms
  - MERT (Och and Ney, 2004)
  - PRO (Hopkins and May, 2011)
  - MIRA (Chiang, 2012)

- 11,001 New Features for Statistical Machine Translation (Chiang et. al 2009)
Log-linear Model (Och and Ney 2004)
Open Problems in Machine Translation

• Evaluation
  – How good is a given machine translation system?
  – Hard problem, since many different translations acceptable
  – Evaluation metrics
    • Subjective judgments by human evaluators
    • Automatic evaluation metrics

• Automatic Evaluation Metrics
  – BLEU (Papineni et. al 2002)
  – METEOR (Banerjee and Lavie 2005)
  – WER/TER (Error rate)
Open Problems in Machine Translation

这个 机场 的 安全 工作 由 以色列 方面 负责．

Israeli officials are responsible for airport security.
Israel is in charge of the security at this airport.
The security work for this airport is the responsibility of the Israel government.
Israeli side was in charge of the security of this airport.
Israel is responsible for the airport’s security.
Israel is responsible for safety work at this airport.
Israel presides over the security of the airport.
Israel took charge of the airport security.
The safety of this airport is taken charge of by Israel.
This airport’s security is the responsibility of the Israeli security officials.

(a typical example from the 2001 NIST evaluation set)
Open Problems in Machine Translation

• Human judgment
  – given: machine translation output
  – given: source and/or reference translation
  – task: assess the quality of machine translation output

• Metrics
  – Adequacy: Does the output convey the same meaning as the input sentence? Is part of the message lost, added, or distorted?
  – Fluency: Is the output good fluent English?
Open Problems in Machine Translation

• Domain Adaptation
  – Training data (News corpus, Europarl, Common Crawl Data)
  – Test data (Education domain, Medical domain)
  – Interpolation Models (Foster and Kuhn 2007)
  – MML Filter (Axelrod et. al 2011)
  – Domain Features (Hasler et. al 2012)

• OOV word translation
  – NE translation (Onaizan and Knight 2002)
  – NE disambiguation (Hermjakob et. al 2008)
  – Unsupervised Transliteration (Sajjad et. al 2012, Durrani et. al 2014)
    • Closely related languages (Durrani et. al 2011, Durrani and Koehn 2014)
Open Problems in Machine Translation

• Decoding Algorithms
  – Stack Decoding (Tillmann et. al 1997)
  – Efficient A* Decoding (Och et. al 2001)
  – Pruning Methods (Moore and Quirk 2007)

• Language Model
  – The house is big (good)
  – The house is xxl (worse)
  – House big is the (bad)
  – Markov-based language models with Kneser-Ney Smoothing
    • Considers history of 4 previous words
  – Syntax-based Language Models (Charniak et. al 2003)
Open Problems in Machine Translation

- Big Data and Scaling to Big Data
  - Parallel data (Billions of words) (Smith et. al 2013)
  - English monolingual data (trillions of words)
  - Randomized data structures (Talbot and Osborne 2007)
    - Developed at Edinburgh now used at Google
  - Distributed Systems
    - Distribute models over 100 machines
  - Efficient data-structures
    - Compact Phrase-tables (Junczys-Dowmunt 2012)
    - Scalable Language Model estimation (Heafield 2013)
      - Prefixes, back-off links in language models, binarization
Open Problems in Machine Translation

• Computer Assisted Translation
  – Machine Translation makes inroads in human translation industry
  – CASMACAT/MateCat Projects in Edinburgh
Why Do Machine Translation?

- Assimilation – reader initiates translation, wants to know the content (Gistable)
- Translation in Hand-held devices
- Post-editing (editable)
- User manuals in different languages, high quality translation (publishable)
- Integration with other NLP applications
  - Speech Technologies
  - Cross lingual information retrieval
- US Defense
  - Arabic-English post 9/11
  - Urdu-English, Pashto-English 2008
  - Dialectal Arabic (Egyptian, Labenese, Iraqi 2009-present)
  - Russian-English (2013-2014)
Open Source Resources

• Toolkits
  – Moses (Koehn et. al 2007), Phrasal (Cerr et. al 2010), NCode (Crego et. al 2011)
  – GIZA++ (Word Alignments)
  – SRILM, IRSTLM, KENLM, LMPLZ (Language Model)

• Data
  – French-English 39M
  – Chinese-English Spanish-English, Czech-English 15M
  – Arabic-English
  – German-English 5.5M
  – Urdu-English/Hindi-English ~300K

• Parsers
  – English, French, German
Thank you !!!

- Most of the slides are borrowed from Philipp Koehn
References


