MT – Final thoughts

David Kauchak
CS159 – Spring 2023

Some slides adapted from
Philipp Koehn
School of Informatics
University of Edinburgh

Kevin Knight
USC/Information Sciences Institute

Dan Klein
USC/Computer Science Department

MT – Language translation

Yo quiero
Yo quiero
Taco Bell
El viejo está feliz porque ha pescado muchas veces.

Language translation

https://www.youtube.com/watch?v=Q6jzL_Oy2lQ
https://www.youtube.com/watch?v=vV1SkTdizzI

Word-alignment Evaluation

The old man is happy. He has fished many times.

Word-alignment Evaluation

How good of an alignment is this?
How can we quantify this?
The old man is happy. He has fished many times.

El viejo está feliz porque ha pescado muchos veces.

How can we quantify this?

Precision: $\frac{6}{7}$  
Recall: $\frac{6}{10}$

What kind of Translation Model?

Mary did not slap the green witch

- Word-level models
- Phrasal models
- Syntactic models
- Semantic models

Maria no dio una bofetada a la bruja verde
Phrasal translation model

The models define probabilities over inputs

\[ p(f \mid e) \]

1. Sentence is divided into phrases

<table>
<thead>
<tr>
<th>Morgen</th>
<th>fliege</th>
<th>ich</th>
<th>nach Kanada</th>
<th>zur Konferenz</th>
</tr>
</thead>
</table>

2. Phrases are translated (avoids a lot of weirdness from word-level model)

| Tomorrow | will fly | in Canada | to the conference |

3. Phrases are reordered

Phrase table

naturerlich

<table>
<thead>
<tr>
<th>Translation</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>of course</td>
<td>0.5</td>
</tr>
<tr>
<td>naturally</td>
<td>0.3</td>
</tr>
<tr>
<td>of course,</td>
<td>0.15</td>
</tr>
<tr>
<td>, of course,</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Phrasal translation model

The models define probabilities over inputs

\[ p(f | e) \]

Morgen fliege ich nach Kanada zur Konferenz

Tomorrow I will fly to the conference in Canada

Advantages?

Advantages of Phrase-Based

Many-to-many mappings can handle non-compositional phrases

Easy to understand

Local context is very useful for disambiguating

– “Interest rate” \( \rightarrow \) …
– “Interest in” \( \rightarrow \) …

The more data, the longer the learned phrases

– Sometimes whole sentences!

Syntax-based models

Benefits?
Syntax-based models

Benefits
– Can use syntax to motivate word/phrase movement
– Could ensure grammaticality

Two main types:
• p(foreign string | English parse tree)
• p(foreign parse tree | English parse tree)

Why always English parse tree?

Tree to string rules examples

1. DT(these) → 这
2. VBP(include) → 中包括
3. VBP(includes) → 中包括
4. NNP(France) → 法国
5. CC(and) → 和
6. NNP(Russia) → 俄罗斯
7. IN(of) → 的
8. NP(NNS(astronauts)) → 宇航员
9. PUNC(,) → 。
10. NP(x0:DT, CD(7), NNS(people)) → x0 , 7人
11. VP(VBG(coming), PP(In(from), x0:NP)) → 来自 , x0
12. IN(from) → 来自
13. NP(x0:NNP, x1:CC, x2:NNP) → x0 , x1 , x2
14. VP(x0:VBP, x1:NP) → x0 , x1
15. S(x0:NP, x1:VP, x2:PUNC) → x0 , x1 , x2
16. NP(x0:NP, x1:VP) → x1 , 的 , x0
17. NP(DT("the"), x0:JJ, x1:NN) → x0 , x1

Tree to string rules examples

Both VBP("include") and VBP("includes") will translate to "中包括” in Chinese.

Contiguous phrase pair substitution rules

Higher-level rules
**Tree Transformations**

1. **DT(these) → 这**
2. **VBP(include) → 中包括**
3. **VBP(includes) → 中包括**
4. **NNP(France) → 法国**
5. **CC(and) → 和**
6. **NNP(Russia) → 俄罗斯**
7. **IN(of) → 的**
8. **NP(NNS(astronauts)) → 宇航员**
9. **PUNCT(.) → .**
10. **NP(x0:DT, CD(7), NNS(people) → x0, 7人**
11. **VP(VBG(coming), PP(IN(from), x0:NP)) → 来自, x0**
12. **IN(from) → 来自**
13. **NP(x0:NNP, x1:CC, x2:NNP) → x0, x1, x2**
14. **VP(x0:VBP, x1:NP) → x0, x1**
15. **S(x0:NP, x1:VP, x2:PUNC) → x0, x1, x2**
16. **NP(x0:NP, x1:VP) → x1, 的, x0**
17. **NP(DT("the"), x0:JJ, x1:NN) → x0, x1**

---

**Tree Transformations**

1. **DT(these) → 这**
2. **VBP(include) → 中包括**
3. **VBP(includes) → 中包括**
4. **NNP(France) → 法国**
5. **CC(and) → 和**
6. **NNP(Russia) → 俄罗斯**
7. **IN(of) → 的**
8. **NP(NNS(astronauts)) → 宇航员**
9. **PUNCT(.) → .**
10. **NP(x0:DT, CD(7), NNS(people) → x0, 7人**
11. **VP(VBG(coming), PP(IN(from), x0:NP)) → 来自, x0**
12. **IN(from) → 来自**
13. **NP(x0:NNP, x1:CC, x2:NNP) → x0, x1, x2**
14. **VP(x0:VBP, x1:NP) → x0, x1**
15. **S(x0:NP, x1:VP, x2:PUNC) → x0, x1, x2**
16. **NP(x0:NP, x1:VP) → x1, 的, x0**
17. **NP(DT("the"), x0:JJ, x1:NN) → x0, x1**
MT Evaluation

How do we do it?

What data might be useful?

Source only

Manual:
- SSER (subjective sentence error rate)
- Correct/Incorrect
- Error categorization

Extrinsic:
Objective usage testing

Automatic:
- WER (word error rate)
- BLEU (Bilingual Evaluation Understudy)
- NIST

MT Evaluation exercise

Play with an MT system

1. Find a few examples of the system doing interesting (surprising?) “good” translations.

2. Find some examples of the system making mistakes (consider, idioms and common expressions)

Automatic Evaluation

Common NLP/machine learning/AI approach

Train sentence pairs

Testing sentence pairs
The U.S. island of Guam is maintaining a high state of alert after receiving an email from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological/chemical attack against public places such as the airport.

The U.S. island of Guam is maintaining a high state of alert after receiving an email from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological/chemical attack against public places such as the airport.

Machine translation: The American (?) international airport and its the office all receives one calls self the sand Arab rich business (?) and so on electronic mail , which sends out: The threat will be able after public place and so on the airport to start the biochemistry attack ; (?) highly alerts after the maintenance.

Reference translation: The American (?) international airport and its the office all receives one calls self the sand Arab rich business (?) and so on electronic mail, which sends out: The threat will be able after public place and so on the airport to start the biochemistry attack ; (?) highly alerts after the maintenance.

Reference translation: The American (?) international airport and its the office all receives one calls self the sand Arab rich business (?) and so on electronic mail, which sends out: The threat will be able after public place and so on the airport to start the biochemistry attack ; (?) highly alerts after the maintenance.

Machine translation: It is a guide to action which ensures that the military always obey the commands of the party.

Candidate 1: It is a guide to action which ensures that the military always obey the commands of the party.

Reference 1: It is a guide to action which ensures that the military always obey the commands of the party.

Reference 2: It is the guiding principle which guarantees the military forces always being under the command of the Party.

Reference 3: It is the practical guide for the army always to heed directions of the party.

What percentage of machine n-grams can be found in the reference translations? Do unigrams, bigrams and trigrams.
N-gram precision example

Candidate 1: It is a guide to action which ensures that the military always obey the commands of the party.

Reference 1: It is a guide to action that ensures that the military will forever heed Party commands.
Reference 2: It is the guiding principle which guarantees the military forces always being under the command of the Party.
Reference 3: It is the practical guide for the army always to heed directions of the party.

Unigrams: 17/18

N-gram precision example

Candidate 1: It is a guide to action which ensures that the military always obey the commands of the party.

Reference 1: It is a guide to action that ensures that the military will forever heed Party commands.
Reference 2: It is the guiding principle which guarantees the military forces always being under the command of the Party.
Reference 3: It is the practical guide for the army always to heed directions of the party.

Unigrams: 17/18
Bigrams: 10/17

N-gram precision example 2

Candidate 2: It is to ensure the army forever hearing the directions guide that party commands.

Reference 1: It is a guide to action that ensures that the military will forever heed Party commands.
Reference 2: It is the guiding principle which guarantees the military forces always being under the command of the Party.
Reference 3: It is the practical guide for the army always to heed directions of the party.

Unigrams: 17/18
Bigrams: 10/17
Trigrams: 7/16
N-gram precision example 2

Candidate 2: *It is to ensure the army forever hearing the directions guide that party commands.*

Reference 1: It is a guide to action that ensures that the military will forever heed Party commands.
Reference 2: It is the guiding principle which guarantees the military forces always being under the command of the Party.
Reference 3: It is the practical guide for the army always to heed directions of the party.

Unigrams: 12/14
Bigrams: 4/13
Trigrams: 1/12

N-gram precision example 2

Candidate 2: *It is to ensure the army forever hearing the directions guide that party commands.*

Reference 1: It is a guide to action that ensures that the military will forever heed Party commands.
Reference 2: It is the guiding principle which guarantees the military forces always being under the command of the Party.
Reference 3: It is the practical guide for the army always to heed directions of the party.

Unigrams: 12/14
Bigrams: 4/13
Trigrams: 1/12

Any problems/concerns?
N-gram precision example

Candidate 3: the
Candidate 4: It is a

Reference 1: It is a guide to action that ensures that the military will forever heed Party commands.
Reference 2: It is the guiding principle which guarantees the military forces always being under the command of the Party.
Reference 3: It is the practical guide for the army always to heed directions of the party.

What percentage of machine n-grams can be found in the reference translations? Do unigrams, bigrams and trigrams.

BLEU Evaluation Metric
(Papineni et al, ACL-2002)

Reference (human) translation:
- The U.S. island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological attack against public places such as the airport.

Machine translation:
- The American [?]: international airport and its office all receives one calls self the sand Arab rich business [?] and so on electronic mail , which sends out ; The threat will be able after public place and so on the biochemistry attack . [?] Highly alerts after the maintenance.

N-gram precision (score is between 0 & 1)
- What percentage of machine n-grams can be found in the reference translation?
- Not allowed to use same portion of reference translation twice (can’t cheat by typing out “the the the the the”)

Brevity penalty
- Can’t just type out single word “the” (precision 1.0)

*** Amazingly hard to “game” the system (i.e., find a way to change machine output so that BLEU goes up, but quality doesn’t)

BLEU Tends to Predict Human Judgments

slide from G. Doddington (NIST)

BLEU: Problems?

Doesn’t care if an incorrectly translated word is a name or a preposition
- gave it to Albright (reference)
- gave it at Albright (translation #1)
- gave it to altar (translation #2)

What happens when a program reaches human level performance in BLEU but the translations are still bad?
- maybe sooner than you think …
Appendix A

Input: corpus of English/Foreign sentence pairs (no alignment)

for some number of iterations:
  for (E, F) in corpus:
    for e in E:
      for f in F:
        \[ p(f \rightarrow e) = \frac{p(f|e)}{\sum_{e' \in E} p(f|e')} \]
        count(e,f) += p(f \rightarrow e)
        count(e) += p(f \rightarrow e)

for all \((e,f)\) in count:
  \[ p(f|e) = \frac{\text{count}(e,f)}{\text{count}(e)} \]

48

49

Step 1: calculate \(p(f \rightarrow e)\) for all pairs of words in the two sentences (assume \(p(f|e)\) is a constant for all \(f,e\))

Pair 1:
- E: green house
  - F: casa verde

Pair 2:
- E: the house
  - F: la casa

50

51

Step 2: aggregate the counts

for (E, F) in corpus:
  for e in E:
    for f in F:
      \[ p(f \rightarrow e) = \frac{p(f|e)}{\sum_{e' \in E} p(f|e')} \]
      count(e,f) += p(f \rightarrow e)
      count(e) += p(f \rightarrow e)

for all \((e,f)\) in count:
  \[ p(f|e) = \frac{\text{count}(e,f)}{\text{count}(e)} \]

Step 3: recalculate \(p(e|f)\) 
Appendix A

Input: corpus of English/Foreign sentence pairs (no alignment)

for some number of iterations:

\[
\text{for } (E, F) \text{ in corpus: for } e \text{ in } E:\n\frac{p_f \rightarrow e}{\sum_p p_f \rightarrow e} = \frac{\text{count}(e, f)}{\text{count}(e)}
\]

for all \((e, f)\) in count:

\[
p_f \rightarrow e = \frac{\text{count}(e, f)}{\text{count}(e)}
\]

Worksheet

Pair 1:

\[
\begin{align*}
\text{p(casa \rightarrow green)} &= 1/2 \\
\text{p(casa \rightarrow house)} &= 1/2 \\
\text{p(verde \rightarrow green)} &= 1/4 \\
\text{p(verde \rightarrow house)} &= 1/4 \\
\end{align*}
\]

\[
\text{count(green, casa)} = 1/4 + 1/4 = 1/2
\]

\[
\text{count(verde, casa)} = 1/4 + 1/4 = 1/2
\]

\[
\text{count(house, casa)} = 1/4 + 1/4 + 1/4 + 1/4 = 1
\]

Then, calculate the probabilities by normalizing the counts.