Machine Translation Systems

Christopher Manning
CS224N / Ling 284
[Based on slides by Kevin Knight, Dan Klein, Dan Jurafsky]

Machine translation:
- The U.S. island of Guam is maintaining a high state of alert. The airport and other public places as airport. Guam International Airport and its offices are maintaining a high state of alert. The airport.
- A high state of alert.

Reference (human) translation:
- The Saudi Arabian Osama bin Laden named Laden, which threatens to launch a biochemical attack on such places as airport. Guam
- The US International Airport of Guam is maintaining a high state of alert.
- The U.S. island of Guam is maintaining a high state of alert.

Multiple Reference Translations

Multiple Reference Translations

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

Reference (human) translation:
- The U.S. island of Guam is maintaining a high state of alert.
- The airport and other public places as airport. Guam
- A high state of alert.
- The airport.

Machine translation:
- The U.S. island of Guam is maintaining a high state of alert. The airport and other public places as airport. Guam
- A high state of alert.
- The airport.

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

Reference (human) translation:
- The U.S. island of Guam is maintaining a high state of alert.
- The airport and other public places as airport. Guam
- A high state of alert.
- The airport.

Machine translation:
- The U.S. island of Guam is maintaining a high state of alert.
- The airport and other public places as airport. Guam
- A high state of alert.
- The airport.

BLEU in Action

341

the gunman was shot to death by the police.

Reference Translation

the gunman was shot to death by the police.

green = 4-gram match
red = word not matched

21

Reference Translation

the gunman was shot to death by the police.

green = 4-gram match
red = word not matched
Initial results showed that BLEU predicts human judgments well

\[ R^2 = 0.88 \%
\]

\[ R^2 = 0.90 \%
\]

Quiz question!

MT Hypothesis: *the gunman was shot dead by police*.

- Ref 1: The gunman was shot to death by the police.
- Ref 2: The cops shot the gunman dead.

- What is the:
  - Unigram precision?
  - Trigram precision?

Note: punctuation tokens are counted in calculation but not sentence boundary tokens

Automatic evaluation of MT

- People started optimizing their systems to maximize BLEU score
  - BLEU scores improved rapidly
  - The correlation between BLEU and human judgments of quality went way, way down
  - StatMT BLEU scores now approach those of human translations but their true quality remains far below human translations
- Coming up with automatic MT evaluations has become its own research field
  - There are many proposals: TER, METEOR, MaxSim, SEPIA, our own RTE-MT
  - TERpA is a representative good one that handles some word choice variation.
- MT research really requires some automatic metric to allow a rapid development and evaluation cycle.

A complete translation system

Decoding for IBM Models

- Of all conceivable English word strings, find the one maximizing \( P(e) \times P(f | e) \)
- Decoding is NP hard
  - (Knight, 1999)
- Several search strategies are available
  - Usually a beam search where we keep multiple stacks for candidates covering the same number of source words
- Each potential English output is called a hypothesis.

Search for Best Translation

voulez – vous vous taire !
Search for Best Translation

voulez – vous vous tairé !

you – you you quiet !

Search for Best Translation

voulez – vous vous tairé !

quiet you – you you !

Search for Best Translation

voulez – vous vous tairé !

you shut up !

Dynamic Programming Beam Search

Each partial translation hypothesis contains:
- Last English word chosen + source words covered by it
- Next-to-last English word chosen
- Entire coverage vector (so far) of source sentence
- Language model and translation model scores (so far)

The “Fundamental Equation of Machine Translation” (Brown et al. 1993)

\[
\hat{e} = \arg\max_{e} \frac{P(e \mid f)}{P(f)}
\]

\[
= \arg\max_{e} P(e) \times \frac{P(f \mid e)}{P(f)}
\]

\[
= \arg\max_{e} P(e) \times P(f \mid e)
\]
What StatMT people do in the privacy of their own homes

\[
\arg\max_e P(e | f) = e
\]

\[
\arg\max_e P(e) \times P(f | e) / P(f) = e
\]

\[
\arg\max_e P(e)^{1.9} \times P(f | e) \quad \text{… works better!}
\]

\[
\arg\max_e P(e)^{1.9} \times P(f | e) \times 1.1^{\text{length}(e)}
\]

Which model are you now paying more attention to?

\[
P(e) \times P(f | e) / P(f) \quad \text{Rewards longer hypotheses, since these are ‘unfairly’ punished by } P(e)
\]

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?

Which model are you now paying more attention to?
Advantages of Phrase-Based

- Many-to-many mappings can handle non-compositional phrases
- Local context is very useful for disambiguating
  - “interest rate” ➔ …
  - “interest in” ➔ …
- The more data, the longer the learned phrases
  - Sometimes whole sentences

How to Learn the Phrase Translation Table?

- Main method: “alignment templates” (Och et al, 1999)
- Start with word alignment, build phrases from that.

How to Learn the Phrase Translation Table?

- One method: “alignment templates” (Och et al, 1999)
- Start with word alignment, build phrases from that.

IBM Models are 1-to-Many

- Run IBM-style aligner both directions, then merge:

Phrase Pair Probabilities

- A certain phrase pair (f-f-f, e-e-e) may appear many times across the bilingual corpus.
- No EM training
- Just relative frequency:

\[ P(f-f-f \mid e-e-e) = \frac{\text{count}(f-f-f, e-e-e)}{\text{count}(e-e-e)} \]
Phrase-Based Translation

Scoring: Try to use phrase pairs that have been frequently observed.
Try to output a sentence with frequent English word sequences.

Syntax and Semantics in Statistical MT
Why Syntax?

- Need much more grammatical output
- Need accurate control over re-ordering
- Need accurate insertion of function words
- Word translations need to depend on grammatically-related words

Yamada and Knight (2001): The need for phrasal syntax

- He adores listening to music.

Parse Tree(E) → Sentence (J)

Parse Tree (English) => Translation model => Sentence (Japanese)

- Preprocess English by a parser
- Probabilistic Operations on a parse-tree
  1. Reorder child nodes
  2. Insert extra nodes
  3. Translate leaf words

Syntax-based Model

Experiment

- Training Corpus: J-E 2K sentence pairs
- J: Tokenized by Chasen [Matsumoto et al., 1999]
- E: Parsed by Collins Parser [Collins, 1999]
  - Trained: 40K Treebank, Accuracy: ~90%
- E: Flatten parse tree
  - To capture word-order difference (SVO->SOV)
- EM Training: 20 Iterations
  - 50 min/iter (Sparc 200Mhz 1-CPU) or
  - 30 sec/iter (Pentium3 700Mhz 30-CPU)

Result: Alignments

<table>
<thead>
<tr>
<th></th>
<th>Ave. Score</th>
<th># perf sent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y/K Model</td>
<td>0.582</td>
<td>10</td>
</tr>
<tr>
<td>IBM Model</td>
<td>0.431</td>
<td>0</td>
</tr>
</tbody>
</table>

- Ave. by 3 humans for 50 sents
- okay(1.0), not sure(0.5), wrong(0.0)
- precision only
MT Applications

Christopher Manning
CS 224N
2009

MT Applications: 1. Traditional

- Traditional scenario:
  - Documents had to be translated for your company/organization. Document production for organization
  - Generally, the quality/accuracy demands are high
  - High cost
    - Though most of it is now done as outsourced piecework
- MT tends to be ineffective: The cost of post-translation error correction is too high
- Main technology in the game: translation memory/translation workbench/terminology management
  - E.g., TRADOS
  - Very slowly, MT technology is starting to be incorporated, but most of the action is in terminology lexicon management
Bad TRADOS Screenshot...

Trados is relatively pricey (high hundreds for PC versions, thousands for server version); seen as necessary productivity tool (Photoshop for translators).

MT Applications: 2. Web

- Web applications:
  - Dominant scenario: User-initiated translation
    - Crucial difference: The quality doesn’t have to be great. The user is usually okay if they can understand the gist of what is going on
  - Second scenario
    - Somehow on the web people will accept medium quality results. Accessible information is better than no information
  - MT is saved!!! “It’s the web, stupid.”
    - (But is there money in it?)

AltaVista
BabelFish
1997:
Free, automatic translation for the masses.
Revolutionary.

But, what was the underlying technology?
SYSTRAN.

MacOS Dashboard?
SYSTRAN
Google until 2006?
SYSTRAN
Machine Translation Summary

- **Usable Technologies**
  - "Translation memories" to aid translator
  - Low quality screening/web translators

- **Technologies**
  - Traditional: Systran (Altavista Babelfish, what you got till mid-2006 on Google) is now seen as a limited success
  - Statistical MT over huge training sets is successful (ISI/LanguageWeaver, Microsoft, Google)

- **Key ideas of the present/future**
  - Statistical phrase based models
  - Syntax based models
  - Better language models (e.g., bigger, using grammar)
  - Better decoding models (e.g., by restricting model?)