L3: Statistical machine translation in a few slides

Mikel L. Forcada[1,2]

[1]Departament de Llenguatges i Sistemes Informàtics, Universitat d'Alacant, E-03071 Alacant (Spain)

[2]Prompsit Language Engineering, S.L., Edifici Quorum III, Av. Universitat s/n, E-03202 Elx, Spain

Crash Course on Machine Translation
IBL, Bulgarian Acad. of Sci.,
Sofia, 2–4 July 2014

Translation as probability/1

- Instead of saying that
  - a source-language (SL) sentence s in a SL text
  - and a target-language (TL) sentence t

  as found in a SL–TL *bitext* are or are not a translation of each other,
- in SMT one says that they are a translation of each other with a probability $p(s,t) = p(t,s)$ (a *joint* probability).
- We'll assume we have such a probability model available. Or at least a reasonable estimate.

The "canonical" model

- We can rewrite eq. (1) as

$$p(t|s) = \frac{p(s|t)p(t)}{p(s)} \tag{3}$$

- and then with (2) to get

$$t^\star = \arg\max_t p(s|t)p(t) \tag{4}$$

"Decoding"/2

- In SMT parlance, the process of finding t^* is called *decoding*.[1]
- Obviously, it does not explore all possible translations t in the *search space*. There are infinitely many.
- The *search space* is *pruned*.
- Therefore, one just gets a reasonable $t^{\simeq\star}$ instead of the ideal $t^\star$
- Pruning and search strategies are a very active research topic.

Free/open-source software: **Moses**.

[1] Reading SMT articles usually entails deciphering jargon which may be very obscure to outsiders or newcomers

Training/2

- The lexical model and the alignment model are estimated using a large sentence-aligned bilingual corpus through a complex iterative process.
- An initial set of lexical probabilities is obtained by assuming, for instance, that any word in the TL sentence aligns with any word in its SL counterpart. And then:
  - Alignment probabilities in accordance with the lexical probabilities are computed.
  - Lexical probabilities are obtained in accordance with the alignment probabilities

  This process ("expectation maximization") is repeated a fixed number of times or until some convergence is observed (free/open-source software: **Giza++**).

"Log-linear"/1

- More SMT jargon!
- It's short for *linear* combination of *log*arithms of probabilities.
- And, sometimes, even features that aren't logarithms or probabilities of any kind.
- OK, let's take a look at the maths.

"Log-linear"/3

- "Feature selection is a very open problem in SMT" (Lopez 2008)
- Other possible functions include length penalties (discouraging unreasonably short or long translations), "inverted" versions of $p(s|t)$, etc.
- Where do we get the λ_k's from?
- They are usually *tuned* so as to optimize the results on a *tuning set*, according to a certain objective function that
  - is taken to be an indicator that correlates with translation quality
  - may be automatically obtained from the output of the SMT system and the translation in the corpus.

  This is called MERT (*minimum error rate training*) sometimes (free/open-source software: the **Moses** suite).

The SMT lifecycle

Development:

- Training: monolingual and sentence-aligned bilingual corpora are used to estimate probability models (features)
- Tuning: a held-out portion of the sentence-aligned bilingual corpus is used to tune the coeficients λ_k

Decoding: sentences s are fed into the SMT system and "decoded" into their translations t.

Evaluation: the system is evaluated against a reference corpus.

L4: Machine translation evaluation

Mikel L. Forcada
(slides by Felipe Sánchez-Martínez)

Departament de Llenguatges i Sistemes Informàtics
Universitat d'Alacant
E-03071 Alacant, Spain

Crash Course on Machine Translation
IBL, Bulgarian Acad. of Sci.,
Sofia, 2–4 July 2014

Machine translation evaluation

The evaluation of the appropriateness of an MT system, is independent of the purpose of the translation?

- **Match difficult, supporters giving very support inconditional, players very motivation* ←Assimilation
- **Eat you were not coming we left*[a] ←Dissemination

[a] Spanish *como* may be *eat* or *as*

Machine translation evaluation

Assessment of the appropriateness of an MT system for dissemination

Expenses incurred:

- ***Operating* costs** (effective cost per word):
  - return of the investment to acquire/develop the MT system
  - technical service and maintenance
  - migration (customisation of software, acquisition of systems) (initial)
  - cost of the evaluation (initial)
- ***Pre-editing* and preparation costs**: the texts may need to be prepared and even pre-edited
- **Post-editing cost**: it depends on the *quality* of the raw translation and on the training of the post-editors
- **Training cost** (initial) (users need learn how to use a new technology)

Machine translation evaluation

Assessment of the appropriateness of an MT system for dissemination

A case study: Autodesk

Autodesk productivity test
`http://langtech.autodesk.com/productivity.html`

Machine translation evaluation

Manual evaluation of machine translation

- High cost
- Takes too much time

For the efficient development of machine translation systems, fast, cheap and regular evaluations need to be performed

Machine translation evaluation

Automatic evaluation of machine translation
Automatic evaluation measures of machine translation :

Are not well suited to . . .

. . . make a decision on the adoption of an MT system for dissemination

Are useful to ...

... compare different version of the same MT system or MT systems following the same approach to translation

Machine translation evaluation

Automatic evaluation of machine translation

- **Position -independent error rate (PER)**: Percentage of words in the reference translation that do not appear in the raw machine translation

$$\text{PER} = 1 - \frac{\#\text{ok} - \max(0, \text{length}(\text{trans}) - \text{length}(\text{ref}))}{\text{length}(\text{ref})}$$

MT: the Israeli officials responsability of airport safety
Ref.: Israeli officials are responsible for the airport security

$$\text{PER} = 1 - \frac{4 - \max(0,7-8)}{8} = 1 - \frac{4}{8} = 0{,}5$$

Machine translation evaluation

Automatic evaluation of machine translation

- **METEOR**: Introduces the use of dictionaries of synonyms and *stemmers* to avoid penalising those words that do not appear in the reference translation but are similar

MT: isra offici respons of airport safet
Ref.: isra offici are respons for airport {secur,safet}