

Problems for Statistical MT

- Preprocessing
- Language modeling
- Translation modeling
- Decoding
- Parameter optimization
- Evaluation

MT Evaluation

- Source only
- Manual:
 - SSER (subjective sentence error rate)
 - Correct/Incorrect
 - Error categorization
- Objective usage testing



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

Reference Evaluation

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/chemical attack against public places such as **the airport**.

Machine translation:
 The American [?] international **airport and its** the office at 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 Evaluation

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/chemical attack against public places such as **the airport**.

Machine translation:
 The American [?] international **airport and its** the office at 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.

n-gram precision?

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/chemical attack against public places such as the airport.

Machine translation:

The American (?) international airport and its the office at receives one calls self the sand Arab ric, 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.

- N-gram precision (score is between 0 & 1)
 - What percentage of machine n-grams can be found in the reference translation?
 - An n-gram is a sequence of n words
 - Not allowed to use same portion of reference translation twice (can't cheat by typing out "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 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/chemical attack against public places such as the airport.

Machine translation:

The American (?) international airport and its the office at receives one calls self the sand Arab ric, 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.

- BLEU formula
 - Generally N=4
 - $w_i=1/N$ (uniform weights)

$$BLEU = \prod_{i=1}^N \sqrt[p_i]{w_i} BP$$

BP=brevity penalty
 p_i =i-gram precision

$$BLEU = BP \cdot \exp\left(\sum_{i=1}^N w_i \log p_i\right)$$

Multiple Reference Translations

Reference translation 1:

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

Reference translation 2:

Guam (International Airport and its) offices are maintaining a high state of alert after receiving an e-mail that was from a person claiming to be the Saudi Arabian businessman Bin Laden and had threatened to launch a biological and chemical attack on the airport and other public places.

Machine translation:

The American (?) international airport and its the office at receives one calls self the sand Arab ric, 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 3:

The US International Airport of Guam and its office has received an e-mail from a self-called man whose name named Laden, which threatens to launch a biochemical attack on such public places as airport. Guam authority has been alert.

Reference translation 4:

US Guam (International Airport and its) offices received an e-mail from Mr. Bin Laden the other businessman from Saudi Arabia. They said there would be a biochemical air raid to Guam Airport and other public places. Guam needs to be in high precaution about this matter.