Multi-Objective Optimization Problems in Statistical Machine Translation

Kevin Duh (Nara Institute of Science and Technology, Japan)
Baskaran Sankaran, Anoop Sarkar (Simon Fraser U., Canada)

Introduction

There are 6000 languages in the world.

⇩

MACHINE TRANSLATION

Our interest:
Multi-objective Optimization for building these software systems

⇩

Hay 6.000 lenguas en el mundo.

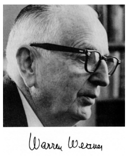
Statistical Machine Translation

A bit of History

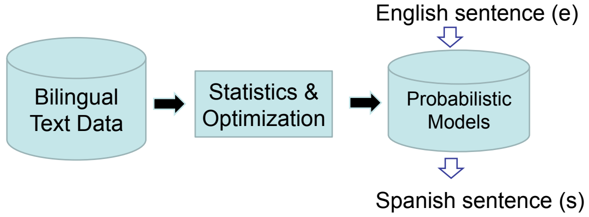
1960s-now:
Rule-Based Machine Translation
e.g. SYSTRAN

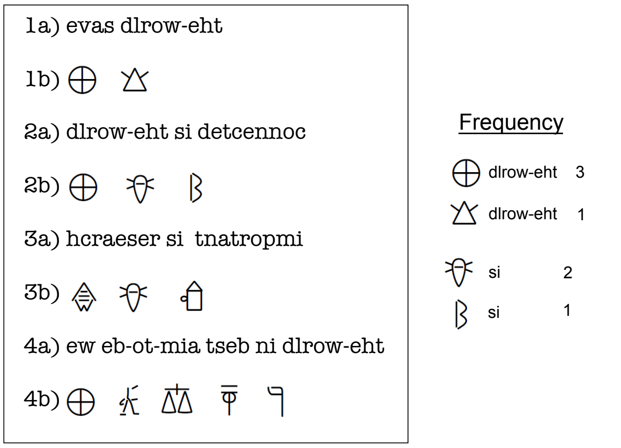
2000s-now:
Statistical Machine Translation
e.g. Google Translate, Bing

When I look at an article in Russian, I say: "This is really written in English, but has been coded in some strange symbols. I will now proceed to decode. [1947]


Warren Weaver

Architecture of Statistical Machine Translation


English sentence (e)
Bilingual Text Data → Statistics & Optimization → Probabilistic Models
Spanish sentence (s)


1a) evas dlrow-eht
1b)
2a) dlrow-eht si detcennoc
2b)
3a) hcraeser si tnatropmi
3b)
4a) ew eb-ot-mia tseb ni dlrow-eht
4b)

Frequency

⊕	dlrow-eht	3
	dlrow-eht	1
	si	2
	si	1

Where is the Optimization Problem?

Optimize weights w_k so that s_{pred} is similar to s_{true}
Usually, non-convex piecewise linear objective

$$s_{predict} = \operatorname*{argmax}_{\substack{s\in spanish\\ sentences}} prob(s\,|\,e) = \operatorname*{argmax}_{\substack{s\in spanish\\ sentences}} \sum_{k=1}^{K} w_k f_k(s,e)$$

$$\text{MAXIMIZE}\ \ similarity(s_{pred}, s_{true})$$

Please give us advice!

Better Techniques?

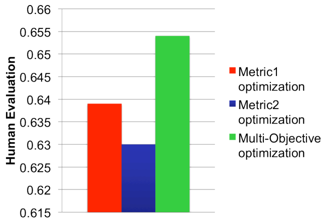
- ________________________________
- ________________________________
- ________________________________

Better Evaluation?

- How to visualize/compare methods with 3+ objectives?
- What to conclude when Pareto Frontiers of diff. methods cross?

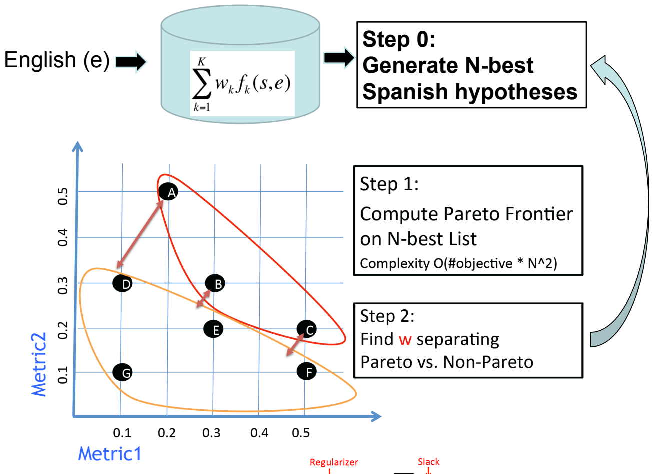
Motivation for Multiple Objectives

1. Ideally, humans determine similarity(s_{pred}, s_{true})
2. But humans cost $$$
3. So we resort to automatic similarity metrics on strings
4. Each metric has pros/cons, so we hope to optimize all



Multi-Objective Optimization Techniques

1. **Lateen Technique:**
   Alternate among single-objective problems.
2. **Linear combination:**
   Combination weights are set to correlate w/ human scores
3. **Pareto Support Vector Machine:**


English (e) → $\sum_{k=1}^{K} w_k f_k(s,e)$ →

Step 0: Generate N-best Spanish hypotheses

Step 1: Compute Pareto Frontier on N-best List
Complexity O(#objective * N^2)

Step 2: Find w separating Pareto vs. Non-Pareto

$$\min_{w} \| \overset{\text{Regularizer}}{w} \|^2 + c \sum_{ij} \overset{\text{Slack}}{\xi_{ij}}$$

$$\text{s.t. } w^T \overset{\text{Feature vector}}{\Phi}(\underset{\text{Input sentence}}{x}, \underset{\text{Good hypothesis}}{y_i}) - w^T \Phi(x, \underset{\text{Poor hypothesis}}{y_j}) \geq 1 - \zeta_{ij}$$

$$\forall y_i \in ParetoFront, y_j \notin ParetoFront$$

i.e. score of pareto hypothesis should be higher than non-pareto hypotheses

Intuition:
Translations on Pareto Front deserve higher probabilities

References

- K. Duh+, Learning to Translation with Multiple Objectives, Proc. of Association for Computational Linguistics (ACL2012)
- B. Sankaran+, Multi-metric Optimization using Ensemble Tuning, Proc. of North American Assoc. for Computational Linguistics (NAACL 2013)