Cunei Machine Translation Platform: System Description

Aaron B. Phillips and Ralf Brown

Language Technologies Institute
Carnegie-Mellon University

November 12th, 2009
3rd Workshop on Example-Based Machine Translation

The Motivation

All examples of translations are equal...

The Motivation

All examples of translations are equal...

but some are *more equal* than others

Outline

Example-Based Machine Translation

- Queries the corpus to extract translations that have a high degree of similarity with the input
- The simplest representation or model of the translation process *is itself* the training data

A Bottom-Up Approach

Modeling

Each example of a translation is scored individually allowing for many avenues of exploration and improvement

- Morphological generalization
- Structural matching and substitution
- Context-sensitivity at the sentence and document level

Modeling

Each example of a translation is scored individually allowing for many avenues of exploration and improvement

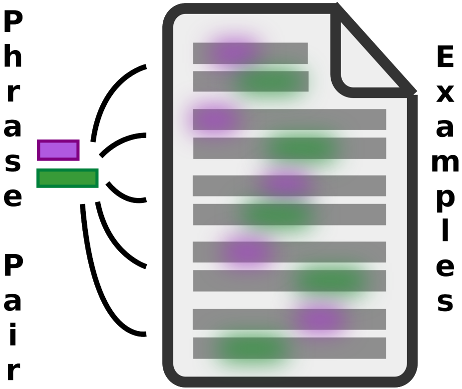
- Morphological generalization
- Structural matching and substitution
- Context-sensitivity at the sentence and document level

But modeling is often heuristic and optimization difficult

Statistical Machine Translation

- Model a limited quantity of information that can adequately represent the translation process
- Promote consistent models that can easily be optimized

A Top-Down Approach



Modeling: Log-Linear Model

Log-linear models are the defacto modeling approach in SMT

- Easily extended with more features
- The system builder does not need to understand how all the features interact
- Straight-forward to optimize

Blurring the Line: EBMT-like Lookups

SMT systems that store the entire corpus for runtime-lookup

- [Vogel, 2005] uses a run-time "Alignment as Sentence Splitting" algorithm for phrases that were not pre-computed and stored in the phrase table
- [Callison-Burch et al., 2005] simplifies the translation model to only include $\psi(f|e)$ and $lex(f|e)$ which are sampled over the corpus
- [Lopez, 2008] calculates a Heiro-style grammar and weights on-the-fly.

Blurring the Line: EBMT-like Lookups

SMT systems that store the entire corpus for runtime-lookup

- [Vogel, 2005] uses a run-time "Alignment as Sentence Splitting" algorithm for phrases that were not pre-computed and stored in the phrase table
- [Callison-Burch et al., 2005] simplifies the translation model to only include $\psi(f|e)$ and $lex(f|e)$ which are sampled over the corpus
- [Lopez, 2008] calculates a Heiro-style grammar and weights on-the-fly.

These systems address the inefficiencies of using very large phrase tables, but do not fundamentally change the modeling approach.

Blurring the Line: EBMT-like Features

SMT systems that incorporate context features

- [Carpuat and Wu, 2007] incorporates scores from a WSD system (trained separately) that generates a new phrase table for each sentence
- [Gimpel and Smith, 2008] adds a new entry in the phrase table for every context (along with a series of context-dependent features)

Blurring the Line: EBMT-like Features

SMT systems that incorporate context features

- [Carpuat and Wu, 2007] incorporates scores from a WSD system (trained separately) that generates a new phrase table for each sentence
- [Gimpel and Smith, 2008] adds a new entry in the phrase table for every context (along with a series of context-dependent features)

Still uses a top-down perspective to calculate features over all examples treated equally. Each phrase-pair includes more dependencies, further fragmenting the search space and making MLE estimates unreliable.

In this case, both approaches are also hampered by the inability to dynamically lookup examples and generate very large phrase tables.

Outline

Data-Driven Machine Translation
- EBMT
- SMT

Cunei: A Hybrid Approach

Experiments

Conclusions

Approach

Stolen from EBMT

- Calculate the features for each example individually

Stolen from SMT

- Collect this information into a single log-linear model that is straight-forward to optimize

Approach

Stolen from EBMT

- Calculate the features for each example individually

Stolen from SMT

- Collect this information into a single log-linear model that is straight-forward to optimize

The glue that holds it together

- Model collections of translation examples

What is a collection?

Every possible subset of examples form a collection

- An SMT phrase-pair is one such collection that contains all the examples
- At the other extreme, every example defines its own collection

What is a collection?

Each collection is defined by a series of feature-values forming constraint on the space of all possible collections

- Contains all examples that satisfy the constraint (whose feature-values are greater than or equal to it)
- The constraint can be used to model the collection
- Able to model joint dependencies among features
- Able to incorporate top-down *or* bottom-up features

Formalism

Given translation t, feature f, weights w, example e, and collection c:

SMT

$$score(t_i) = \prod_k f_{i,k}^{w_k} \tag{1}$$

EBMT

$$score(t_i) = \sum_{\forall e_m \in t_i} \prod_k f_{m,k}^{w_k} \tag{2}$$

Cunei's Approach

$$score(t_i) = \max_{\forall c_j \subseteq t_i} |c_j|^{w_{matches}} \prod_k \min_{\forall e_m \in c_j} f_{m,k}^{w_k} \tag{3}$$

How is it modeled?

Log-Linear Model

- Count of matching examples (m)
- Constraints ($c_{1\ldots n}$)

m	c_1	c_2	c_3	$\ldots$	c_n

How is it modeled?

Log-Linear Model

- Count of matching examples (m)
- Count of phrases in corpus (C)
- Lexical weighting (lex)
- Phrase penalty (e)
- Alignment scores ($a_{1...n}$)

m	C_f	$lex(f\|e)$	C_e	$lex(e\|f)$	e	a_1	a_2	...	a_n

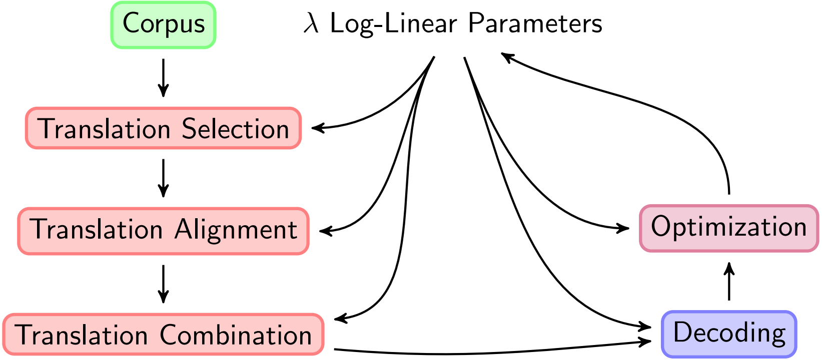
Optimization

Using a log-linear model, but the search space is much larger

How each translation is modeled changes based on the weights, but translations modeled by different constraints are still valid (just not maximal)

Could optimize using Moses' MERT, but built-in support for [Smith and Eisner, 2006]'s annealing method

System Diagram



Advantages

- Easy to model non-local features dependent on the particular input or surrounding translations
- Weights inform phrase-extraction producing a more consistent translation model
- Compactly and efficiently search a larger space of possible translations

Outline

Language Pairs

- Finnish-to-English
- French-to-English
- German-to-English

Training Data

Bilingual Europarl corpora

- Applied light pre-processing, filtering, and tokenization suitable for Western languages
- Word-aligned with GIZA++ in both directions

243 million word English language model

- All Europarl proceedings in English
- Portion of the English newswire released by the 2009 WMT

Showdown: Cunei vs. Moses

	Finnish-to-English	
	Dev	Test
Moses	0.2445	0.2361
Cunei	**0.2456**	**0.2369**

	French-to-English	
	Dev	Test
Moses	0.3207	0.3219
Cunei	**0.3215**	**0.3225**

	German-to-English	
	Dev	Test
Moses	0.2746	0.2546
Cunei	**0.2813**	**0.2634**

Finnish-to-English Output

Moses	mr president, indeed himaren orgy of violence and electoral fraud in local elections in the province, who were living in the region kreikkalaisvhemmistn.
Cunei	mr president, indeed in the province of violence and electoral fraud in local elections, which were against the greek minority in the area.
Reference	madam president, it is quite right that the municipal elections in himara were marked by violence and fraud at the expense of the greek minority living there.

German-to-English Output

Moses	i would like to criticise the lack of initiatives on the new challenges in the safety of employee participation and in industrial relations.
Cunei	i share the criticism of initiatives to the new challenges in the field of health and safety, and worker participation in the labour relations.
Reference	i would like to raise a criticism in connection with the lack of initiatives produced in response to the new challenges we face in health and safety at work, employee participation and labour relations.

Outline

Challenges

The search space is *much* larger

Challenges

The search space is *much* larger

...but this is also what allows us to select better translations

Cunei Machine Translation Platform

- Bridges the gap between EBMT and SMT by statistically modeling each example
- Competitive with state-of-the-art SMT systems like Moses
- Open-source with permissive license

Cunei Machine Translation Platform

- Bridges the gap between EBMT and SMT by statistically modeling each example
- Competitive with state-of-the-art SMT systems like Moses
- Open-source with permissive license

Try it out!

`http://www.cunei.org`

End

Europarl Statistics

	Finnish-to-English	
Types	499,770	84,257
Tokens	21,492,772	29,744,581
Sentences	1,121,312	

	French-to-English	
Types	106,862	87,083
Tokens	34,979,287	32,001,553
Sentences	1,207,184	

	German-to-English	
Types	273,960	86,671
Tokens	29,730,317	31,156,576
Sentences	1,165,545	

French-to-English Output

Moses	for some reason, i know that my name is not on the attendance register.
Cunei	for some reason i do not know, my name is not on the attendance register.
Reference	for some strange reason, my name is missing from the register of attendance.

Callison-Burch, C., Bannard, C., and Schroeder, J. (2005).
Scaling phrase-based statistical machine translation to larger corpora and longer phrases.
In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 255–262, Ann Arbor, USA.

Carpuat, M. and Wu, D. (2007).
Context-dependent phrasal translation lexicons for statistical machine translation.
In Maegaard, B., editor, *Machine Translation Summit XI Proceedings*, pages 73–80, Copenhagen, Denmark.

Gimpel, K. and Smith, N. A. (2008).
Rich source-side context for statistical machine translation.
In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 9–17, Columbus, USA.

Lopez, A. (2008).
Tera-scale translation models via pattern matching.
In *Proceedings of the 22nd International Conference on Computational Linguistics*, pages 505–512, Manchester, United Kingdom.

Smith, D. A. and Eisner, J. (2006).
Minimum risk annealing for training log-linear models.
In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 787–794, Sydney, Australia.

Vogel, S. (2005).
Pesa: Phrase pair extraction as sentence splitting.
In *Machine Translation Summit X Proceedings*, pages 251–258, Phuket, Thailand.