Outline

1. Multi-Engine Machine Translation
   - Alignment
   - Search Space
   - Features
     - Match

2. Model Combination

3. Other Combination Approaches
Multi-Engine Machine Translation
Model Combination
Other Combination Approaches

Alignment
Search Space
Features

Individual Systems

Translate

Align

Decode

Output

This Work: MEMT

METEOR

Translate

Translate

Translate

Input

Kenneth Heafield
System Combination
Arabic-English Example Combination

**System 1:** So even if that was meaningful, it is because you were late

**System 2:** Even if feasible, it is because you have been delayed

**Combined:** Even if feasible, it is because you were late

**Compare:** Compare

**Reference:** And even if that was useful, it was because you were late
Sentence Pair Alignment

**Match** surface, stems, WordNet synsets, and automatic paraphrases

**Minimize** crossing alignments

Twice that produced by nuclear plants

Double that that produce nuclear power stations

Overall Alignment: Urdu-English Example

1 Russian President Putin Mir ولادی it for a big success.

2 The Russian president ولادی the result of a big victory for Putin.
Overall Alignment: Urdu-English Example

1 Russian President Putin 

2 The Russian president

3 For the result Russian President

1 ولادی it for a big success.

2 ولادی the result of a big victory for Putin.

3 ولادی Mir Putin is a great success.

2 ولادی the result of a big victory for Putin.

3 ولادی Mir Putin is a great success.
Search Space

Algorithm

Start at the beginning of each sentence
Branch by appending the first unused word from a system

Example

System 1: Now can know why.

System 2: Now we can now know why.

Partial Hypothesis

{Now
Now}
Search Space

Algorithm

Start at the beginning of each sentence
Branch by appending the first unused word from a system
Use the appended word and those aligned with it

Example

System 1: Now can know why.

System 2: Now we can now know why.

Partial Hypothesis

\[
\begin{cases}
\text{can} \\
\text{we}
\end{cases}
\]
Algorithm

**Start** at the beginning of each sentence

**Branch** by appending the **first unused word** from a system

**Use** the **appended word** and those aligned with it

**Loop** until all hypotheses reach end of sentence

Example

**System 1:** Now can know why.

**System 2:** Now we can now know why.

**Partial Hypothesis**

Now we \{\text{can, can}\}
Algorithm

Start at the beginning of each sentence
Branch by appending the first unused word from a system
Use the appended word and those aligned with it
Loop until all hypotheses reach end of sentence

Example

System 1: Now can know why.

System 2: Now we can now know why.

Partial Hypothesis

Now we can \{ know, now \}
Outline

1. Multi-Engine Machine Translation
   - Alignment
   - Search Space
   - Features
     - Match

2. Model Combination

3. Other Combination Approaches
### Features

#### Length

Length of hypothesis

#### Language Model

- **Model**: log probability from an ARPA language model
- **OOV**: count of words not found in the model

#### Match

Count of $n$-grams matching each system
Feature Rationale

**Length**

Length of hypothesis

Compensate for length’s impact on other features

**Language Model**

**Model:** log probability from an ARPA language model

**OOV:** count of words not found in the model

Fluent output with tuned OOV penalty

**Match**

Count of n-grams matching each system

Agreement with translation systems
Match Features

**System 1:** Supported Proposal of France

**System 2:** Support for the Proposal of France

**Hypothesis:** Support for Proposal of France

<table>
<thead>
<tr>
<th></th>
<th>Unigram</th>
<th>Bigram</th>
<th>Trigram</th>
<th>Quadgram</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>System 1</strong></td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>System 2</strong></td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
What’s in a match?

**Exact matches**
- Lexical choice
- Choosing between aligned alternatives

**Approximate matches**
- Vote to include/exclude text
- Word order

**Answer**
- Use both types of features
- Exact matches effectively get a tunable bonus
Multi-Engine Machine Translation
Model Combination
Other Combination Approaches

Input
Hypergraph
Hypergraph
Hypergraph
Individual Systems
Model Combination
Select
Output

Kenneth Heafield
System Combination
Model Combination is Hypothesis Selection

The Search Space

- Union of search spaces from each system
- Combined sentence must be in one system’s hypergraph

Formally

- Every system outputs a hypergraph
- Phrasal lattice is just a special-case hypergraph
- Add a root node and an edge to each system root
Model Combination is Hypothesis Selection

The Search Space
- Union of search spaces from each system
- Combined sentence must be in one system’s hypergraph

Formally
- Every system outputs a hypergraph
- Phrasal lattice is just a special-case hypergraph
- Add a root node and an edge to each system root

Source Alignment
Hypergraphs retain alignment to source
Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>Length of hypothesis</td>
</tr>
<tr>
<td>Model score</td>
<td>Score given by the underlying system</td>
</tr>
<tr>
<td>System indicator</td>
<td>Each system has a feature:</td>
</tr>
<tr>
<td></td>
<td>1 if derived from that system</td>
</tr>
<tr>
<td></td>
<td>0 otherwise</td>
</tr>
<tr>
<td>N-gram support</td>
<td>Support from each system for n-grams</td>
</tr>
</tbody>
</table>
N-gram support

Posterior of $n$-gram

What fraction of system $i$’s translations include “crack rocks”?

Formally

$$v^n_i(g) = \mathbb{E}_{P_i(d|f)} h(d, g)$$

$v^n_i(g)$ System $i$’s vote for $n$-gram $g$

$P_i(d|f)$ Probability of a derivation $d$ in hypergraph $f$ from system $i$

$h(d, g)$ 1 if the derivation $d$ contains $n$-gram $g$; 0 otherwise
## Performance

<table>
<thead>
<tr>
<th></th>
<th>ar-en</th>
<th>zh-en</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best individual</td>
<td>43.9</td>
<td>28.4</td>
</tr>
<tr>
<td>Combined</td>
<td>45.3</td>
<td>29.0</td>
</tr>
</tbody>
</table>

**Table:** Performance (BLEU) on NIST 2008 task using three systems.
Serial System Combination

Input → Translate (SYSTRAN) → Post-edit → Output
### Input Comparison

### Input to System Combination

<table>
<thead>
<tr>
<th>Method</th>
<th>Output Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBR</td>
<td>N-best list</td>
</tr>
<tr>
<td>Hyposel</td>
<td>N-best list</td>
</tr>
<tr>
<td>Confusion Networks</td>
<td>N-best list</td>
</tr>
<tr>
<td>MEMT</td>
<td>1-best</td>
</tr>
<tr>
<td>Model Combination</td>
<td>Hypergraph</td>
</tr>
<tr>
<td>Serial System Combination</td>
<td>Single output</td>
</tr>
</tbody>
</table>
Results Into English

Czech-English

- memt: 1.3
- upv: 0.4
- rwth: 0.6
- bbn: 1.6

German-English

- memt: 1.8
- upv: 0.8
- rwth: 1.6
- bbn: 0.8
- jhu: 0.9
- hypo: -0.6

Spanish-English

- memt: 0.7
- upv: 0.1
- bbn: 1.0

French-English

- memt: 0.2
- upv: -0.2
- rwth: 0.4
- bbn: 0.4
- jhu: 0.9
- dcu: -0.3
- hypo: lium: -0.4

System Combination
Results From English

English-Czech

- memt: 0.4
- upv: 0.9
- rwth: 0.9
- koc: 0.0
- dcu: 2.2

English-Spanish

- memt: 1.4
- upv: 0.4
- rwth: 0.7
- koc: 0.0

English-German

- memt: 0.9
- upv: 0.4
- rwth: 0.4
- koc: 0.3

English-French

- memt: 1.2
- upv: 1.0
- rwth: 1.0
- koc: 0.8