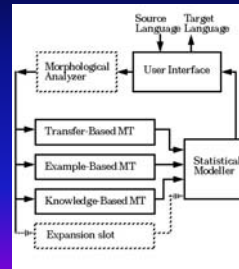


## MEMT: Multi-Engine Machine Translation

11-731  
Machine Translation  
Alon Lavie  
February 19, 2007

## Multi-Engine MT



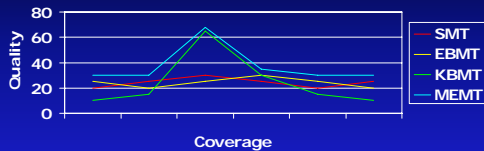
- Apply several MT engines to each input in parallel
- Create a combined translation from the individual translations
- Goal is to combine strengths, and avoid weaknesses.
- Along all dimensions: domain limits, quality, development time/cost, run-time speed, etc.
- Various approaches to the problem

Feb 19, 2007

11-731: MEMT

2

## Multi-Engine MT



Feb 19, 2007

11-731: MEMT

3

## MEMT Goals and Challenges

- **Scientific Challenges:**
  - How to combine the output of multiple MT engines into a selected output that outperforms the originals in translation quality?
  - **Synthetic combination** of the output from the original systems, or **just selecting the best output** (on a sentence-by-sentence basis)?
- **Engineering Challenge:**
  - How to integrate multiple distributed translation engines and the MEMT combination engine in a common framework that supports ongoing development and evaluation

Feb 19, 2007

11-731: MEMT

4

## MEMT Approaches

- Earliest work on MEMT in early 1990s (PANGLOSS) [pre "ROVER"]
- Several Main Approaches:
  - Classification approaches
  - Lattice Combination and "joint" decoding
  - Confusion Networks
  - Alignment-based Synthetic MEMT

Feb 19, 2007

11-731: MEMT

5

## Classification-based Approaches

- **Main Idea:** construct a classifier that given several translations for the same input sentence selects the "best" translation (on a sentence-by-sentence basis)
- Main knowledge source for scoring the individual translations is a standard statistical target-language LM, plus confidence scores for each engine
- Example: [Tidhar & Kuessner, 2000]

Feb 19, 2007

11-731: MEMT

6

## Lattice-based MEMT

- Earliest approach, first tried in CMU's PANGLOSS in 1994, and still active in recent work
- **Main Ideas:**
  - Multiple MT engines each produce a lattice of scored translation fragments, indexed based on source language input
  - Lattices from all engines are combined into a global comprehensive lattice
  - Joint Decoder finds best translation (or n-best list) from the entries in the lattice

Feb 19, 2007

11-731: MEMT

7

## Lattice-based MEMT: Example

<i>El punto de descarga</i> The drop-off point	<i>se cumplirá en</i> will comply with	<i>el puente Agua Fria</i> The cold Bridgewater
<i>El punto de descarga</i> The discharge point	<i>se cumplirá en</i> will self comply in	<i>el puente Agua Fria</i> the "Agua Fria" bridge
<i>El punto de descarga</i> Unload of the point	<i>se cumplirá en</i> will take place at	<i>el puente Agua Fria</i> the cold water of bridge

## Lattice-based MEMT

- **Main Drawbacks:**
  - Requires MT engines to provide lattice output → difficult to obtain!
  - Lattice output from all engines must be **compatible**: common indexing based on source word positions → difficult to standardize!
  - Common TM used for scoring edges may not work well for all engines
  - Decoding does not take into account any **reinforcements** from multiple engines proposing the same translation for any portion of the input

Feb 19, 2007

11-731: MEMT

9

## Consensus Network Approach

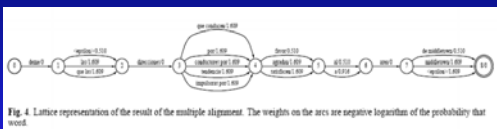
- **Main Ideas:**
  - Collapse the collection of linear strings of translations into a minimal consensus network ("sausage" graph) that represents a finite-state automaton
  - Edges that are supported by multiple engines receive a score that is the sum of their contributing confidence scores
  - Decode: find the path through the consensus network that has optimal score
  - Example: [Bangalore et al, 2001]

Feb 19, 2007

11-731: MEMT

10

## Consensus Network Example



Feb 19, 2007

11-731: MEMT

11

## Alignment-based Synthetic MEMT

Two Stage Approach:

1. Identify common words and phrases across the translations provided by the engines
2. Decode: search the space of synthetic combinations of words/phrases and select the highest scoring combined translation

Example:

1. announced afghan authorities on saturday reconstituted four intergovernmental committees
2. The Afghan authorities on Saturday the formation of the four committees of government

Feb 19, 2007

11-731: MEMT

12

## Alignment-based Synthetic MEMT

### Two Stage Approach:

1. Identify common words and phrases across the translations provided by the engines
2. Decode: search the space of synthetic combinations of words/phrases and select the highest scoring combined translation

### Example:

1. announced **afghan authorities on saturday** reconstituted **four intergovernmental committees**
2. The **Afghan authorities on Saturday** the formation of the **four committees of government**

MEMT: the afghan authorities announced on Saturday the formation of four intergovernmental committees

Feb 19, 2007

11-731: MEMT

13

## The Word Alignment Matcher

- Developed by Satanjeev Banerjee as a component in our METEOR Automatic MT Evaluation metric
- Finds maximal alignment match with minimal "crossing branches"
- Allows alignment of:
  - Identical words
  - Morphological variants of words
  - Synonymous words (based on WordNet synsets)
- Implementation: Clever search algorithm for best match using pruning of sub-optimal sub-solutions

Feb 19, 2007

11-731: MEMT

14

## Matcher Example

the sri lanka prime minister criticizes the leader of the country  
President of Sri Lanka criticized by the country's Prime Minister

Feb 19, 2007

11-731: MEMT

15

## The MEMT Decoder Algorithm

- Algorithm builds collections of partial hypotheses of increasing length
- Partial hypotheses are extended by selecting the "next available" word from one of the original systems
- Sentences are assumed mostly synchronous:
  - Each word is either *aligned* with another word or is an *alternative* of another word
- Extending a partial hypothesis with a word "pulls" and "uses" its aligned words with it, and marks its alternatives as "used"
- Partial hypotheses are scored and ranked
- Pruning and re-combination
- Hypothesis can end if any original system proposes an end of sentence as next word

Feb 19, 2007

11-731: MEMT

16

## Scoring MEMT Hypotheses

- Scoring:
  - Word confidence score [0,1] based on engine confidence and reinforcement from alignments of the words
  - LM score based on suffix-array 6-gram LM
  - Log-linear combination: weighted sum of logs of confidence score and LM score
  - Select best scoring hypothesis based on:
    - Total score (bias towards shorter hypotheses)
    - Average score per word

Feb 19, 2007

11-731: MEMT

17

## The MEMT Algorithm: Further Issues

- Parameters:
  - "lingering word" horizon: how long is a word allowed to linger when words following it have already been used?
  - "lookahead" horizon: how far ahead can we look for an alternative for a word that is not aligned?
  - "POS matching": limit search for an alternative to only words of the same POS
  - "Chunking": phrases in an engine can be marked as "chunks" that should not be broken apart

Feb 19, 2007

11-731: MEMT

18

## Example

**IBM:** korea stands ready to allow visits to verify that it does not manufacture nuclear weapons . 0.7407

**ISI:** North Korea Is Prepared to Allow Washington to Verify that It Does Not Make Nuclear Weapons 0.8007

**CMU:** North Korea prepared to allow Washington to the verification of that is to manufacture nuclear weapons 0.7668

**Selected MEMT Sentence :**  
north korea is prepared to allow washington to verify that it does not manufacture nuclear weapons . 0.8894 (-2.75135)

Feb 19, 2007

11-731: MEMT

19

## Example

**IBM:** victims russians are one man and his wife and abusing their eight year old daughter plus a ( 11 and 7 years ) man and his wife and driver , egyptian nationality . : 0.6327

**ISI:** The victims were Russian man and his wife, daughter of the most from the age of eight years in addition to the young girls ) 11 7 years ( and a man and his wife and the bus driver Egyptian nationality. : 0.7054

**CMU:** the victims Cruz man who wife and daughter both critical of the eight years old addition to two Orient ( 11 ) 7 years ) woman , wife of bus drivers Egyptian nationality . : 0.5293

**MEMT Sentence :**  
**Selected :** the victims were russian man and his wife and daughter of the eight years from the age of a 11 and 7 years in addition to man and his wife and bus drivers egyptian nationality . 0.7647 -3.25376

**Oracle :** the victims were russian man and wife and his daughter of the eight years old from the age of a 11 and 7 years in addition to the man and his wife and bus drivers egyptian nationality young girls . 0.7964 -3.44128

Feb 19, 2007

11-731: MEMT

20

## Example

**IBM:** the sri lankan prime minister criticizes head of the country's : 0.8862

**ISI:** The President of the Sri Lankan Prime Minister Criticized the President of the Country : 0.8660

**CMU:** Lankan Prime Minister criticizes her country: 0.6615

**MEMT Sentence :**  
**Selected:** the sri lankan prime minister criticizes president of the country . 0.9353 -3.27483

**Oracle:** the sri lankan prime minister criticizes president of the country's . 0.9767 -3.75805

Feb 19, 2007

11-731: MEMT

21

## System Development and Testing

- Initial development tests performed on TIDES 2003 Arabic-to-English MT data, using IBM, ISI and CMU SMT system output
- Preliminary evaluation tests performed on three Arabic-to-English systems and on three Chinese-to-English COTS systems
- Recent Deployments:
  - GALE Interoperability Operational Demo (IOD): combining output from IBM, LW and RWTH MT systems
  - Used in joint ARL/CMU submission to MT Eval-06: combining output from several ARL (mostly) rule-based systems

Feb 19, 2007

11-731: MEMT

22

## Internal Experimental Results: MT-Eval-03 Set Arabic-to-English

System	METEOR Score
System A	.4241
System B	.4231
System C	.4405
Choosing best online translation	.4432
MEMT	.5185
Best hypothesis generated by MEMT	.5883

Feb 19, 2007

11-731: MEMT

23

## ARL/CMU MEMT MT-Eval-06 Results Arabic-to-English

NIST Set:

	BLEU	METEOR	NIST	TER
Best Individual Engine	0.1355	0.4604	6.0239	73.422
Best MEMT	0.1417	0.5049	6.0889	79.458

GALE Set:

	BLEU	METEOR	NIST	TER
Best Individual Engine	0.0653	0.5923	4.0166	81.197
Best MEMT	0.0807	0.4400	4.1438	84.113

Feb 19, 2007

11-731: MEMT

24

## Architecture and Engineering

- **Challenge:** How do we construct an effective architecture for running MEMT within large-scale distributed projects?
  - Example: GALE Project
  - Multiple MT engines running at different locations
  - Input may be text or output of speech recognizers, Output may go downstream to other applications (IE, Summarization, TDT)
- **Approach:** Using IBM's UIMA: Unstructured Information Management Architecture
  - Provides support for building robust processing "workflows" with heterogeneous components
  - Components act as "annotators" at the character level within documents

Feb 19, 2007

11-731: MEMT

25

## UIMA-based MEMT

- MEMT engine set up as a remote server:
  - Communication over socket connections
  - Sentence-by-sentence translation
- Java "wrapper" turns the MEMT service into a UIMA-style annotator component
- UIMA supports easy integration of the MEMT component into various processing "workflows":
  - Input is a "document" annotated with multiple translations
  - Output is the same "document" with an additional MEMT annotation

Feb 19, 2007

11-731: MEMT

26

## Conclusions

- New sentence-level MEMT approach with nice properties and encouraging performance results:
  - 15% improvement in initial studies
  - 5-30% improvement in MT-Eval-06 setup
- Easy to run on both research and COTS systems
- UIMA-based architecture design for effective integration in large distributed systems/projects
  - Pilot study has been very positive
  - Can serve as a model for integration framework(s) under GALE and other projects

Feb 19, 2007

11-731: MEMT

27

## Open Research Issues

- Main Open Research Issues:
  - Improvements to the underlying algorithm:
    - Better word and phrase alignments
    - Larger search spaces
  - Confidence scores at the sentence or word/phrase level
  - Engines providing phrasal information
  - Decoding is still suboptimal
    - Oracle scores show there is much room for improvement
    - Need for additional discriminant features
    - Stronger (more discriminant) LMs

Feb 19, 2007

11-731: MEMT

28

## Demo

Feb 19, 2007

11-731: MEMT

29

## References

- 1994, Frederking, R. and S. Nirenburg. "Three Heads are Better than One". In Proceedings of the Fourth Conference on Applied Natural Language Processing (ANLP-94), Stuttgart, Germany.
- 2000, Tishar, Dan and U. Kessner. "Learning to Select a Good Translation". In Proceedings of the 17th International Conference on Computational Linguistics (COLING-2000), Saarbrücken, Germany.
- 2001, Bangalore, S., G. Bordini, and G. Riccardi. "Computing Consensus Translation from Multiple Machine Translation Systems". In Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop, Italy.
- 2005, [Ivanjurovic, S. and A. Ljeto. Multi-Engine Machine Translation Using Large Grammatical Models](#). In Companion Volume of Proceedings of the 43th Annual Meeting of the Association of Computational Linguistics (ACL-2005), Ann Arbor, Michigan, June 2005.
- 2005, [Ivanjurovic, S. and A. Ljeto. Multi-Engine Machine Translation Using Large Grammatical Models](#). In Proceedings of the 10th Annual Conference of the European Association for Machine Translation (EAMT-2005), Budapest, Hungary, May 2005.

Feb 19, 2007

11-731: MEMT

30