Lattice Based Translation Memory for Neural Machine Translation

Contents

- **Translation Memory guided Neural Machine Translation**
- **Two Baselines**
  - *Search Engine Guided Neural Machine Translation*
  - *Guiding Neural Machine Translation with Retrieved Translation Pieces*
- **Our Work**
  - Motivation
  - Model Structure
  - Current Experiment Results
  - Problems and Analysis

Translation Memory Guided Machine Translation

- Translation Memory
  - *Similar past translations especially in a narrow domain*
  - *Tech, Legal, User guide etc.*
- Imitate features like phrases, sentence pattern, etc.
- Put emphasis on specific expressions of TM
- External memory type?
  - *Lexicons, N-gram pieces, Phrase table, Sentence*
- How to define sentence similarity?
  - *Fuzzy Match Score*

$$s_{fuzzy}(X, X') = 1 - \frac{D_{edit}(X, X')}{max(|X|, |X'|)}$$

Source Sentence: Whereas in order to be effective, the rules laid down by this Directive **should** cover all **animals and products** that are subject, **in** intra-Community trade, to veterinary requirements;	**Source TM Sentence (~0.69):** Whereas in order to be effective, the rules laid down by this Directive **must** cover all **goods** that are subject **in the case of** intra-Community trade to veterinary requirements;
Reference Sentence: Considérant que , pour avoir un effet utile, les règles posées par la présente directive **devraient** couvrir l ' ensemble des **animaux et produits soumis** dans les échanges intracommunautaires à des exigences vétérinaires;	**TM Reference Sentence:** Considérant que , pour avoir un effet utile , les règles posées par la présente directive **doivent** couvrir l ' ensemble des **marchandises soumises** dans les échanges intracommunautaires à des exigences vétérinaires;

Search Engine Guided Neural Machine Translation

- **Key-value Memory**
  - *Constructed by retrieved sentence **pairs**.*
  - *TM: Each decoding step* $\{c'_t : (y'_t, h'_t)\}$
- **Retrieve Key-value Memory during translation**
  - *Query* c_t ; *Key* c'_τ ; *Value* h'_t

$$\mathrm{q}_{t,\tau} = \frac{exp(E(c_t, c_{\tau'}))}{\sum_{\tau'} exp(E(c_t, c'_{\tau'}))} \qquad \bar{h}_t = \sum_\tau q_{t,\tau} h'_\tau$$

  - *Deep fusion*

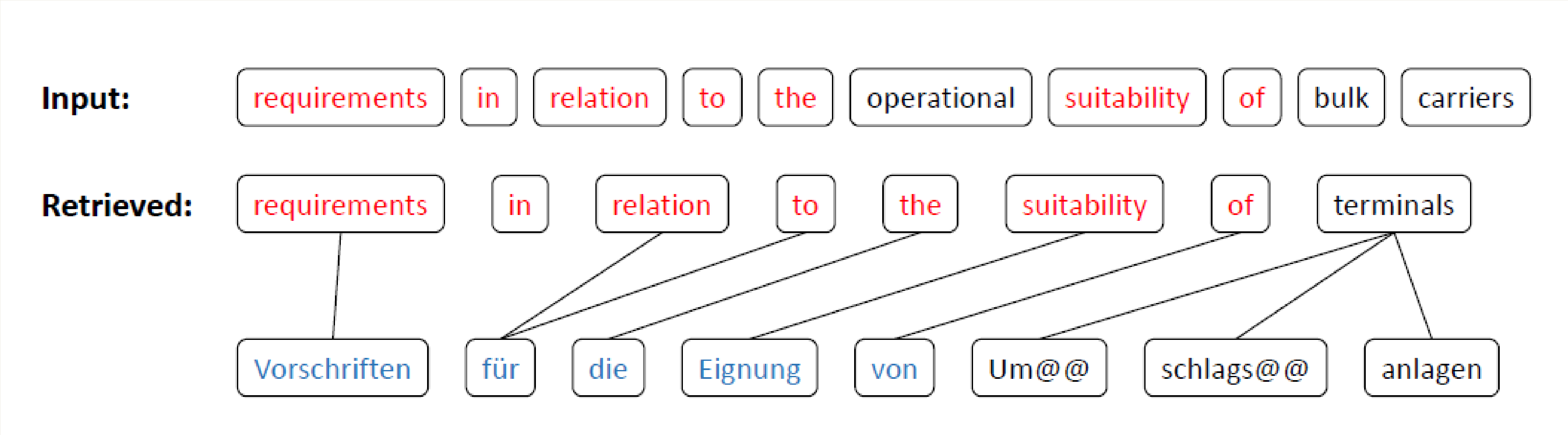
$$h_{fusion} = \zeta_t \cdot \bar{h}_t + (1 - \zeta_t) \cdot h_t$$

  - *Shallow fusion*

$$\mathrm{p}(y_t \mid y_{<t}, X, M) = \zeta_t q_{t',t} + (1 - \zeta_t) p(y_t \mid y_{<t}, X)$$

Guiding Neural Machine Translation with Retrieved Translation Pieces

- **Collecting retrieved translation pieces**
  - *Target-side N-gram pieces (max 4)*
  - *Last word's corresponding source word is an unedited word*
- **Matching score – max similarity**



Comparison of two baselines

- **Parameterized representation for TM**
  - Differentiable during training
  - Model specific
- **Time Consuming**
  - N TM sentences feed into NMT
- **Global Information**
  - Encode the whole sentence
- **Context matching**
  - Include both source and target info

- **Matching during beam search**
  - Only influence the translation process
  - Applicable to all models
- **Easy and Fast**
  - Increase 1/3 decoding time only
- **Local Information**
  - N-gram pieces with sentence-level similarity
- **Target matching**
  - Use alignment to influence prediction

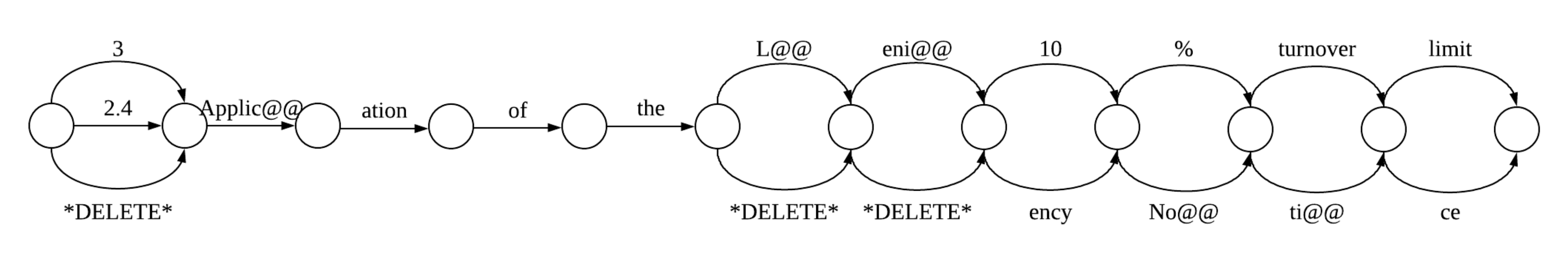
Motivation

- Deep Fusion
  - *Both baselines work well with output distribution manipulation.*
  - *Let the model automatically integrate useful representations of TM*
- Utilize Global Information while Keeping Relatively Efficient
  - *Pack multiple sentences into a lattice*
  - *Graph attention network*
- Work on the STOA – Transformer
  - *Both baselines are RNNSearch based models.*

Model – Pack Lattice

- Pack Three sentences into one
  - *3 Appli@@ ation of the L@@ eni@@ ency No@@ ti@@ ce*
  - *2.4 Applic@@ ation of the La@@ eni@@ ency No@@ ti@@ ce*
  - *Applic@@ ation of the 10 % turnover limit*

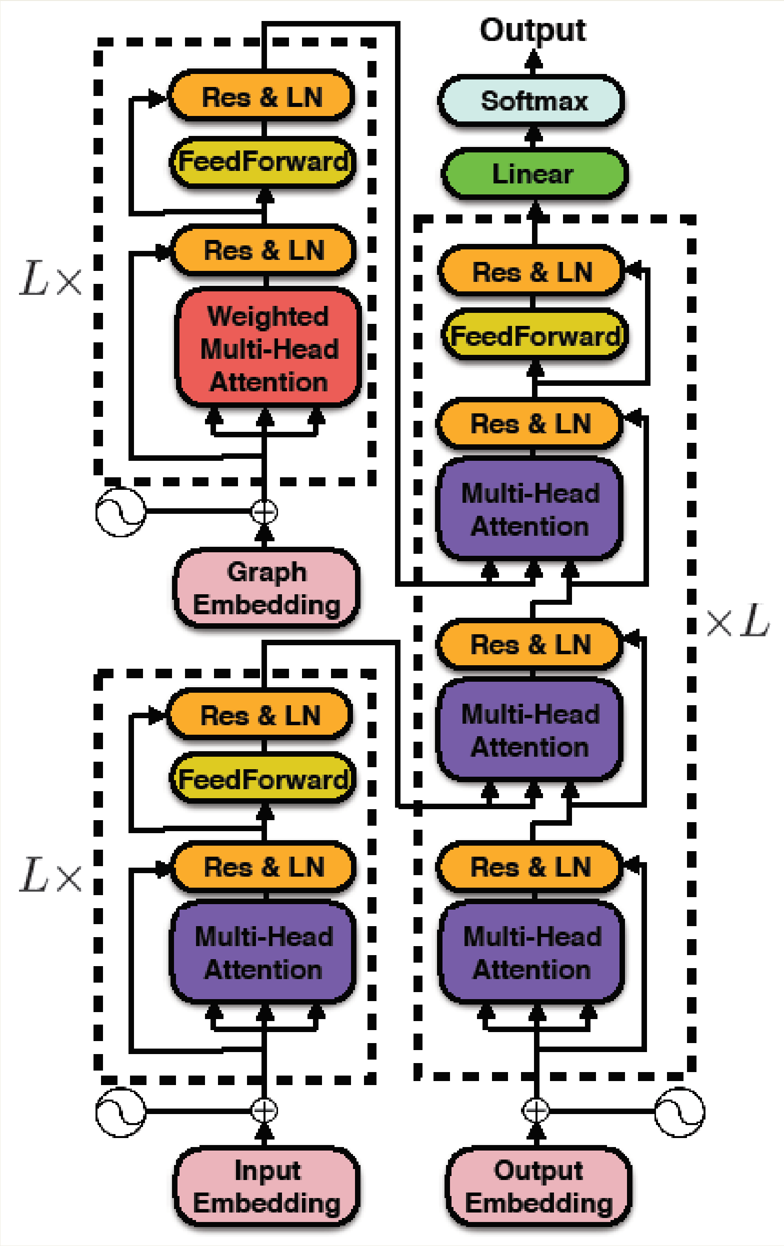
  - *Make classifications on each node*
  - *Nearly half redundant words are removed.*



Model – Lattice Attention

- Lattice Self Attention
  - *Idea from Graph Attention Network*
  - *Attend over neighbor nodes*
  - *Actually basic self attention with an adjacency matrix restriction*

- Lattice Attention
  - *Output of encoder-decoder attention over lattice self attention*
  - *Layer before output*



Credit to: Lemao Liu

Current Experiment Results

		EN-FR	FR-EN	EN-DE	DE-EN	EN-ES	ES-EN
Dev	RNNSearch	59.08	59.69	45.18	50.20	50.71	55.02
	SEG-RNNSearch	64.16	64.64	49.26	55.63	57.62	60.28
	Piece-RNNSearch	65.03	-	50.61	-	57.49	-
	Transfomer	66.33	65.95	51.86	58.54	59.50	61.97
	Piece Transformer	68.46	65.69	**52.78**	59.06	**60.23**	**62.94**
	Lattice Transformer	**70.91**	**67.62**	50.20	**61.52**	58.39	61.59
Test	RNNSearch	59.43	60.11	44.21	49.74	50.61	54.66
	SEG-RNNSearch	64.60	65.11	48.80	55.33	57.27	59.34
	Piece-RNNSearch	65.69	-	50.36	-	57.11	-
	Transfomer	66.36	67.34	51.19	58.86	59.29	61.79
	Piece Transformer	68.70	67.09	**52.47**	58.65	**59.79**	**62.23**
	Lattice Transformer	**70.82**	**68.96**	49.84	**61.42**	58.14	61.22

Source Sentence: Whereas in order to be effective, the rules laid down by this Directive should cover all animals and products that are subject, in intra-Community trade, to veterinary requirements;	**Source TM Sentence (~0.69):** Whereas in order to be effective, the rules laid down by this Directive must cover all goods that are subject in the case of intra-Community trade to veterinary requirements;
Reference Sentence: Considérant que , pour avoir un effet utile, les règles posées par la présente directive devraient couvrir l ' ensemble des animaux et produits soumis dans les échanges intracommunautaires à des exigences vétérinaires;	**TM Reference Sentence:** Considérant que , pour avoir un effet utile , les règles posées par la présente directive doivent couvrir l ' ensemble des marchandises soumises dans les échanges intracommunautaires à des exigences vétérinaires;
Transformer Translation (BLEU 0.42): Considérant que, pour être efficace, les règles prévues par la présente directive doivent couvrir tous les animaux et produits soumis , dans les échanges intracommunautaires, à des exigences vétérinaires;	**Lattice Transformer Translation (BLEU 0.85):** Considérant que, pour avoir un effet utile, les règles fixées par la présente directive devraient couvrir l ' ensemble des animaux et des produits soumis dans les échanges intracommunautaires à des exigences vétérinaires;

Problems and Analysis – Sentence Length & Similarity

Severe overfitting

Length	EN	FR	EN	DE	EN	ES
Train	29.44	33.35	34.43	33.44	32.10	34.95
Dev	29.42	33.17	45.32	48.98	45.33	49.60
Test	29.75	33.51	45.45	49.02	42.61	46.66

Similarity	EN-FR	FR-EN	EN-DE	DE-EN	EN-ES	ES-EN
Train	0.52	0.51	0.49	0.49	0.48	0.49
Dev	0.51	0.48	0.39	0.37	0.40	0.43
Test	0.53	0.48	0.35	0.35	0.39	0.45

After Reshuffle

Length	EN	DE
Train	34	34
Dev	38.67	38.08
Test	38.46	37.58

Similarity	DE-EN
Train	0.49
Dev	0.47
Test	0.48

Nearly all sets have 0.5 average similarity against top-5 TM sentences.

Problem and Analysis

- Deal with sentences that does not have very similar TM? (on average)
  - *Generate a more compact lattice with little noise*
    - Train with dynamically selected TM sentences
    - Remove irrelevant pieces from lattice using alignment
    - Also helps with efficiency
- Tiny batch constraints!
  - *Lattice Length ~= 3 Target Length*
  - *Huge GPU overhead*
  - *Transformer - 4096 tokens/batch*
  - *Lattice Transformer - 256-512 tokens/batch*
  - *Smaller batch results in lower training speed*
- Computational complexity!
  - *Fix lattice (with FFN layer & 256 tokens/batch) ~ 6 times (probably share all target layers)*
  - *Without FFN Layer & 256 tokens/batch ~ 3.1 times*
  - *Without FFN Later & 512 tokens/batch ~ 2.2 times*

Future work

- Run the rest of the experiments on other datasets.
- Explore the most efficient model with comparable results.
- Weight?

- Tm*5 concat baseline is better
- Score
- Irrelevant
- Position
- Other datasets