Apply Chinese Radicals Into Neural Machine Translation: Deeper Than Character Level

Lifeng Han

lifeng.han@adaptcentre.ie
https://github.com/poethan
ADAPT, Dublin City University
Limerick, Ireland, May 24

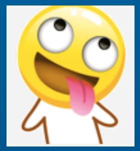
LPRC 2018: Limerick Postgraduate Research Conference

DCU Ollscoil Chathair Bhaile Átha Cliath Dublin City University
sfi Fondúireacht Eolaíochta Éireann Dá bhfuil romhainn Science Foundation Ireland For what's next
European Union European Regional Development Fund
ADAPT Engaging Content Engaging People

Agenda

- Myself
- Topic intro
- Related work
- Proposed idea/model
- Experiments design
- Evaluation results
- Future work

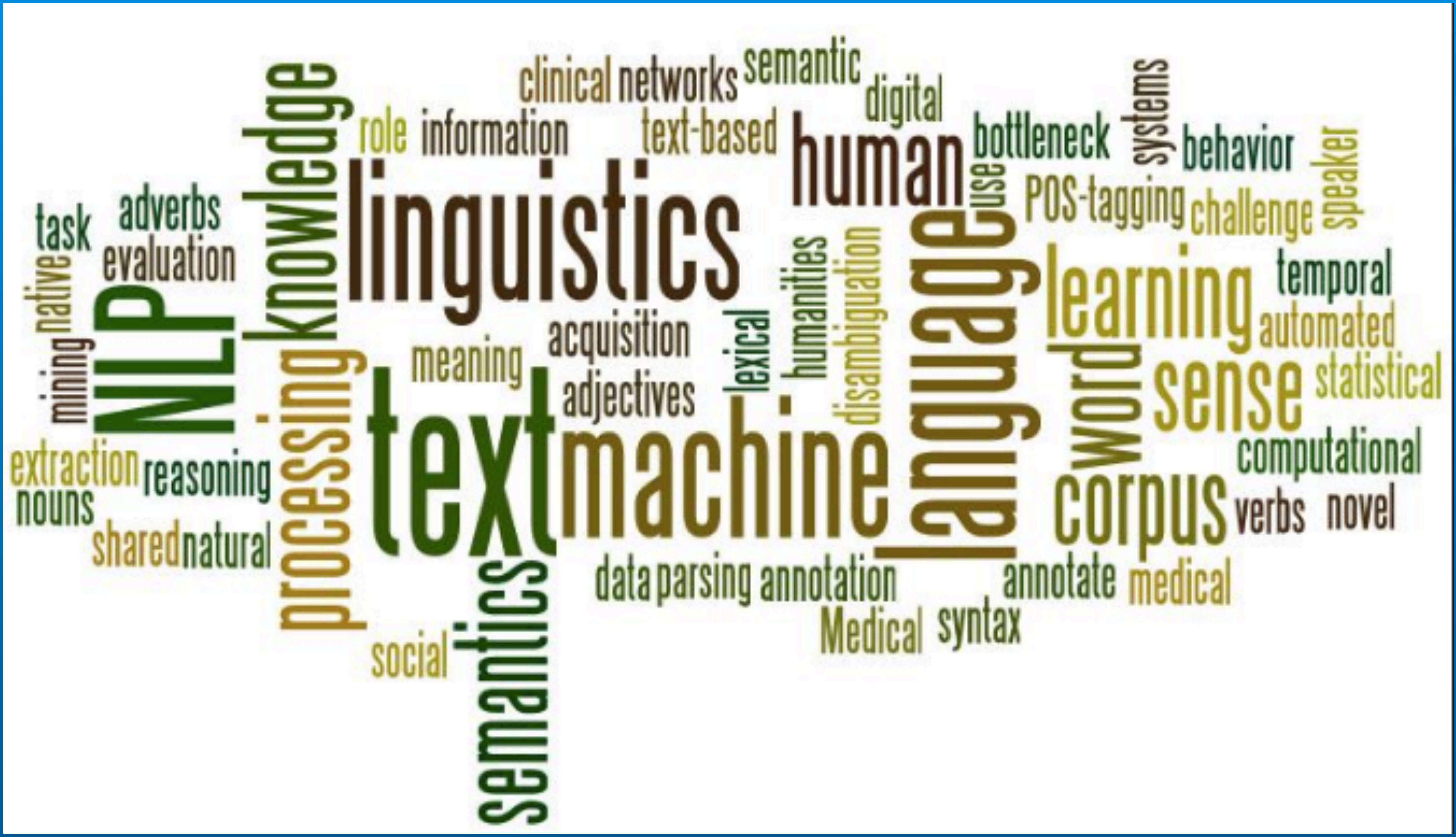
Myself

- PhD student, ADAPT Centre, DCU, Dublin, 2016 on
- Student Researcher, Amsterdam, 2014-16
- Master degree of Sci., Macau, 2011-14
- Bachelor of Maths, Shijiazhuang, 2007-11
- Primary ~ high school, Handan
- No kindergarten 

Intro

- Machine Translation
  - - what I’ m doing. Translate human languages via Machine.
- Natural Language Processing
  - - different processing tasks of human languages
- Artificial Intelligence
  - - teach machine to perform human intelligences

MT-NLP-AI



From: *vikingsna.org*

Related work

- Machine Translation: Rule to Neural
  - - rule, example-based, statistical, phrase-based, hierarchical structure, tree-best, forest, neural models
- Neural MT, sequence to sequence, attention, coverage
  - - word embeddings, sequence to sequence encoding-decoding, attention, coverage, document/discourse level
- Chinese NLP, radical applications
  - - Word Segmentation, Entity recognition, MT, Sentiment Analysis, text mining

Chinese radical: example

Fig. 1: Radical as independent character.

Chinese radical: example

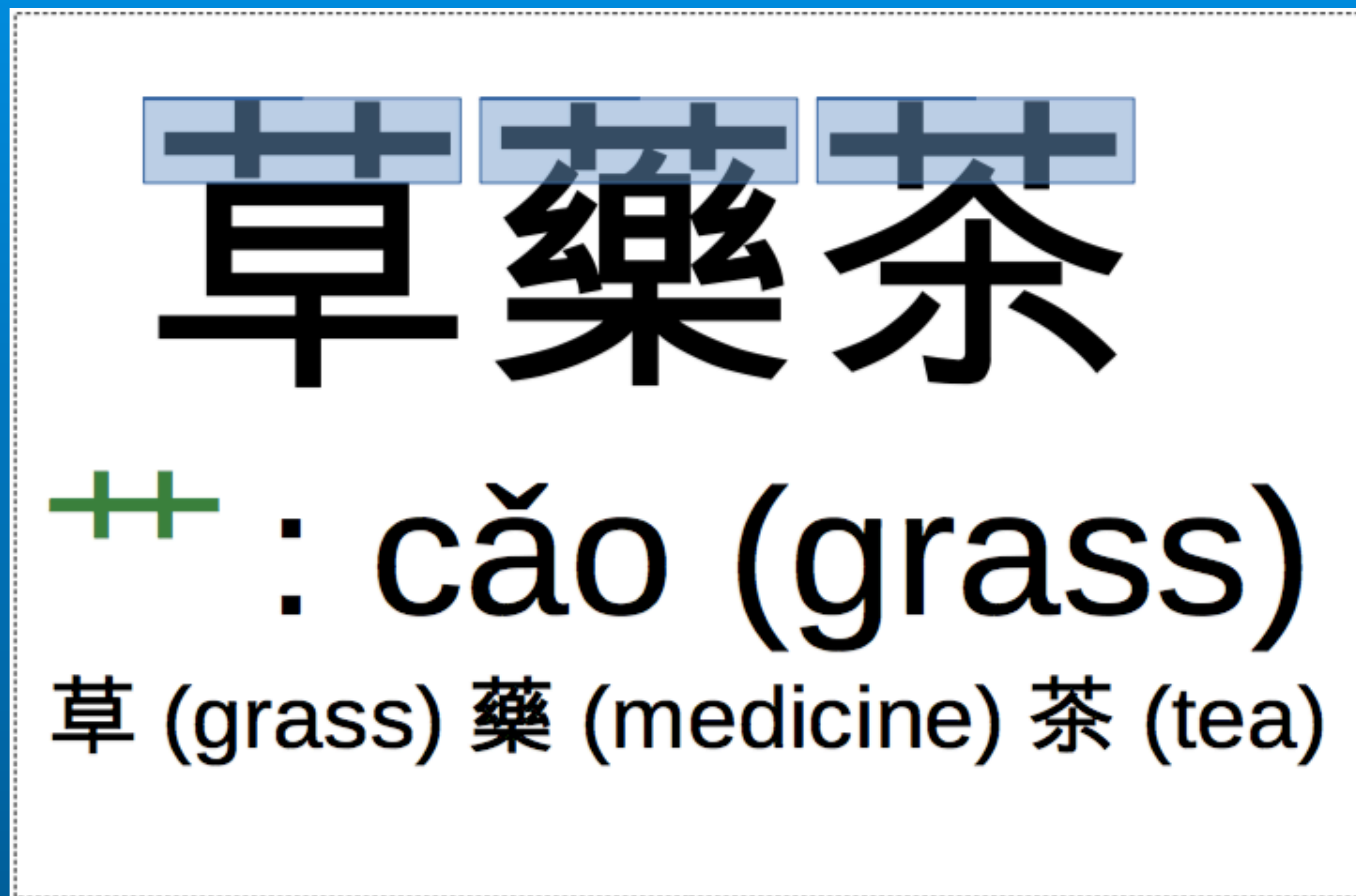


Fig. 2: Radical can not be independent character.

Proposed Model

- Apply Chinese Radical into Translation
  - - how to apply radicals into MT
  - - how to split character into radicals
- Combine radical-level MT with Neural Model
  - - attention-based Neural MT
  - - radical combination into input data

Combinations

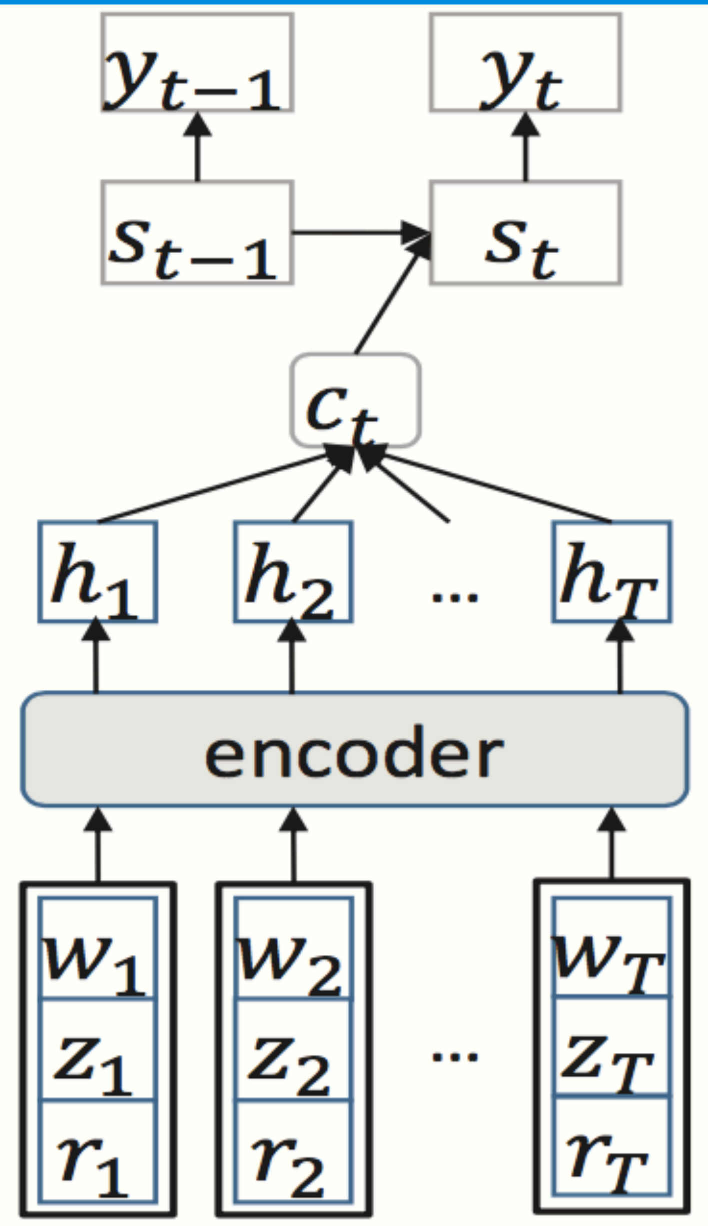


Fig. 3: Architecture of NMT with multi-embedding.

Experiments

- Attention Neural MT
  - Word+Character+Radical
  - Word+Character
  - Character+Radical
  - Word+Radical
- Data prepration
  - Training: 1.25 million parallel Chinese-English sentences / 80.9 millions Chinese words and 86.4 millions English
  - Development / testing: NIST06/NIST08 (National Institute for Standards and Technology, USA)

Settings

Table 1: Model Settings

Settings	Description	abbreviation
Baseline	Words	W
Setting1	Word+Character+Radical	W+C+R
Setting2	Word+Character	W+C
Setting3	Word+Radical	W+R
Setting4	Character+Radical	C+R

Evaluation

- Broader Evaluation Metrics
  - hLEPOR, BEER, CharacTER -> BLEU, NIST
- Evaluation Scores
  - - in-depth analysis

MT evaluation metric LEPOR
Code & WIKI: https://en.wikipedia.org/wiki/LEPOR

Development data BLEU

Table 2: BLEU Scores on NIST06 Development Data

	1-gram	2-gram	3-gram	4-gram
Baseline	.7211	.5663	.4480	.3556
W+C+R	**.7420**	**.5783**	**.4534**	**.3562**
W+C	.7362	.5762	.4524	.3555
W+R	.7346	.5730	.4491	.3529
C+R	.7089	.5415	.4164	.3219

Development data NIST

Table 3: NIST Scores on NIST06 Development Data

	1-gram	2-gram	3-gram	4-gram	5-gram
Baseline	5.8467	7.7916	8.3381	8.4796	8.5289
W+C+R	**6.0047**	**7.9942**	**8.5473**	**8.6875**	**8.7346**
W+C	5.9531	7.9438	8.5127	8.6526	8.6984
W+R	5.9372	7.9021	8.4573	8.5950	8.6432
C+R	5.6385	7.4379	7.9401	8.0662	8.1082

Development data Broader

Table 4: Broader Metrics Scores on NIST06 Development Data

	Metrics on Single Reference		
Models	hLEPOR	BEER	CharacTER
Baseline	.5890	.5112	.9225
W+C+R	.5972	**.5167**	**.9169**
W+C	**.5988**	.5164	.9779
W+R	.5942	.5146	.9568
C+R	.5779	.4998	1.336

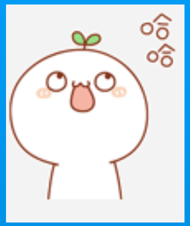
Testing data BLEU 

Table 5: BLEU Scores on NIST08 Test Data

	1-gram	2-gram	3-gram	4-gram
Baseline	.6451	.4732	.3508	.2630
W+C+R	**.6609**	**.4839**	**.3572**	**.2655**
W+C	.6391	.4663	.3412	.2527
W+R	.6474	.4736	.3503	.2607
C+R	.6378	.4573	.3296	.2410

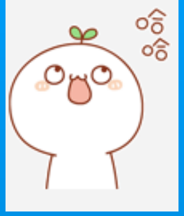
Testing data NIST 

Table 6: NIST Scores on NIST08 Test Data

	1-gram	2-gram	3-gram	4-gram	5-gram
Baseline	5.1288	6.6648	7.0387	7.1149	7.1387
W+C+R	**5.2858**	**6.8689**	**7.2520**	**7.3308**	**7.3535**
W+C	5.0850	6.5977	6.9552	7.0250	7.0467
W+R	5.1122	6.6509	7.0289	7.1062	7.1291
C+R	5.0140	6.4731	6.8187	6.8873	6.9063

Testing data Broader 

Table 7: Broader Metrics Scores on NIST08 Test Data

	Metrics Evaluated on 4-references		
Models	hLEPOR	BEER	CharacTER
Baseline	.5519	.4748	**0.9846**
W+C+R	**.5530**	**.4778**	1.3514
W+C	.5444	.4712	1.1416
W+R	.5458	.4717	0.9882
C+R	.5353	.4634	1.1888

Future work

- Improve parameter optimisation/tuning models
- Include more testing data
- Include different domain data
- Reduce training data and test low-resource scenario
- This paper pre-print: https://arxiv.org/pdf/1805.01565.pdf

Follow the project

- LEPOR: https://github.com/poethan/LEPOR/
- Chinese character decomposition: https://github.com/poethan/MWE4MT/tree/master/radical4mt
- Acknowledgment:

  - The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

Selected references

- ALF Han, DF Wong, LS Chao. 2013. Chinese named entity recognition with conditional random fields in the light of Chinese characteristics. Intelligent Information Systems Symposium, 57-68.

- ALF Han, DF Wong, LS Chao. 2013. LEPOR: A Robust Evaluation Metric for Machine Translation with Augmented Factors Proceedings of the 24th International Conference on Computational Linguistics.

- ALF Han, DF Wong, LS Chao, L He, Y Lu, J Xing, X Zeng . 2013. Language-independent Model for Machine Translation Evaluation with Reinforced Factors. Machine Translation Summit XIV, 215-222.

- L Han. 2018. Machine Translation Evaluation Resources and Methods: A Survey. IPRC: Ireland Postgraduate Research Conference. http://doras.dcu.ie/24493/

- Lifeng Han and Shaohui Kuang. 2018. Apply Chinese radicals into neural machine translation: Deeper than character level. ArXiv pre-print https://arxiv.org/abs/1805.01565v1