Fine-grained Attention Mechanism for Neural Machine Translation

Presenter：Baosong Yang

Fine-grained Attention Mechanism for Neural Machine Translation

- Only one scalar value is assigned
- Each dimension of a context vector should receive a separate attention score

- Context (Context-Dependent
- Word Representation for Neural Machine Translation):

	En-De		En-Fi	
Beam Width	1	12	1	12
Baseline	17.57 (17.62)	20.78 (19.72)	6.07 (7.18)	7.83 (8.35)
+AttY	19.15 (18.82)	21.41 (20.60)	7.38 (8.02)	8.91 (9.20)
+AttY2D	20.49 (19.42)	22.50 (20.83)	8.33 (8.75)	9.32 (9.41)
+Context(C)	19.13 (18.81)	22.13 (21.01)	7.47 (7.93)	8.84 (9.18)
+C+AttY	20.96 (20.06)	23.25 (21.35)	8.67 (9.18)	10.01 (9.95)
+C+AttY2D	22.37 (20.56)	**23.74** (22.13)	9.02 (9.63)	**10.20** (10.90)

$$c^x = \frac{1}{T}\sum_{t=1}^{T} \mathrm{NN}_\theta(x_t)$$

$$x_t \leftarrow x_t \odot \sigma(W_x c^x + b_x)$$
$$y_t \leftarrow y_t \odot \sigma(W_y c^x + b_y)$$

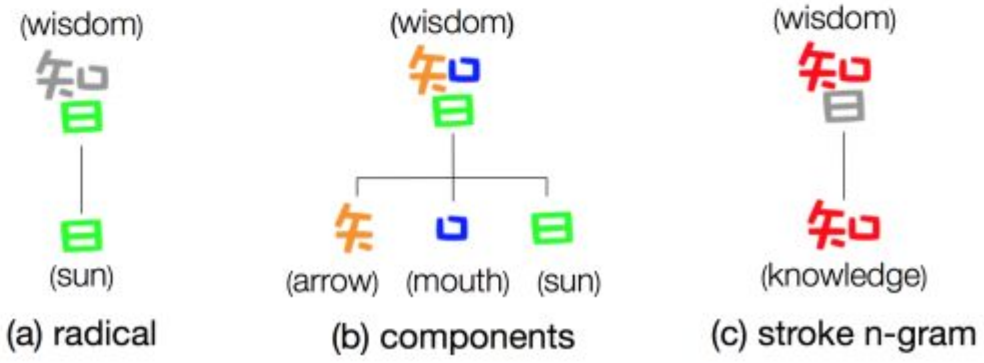
- Similar to Multi-head and Multi-dimention
- Use gate in Context-aware SAN？

cw2vec: Learning Chinese Word Embeddings with Stroke n-gram Information

Presenter：Baosong Yang

Motivation

- Largely focused on European languages
- Latin script and s Chinese employ different writing system

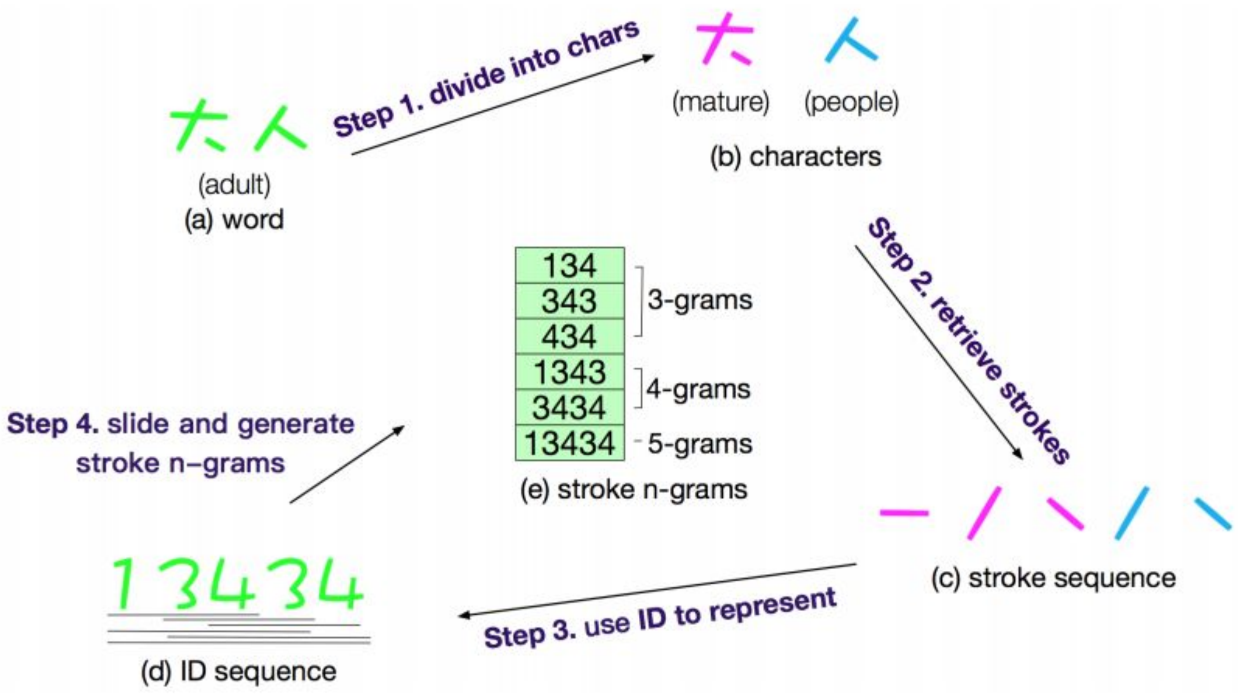


Model

- 1. Word to Stroke n-gram

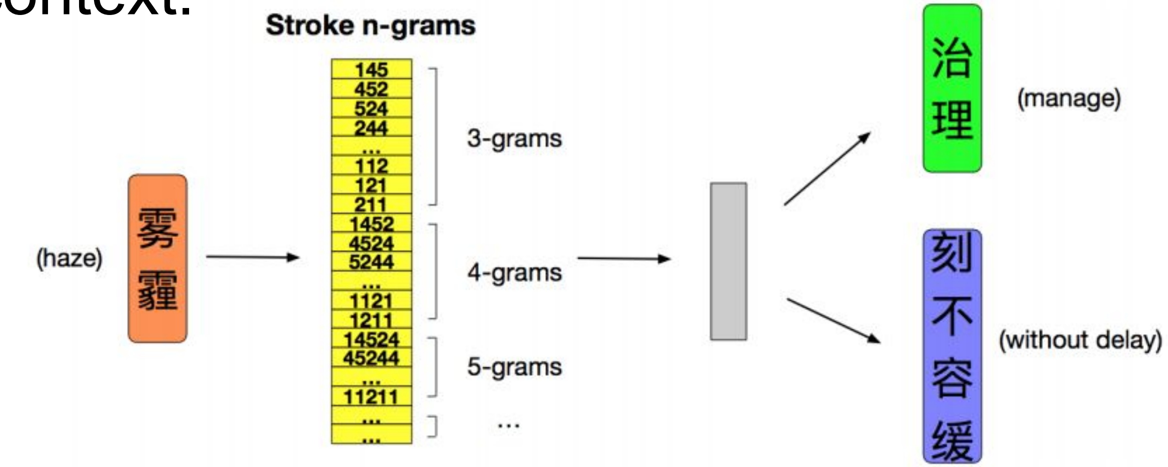
Stroke Name	Horizontal	Vertical	Left-falling	Right-falling	Turning
Shape, ID	一 (㇀), 1	丨 (亅), 2	丿, 3	㇏ (丶), 4	㇆ (㇄㇉), 5

Figure 3: General shapes of Chinese strokes.



Training

- 1. Objective Function: Same to W2V, i.e. Measuring the similarity between a word and its context.



- Stroke n-gram dictionary S, S(w) denote the word w.

$$sim(w,c) = \sum_{q \in S(w)} \vec{q} \cdot \vec{c}$$

- Negative sampling: replace the expensive denominator with a collection of context words "negatively" sampled based on a distribution.

$$\mathcal{L} = \sum_{w \in D} \sum_{c \in T(w)} \log \sigma(sim(w,c)) + \lambda \mathbb{E}_{c' \sim P}[\log \sigma(-sim(w,c'))]$$

Experiment

- **Data:** consists of 265K Chinese Wikipedia articles.
- **Benchmarks:** Word Similarity Task, Word Analogy Task (1,124), Text Classification, Named Entity Recognition
- **Baseline:** W2V, GloVe, CWE (character + word), GWE (font images), JWE (Components)

Model	Word Similarity		Word Analogy		Text Classification	Named Entity Recognition
	wordsim-240	wordsim-296	3CosAdd	3CosMul		
skip-gram (Mikolov et al. 2013b)	44.2	44.4	58.3	58.9	93.4	65.1
cbow (Mikolov et al. 2013b)	47.0	50.2	54.3	53.5	93.4	59.6
GloVe (Pennington, Socher, and Manning 2014)	45.2	44.3	68.8	66.7	94.2	66.0
CWE (Chen et al. 2015)	50.0	51.5	68.5	69.6	93.2	65.8
GWE (Su and Lee 2017)	50.0	49.1	50.8	50.6	94.3	65.5
JWE (Xin and Song 2017)	48.0	**52.7**	74.2	76.3	94.2	67.9
`cw2vec` (stroke n-grams)	**50.4**	**52.7**	**78.1**	**80.5**	**95.3**	**71.7**

Case study

Targets	GWE	JWE	GloVe	CWE	cw2vec
水污染 (water pollution)	污染源(pollutant src)	荒漠化(desertification)	公害(public nuisance)	污染源(pollutant src)	污染(pollution)
	污染(pollution)	污染物(pollutant)	废弃物(garbage)	污染(pollution)	污染物(pollutant)
	水害(water damage)	内涝(waterlogging)	洪涝(flood)	污染物(pollutant)	水质(water quality)
	污泥(sludge)	排污(pollution discharge)	奶制品(dairy product)	水害(water damage)	水资源(water resource)
	沙漠化(desertization)	油污(oil pollution)	循环系统(circulatory sy)	污泥(sludge)	污染源(pollutant src)
	污水(sewage)	沙漠化(desertization)	神经系统(nervous sy)	污水(sewage)	废水(waste water)
	污渍(stain)	地表水(surface water)	市容(city appearance)	污渍(stain)	荒漠化(desertification)
	废水(waste water)	盐碱化(salinization)	职业病(occupational ds)	污物(dirt)	地下水(groundwater)
	渗水(leakage)	渗漏(seepage)	结构性(designability)	废水(waste water)	地表水(surface water)
	污垢(filth)	公害(public nuisance)	污染(pollution)	渗水(leakage)	沙漠化(desertization)
孙悟空 (Sun Wukong)	孙悟天(Son Goten)	唐僧(Monk Tang)	唐僧(Monk Tang)	孙悟天(Son Goten)	沙悟净(Sha Wujing)
	孙悟饭(Son Gohan)	孙悟饭(Son Gohan)	孙悟饭(Son Gohan)	孙悟饭(Son Gohan)	白骨精(Bai Gujing)
	小悟(Xiao Wu)	白骨精(Bai Gujing)	白骨精(Bai Gujing)	小悟(Xiao Wu)	西游记(J. to the West)
	龙珠(Dragon Ball)	沙悟净(Sha Wujing)	西游记(J. to the West)	阿悟(A Wu)	沙僧(Monk Sha)
	甘悟(Gan Wu)	西游记(J. to the West)	龙珠(Dragon Ball)	沙悟净(Sha Wujing)	猴王(Monkey King)
	阿悟(A Wu)	唐三藏(Xuanzang)	三打(three strikes)	甘悟(Gan Wu)	孙悟天(Son Goten)
	玉悟(Yu Wu)	贝吉塔(Vegeta)	沙悟净(Sha Wujing)	董悟(Dong Wu)	唐三藏(Xuanzang)
	天大(extremely big)	红孩儿(Red Boy)	唐三藏(Xuanzang)	玉悟(Yu Wu)	贝吉塔(Vegeta)
	真飞龙(really dragon)	猴王(Monkey King)	色狼(lady-killer)	西游记(J. to the West)	龙珠(Dragon Ball)
	悟(Wu)	沙僧(Monk Sha)	阿哥(A Ge)	龙珠(Dragon Ball)	孙悟饭(Son Gohan)

Conclusion

- Chinese Charater different to Latin script
- The model maybe useful for machine translation:
  - semantic
  - reduce parameter size
  - addressing OOV
- Simplified is better than tranditional?
- Chinese is profound:
  - Stroke: ((八，人), (田，由)) => Radicals => Homomorphic (+BPE)
  - Hieroglyphics (+ font recognition): (恶梦，噩梦)
  - Homophone (+ pinyin)：木(a tree) 林(some trees) 森(a lot of trees)
  - or incorporating above