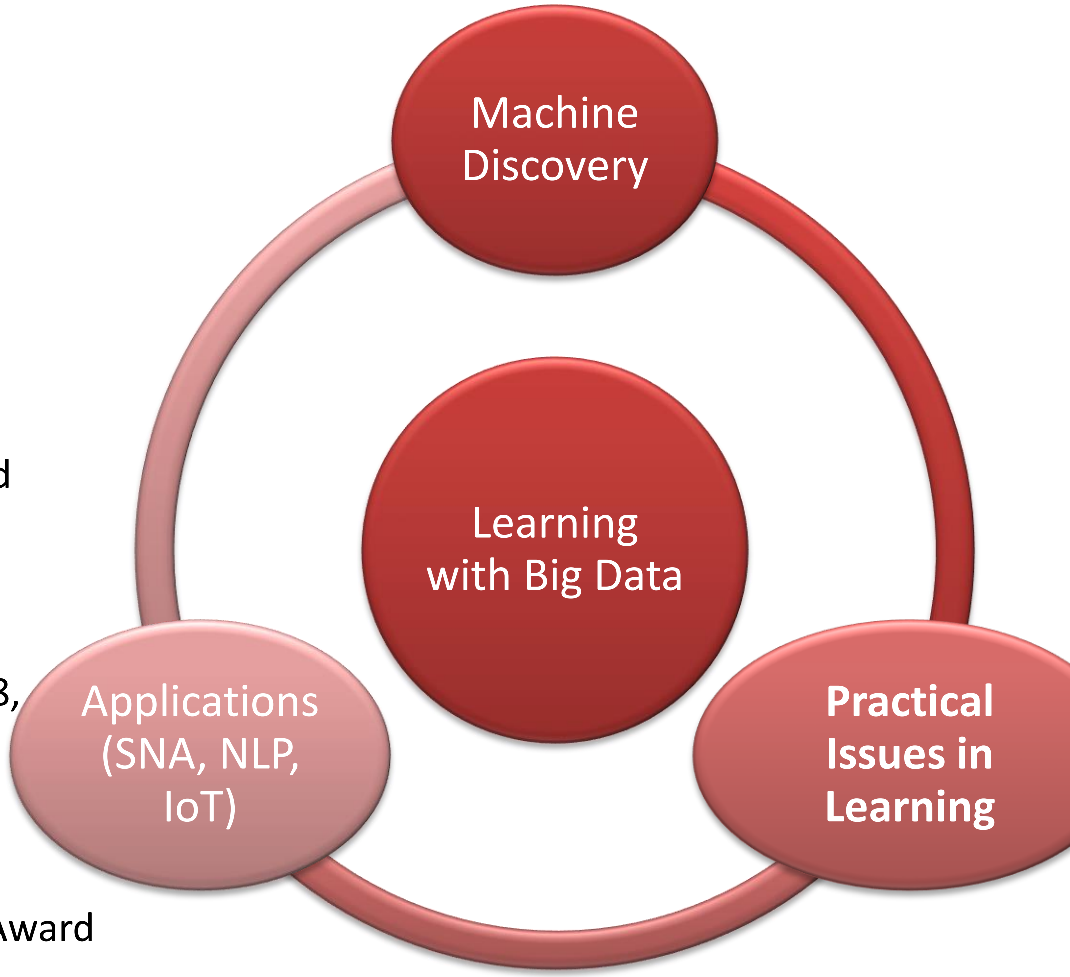
Can Machine Be Creative?

Shou-de Lin （林守德教授）
NTU CSIE
sdlin@csie.ntu.edu.tw

About Machine Discovery and Social Network Mining LAB

- **PI:** Shou-de Lin
  - B.S. in NTUEE
  - M.S. in EECS, UM
  - M.S. in Computational Linguistics, USC
  - Ph.D. in CS, USC
  - Postdoc in LANL
- **Courses:**
  - Machine Discovery
  - Social network Analysis
  - Technical Writing and Research Method
  - Statistical Artificial Intelligence
  - Probabilistic Graphical Model
- **Awards:**
  - All-time ACM KDD Cup Champion (2008, 2010, 2011, 2012,2013)
  - WSDM Cup 2016 Champion
  - Best Paper Award WI2003, ASONAM 2011, TAAI 2010, 2014, 2016
  - US Areospace AROAD Research Grant Award 2011, 2013, 2014, 2015
  - Google Research Award 2008
  - Microsoft Research Award 2009, 2015, 2016
  - IBM University Research Award 2015



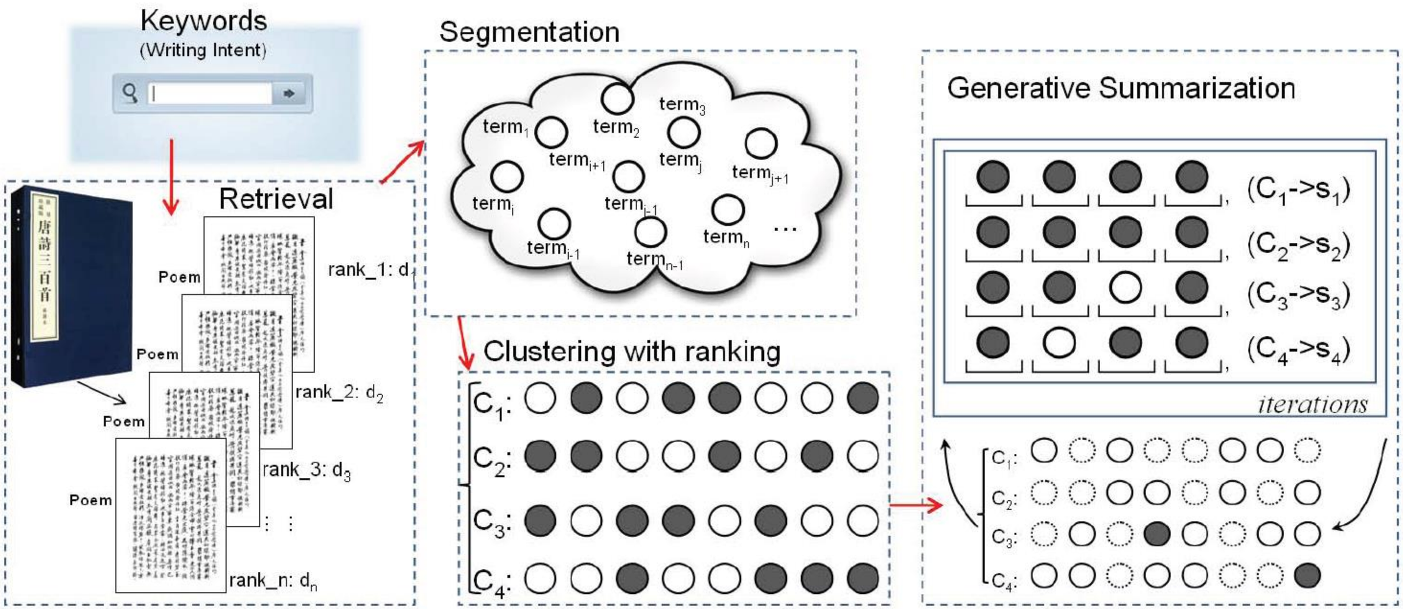
We worked on Two Directions for the Past 1x Years

- Machine Discovery
- Addressing Practical Issues in Machine Learning

從電腦寫詩到電腦寫歌詞

- 小冰寫詩
  - 像每一座城市愧對鄉村
    我才有一個美好的完成
    每個失眠的夜晚
    我是一個花言巧語的人
    隱匿在靈魂最迷失的火
  - 繞出城市的邊緣
    美好的
    在風裡
    最輕微的觸動

電腦創作古典詩(IJCAI13)



故人	思鄉	回文詩
千里行路難	何處不自由	行人路道路人行
平生不可攀	故土木蘭舟	別有心人心有別
相逢無人見	落日一回首	寒風幾度幾風寒
與君三十年	孤城不可留	

歌詞改寫

記得

誰還記得是誰先說
永遠的愛我
以前的一句話是我們
以後的傷口
過了太久沒人記得
當初那些溫柔
我和你手牽手說要一起
走到最後

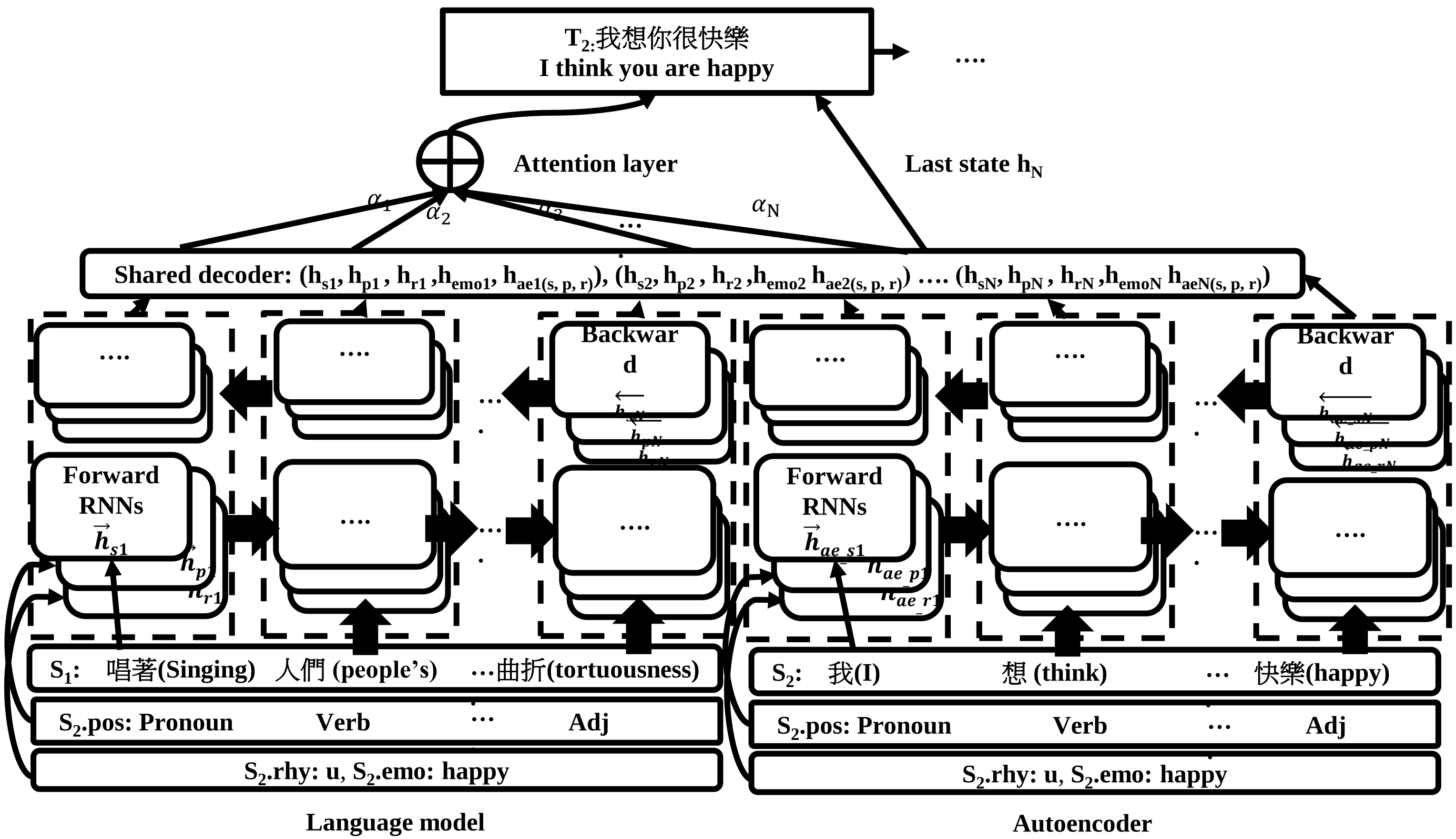
不懂

雨都停了天都亮了
我們還不懂
這愛情路究竟帶我們
到甚麼地方
是要持續仍舊珍惜
還是回到原地
如今此刻的我的確是有
一點疲倦

Lyrics Rewriting

- Goal 1: the re-written lyric has to be performed under the same tune.
- Goal 2: control the rhyme
- Goal 3: control the emotion
- Goal 4: preserve the semantics

Our Multi-task Model



Auto-Evaluation Results

- Average human ratings.

	compatibility with originals (max 5 points)	rhyme preference (max 3 points)	center theme preference (max 3 points)	fluency compared with originals (max 3 points)
single task RNN	2.04	1.70	1.38	1.35
cluster based	2.00	2.12	1.34	1.26
our model without smoothing	2.81	1.90	1.74	1.78
our model	**3.17***	**2.15**	**2.09***	**2.14***

* $p < 0.001$

- Automatic evaluation on POS and rhyme format, accuracy, segmentation error

	POS format accuracy	rhyme accuracy	BLEU	ROUGE-2	segmentation length
single task RNN	0.7627	0.6592	0.3792	**0.055**	0.5138
cluster based	0.1159	0.2351	**0.4424**	0.144	1.446
Our model without smoothing	0.6704	0.6809	0.4223	0.282	0.5600
Our model	**0.8277**	**0.7896**	0.4355	0.106	**0.4066**

- Augmenting pinyin feature to sequential features.

	using only ending rhyme	using rhyme sequence
precision on ending rhyme	80.2%	79.9%
precision on internal rhyme	8.6%	17.8%

Sample Results

Source text	Target 1	Target 2
我重溫午後的陽光 將吉他斜背在肩上 跟多年前一樣 我們輕輕地唱 去任何地方	他像明日的寶藏 也許太陽曬過你在遠方 在很久很久以前一樣 我最深的走廊 等待這地方	我像秋天的張揚 默默珍藏留下你在身上 在很久很久以前一樣 我們最深的地方 讓我們歌唱

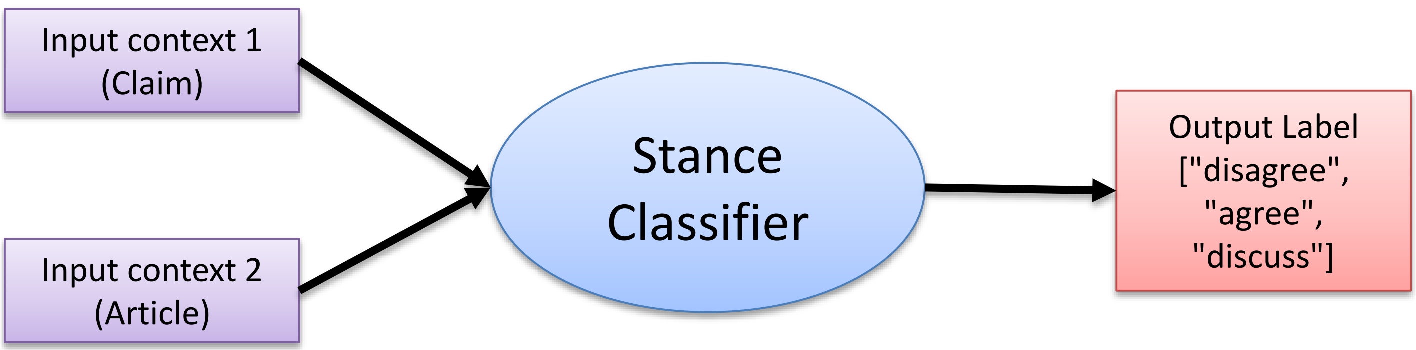
Emotion Transfer

Source text	Transfer to happy	Transfer to sad
從那天起 甜蜜的很輕易 親愛的 別任性 你的眼睛 在說我願意	為你等待 簡單的好疲憊 幸福的 這愛情 我的眼睛 在等待有人開啟	當我放棄 寂寞的好疲憊 傷害的 我證明 你的任性 被欺騙我放棄

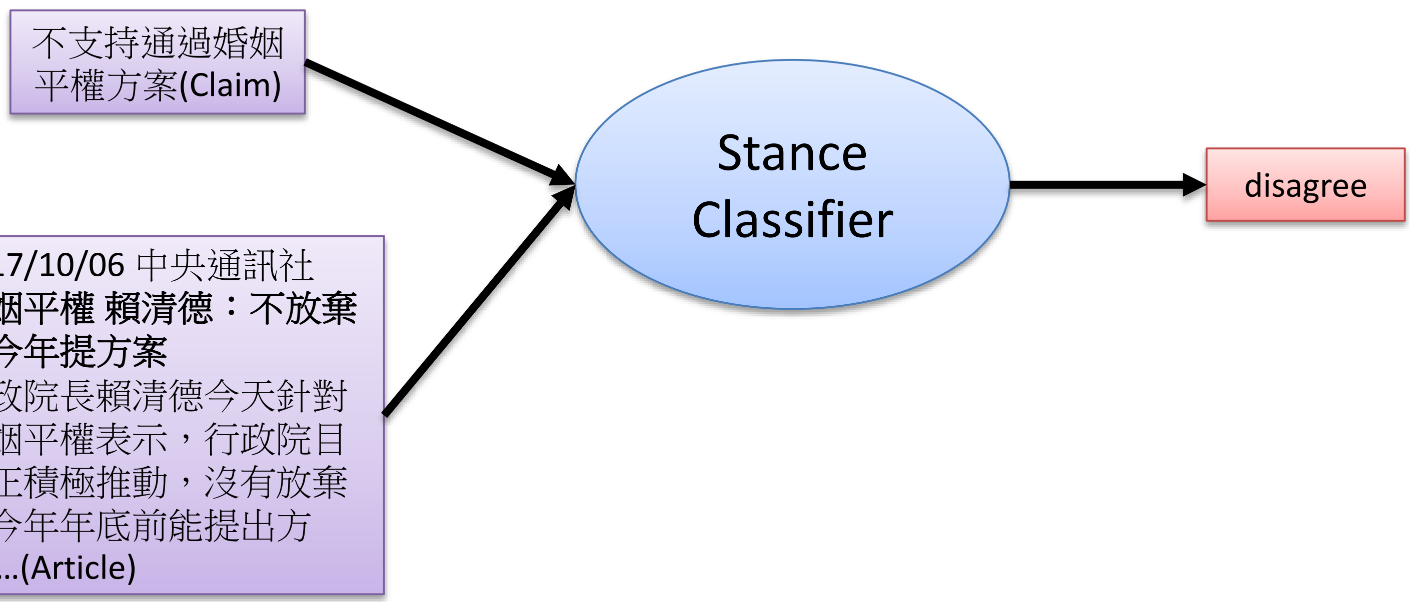
Stance Classification with Attention Mechanism

Joint Work with Te-Wang Chiu

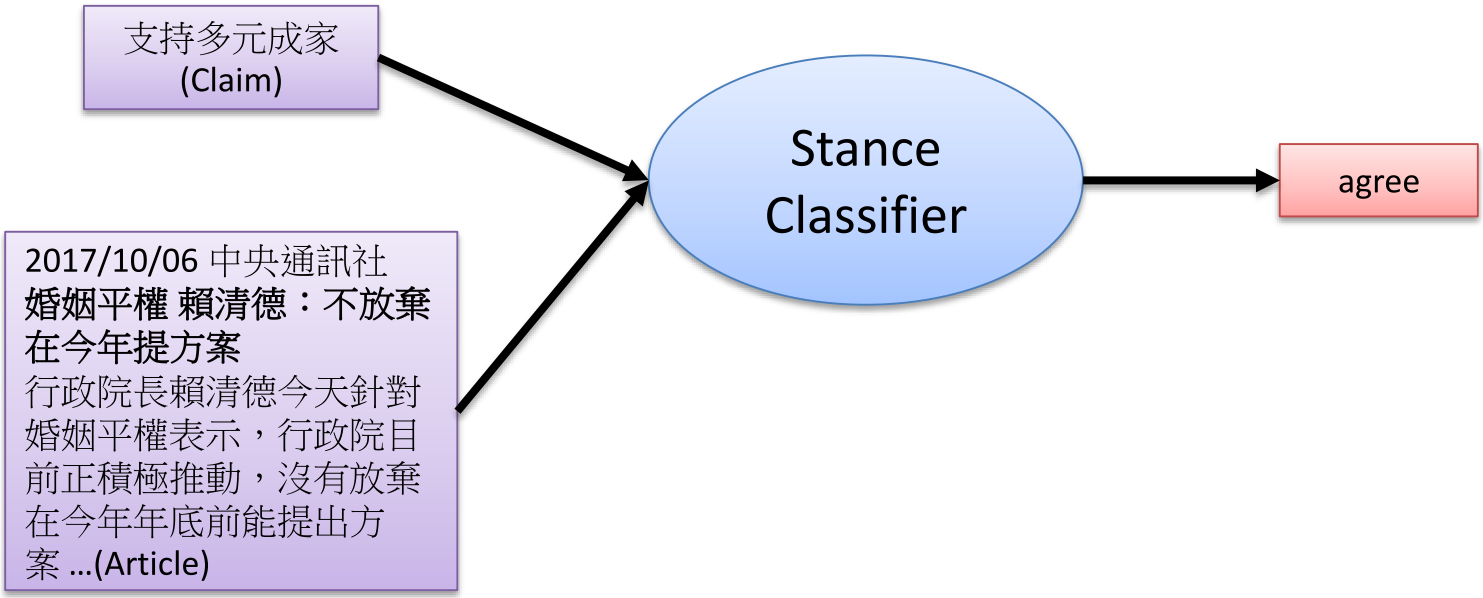
Stance Classification Task



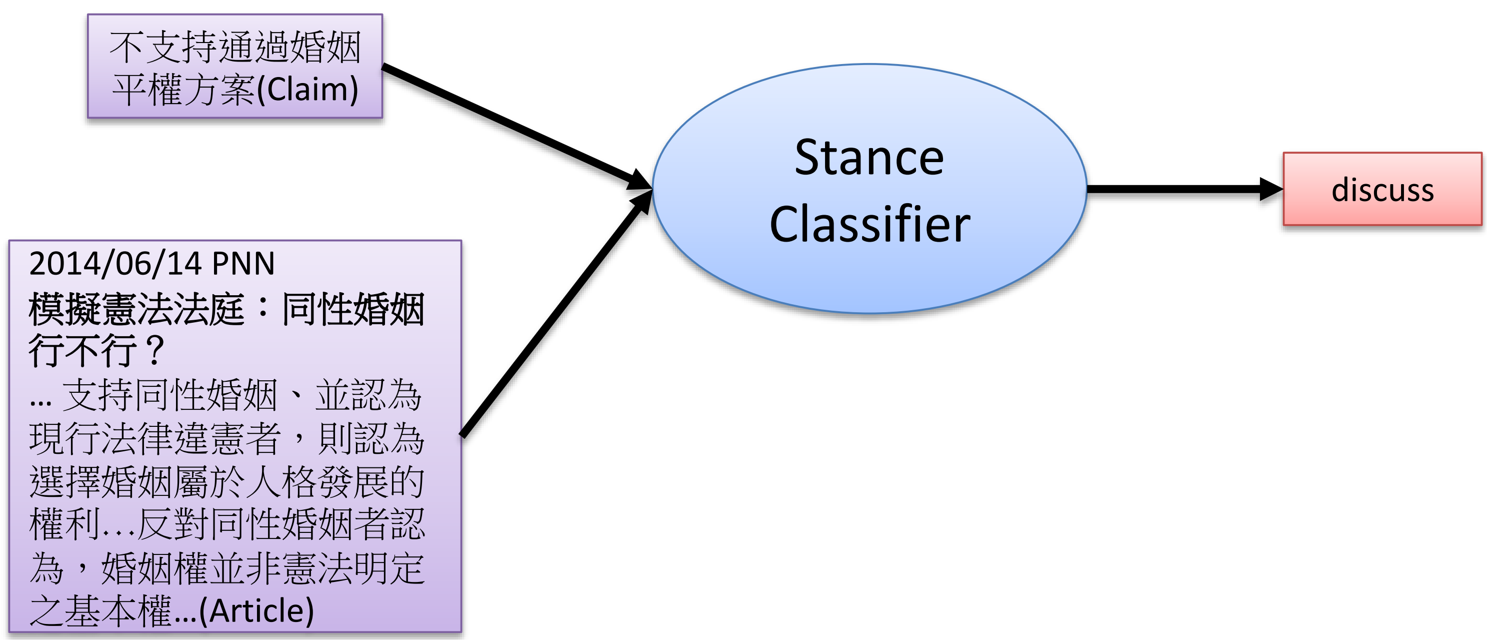
Stance Classification Example(1/3)



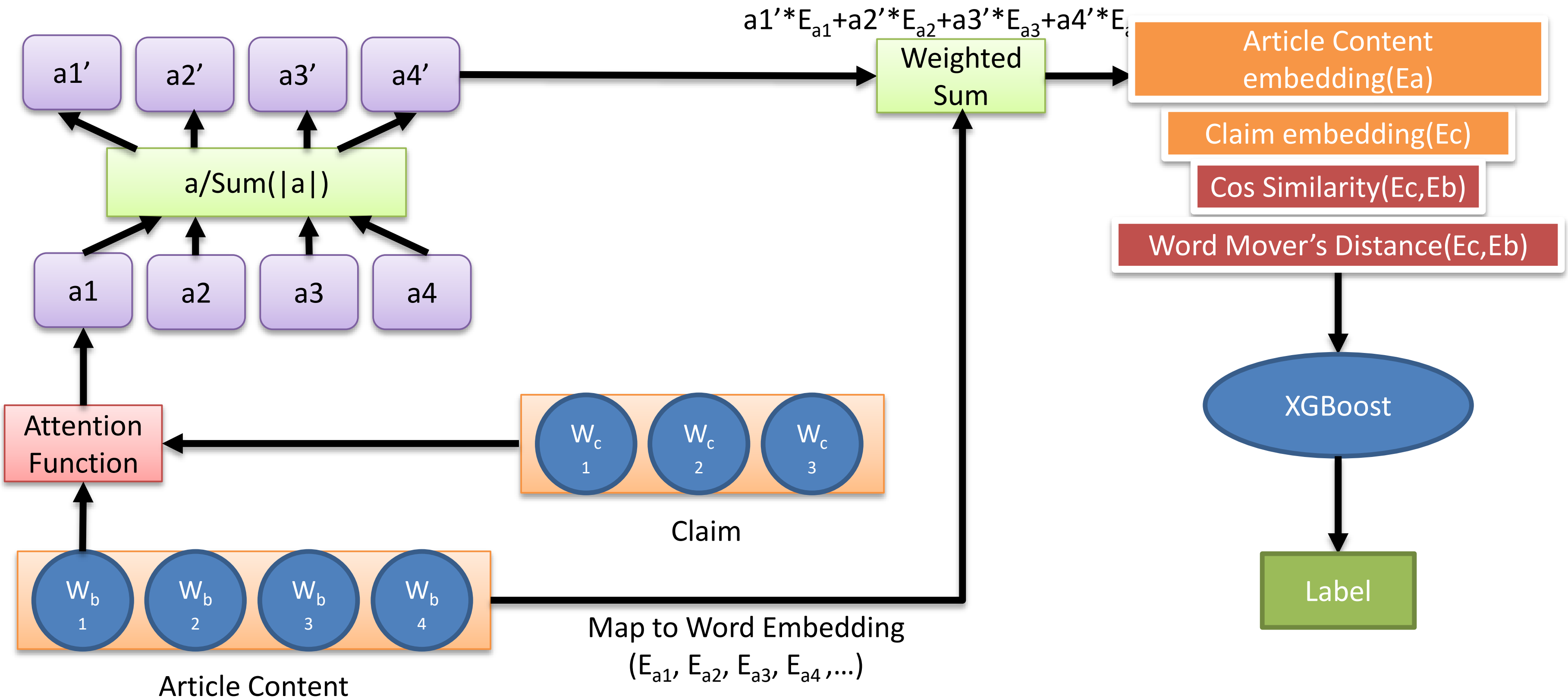
Stance Classification Example(2/3)



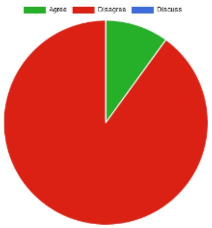
Stance Classification Example(3/3)



Attention-Based Solution



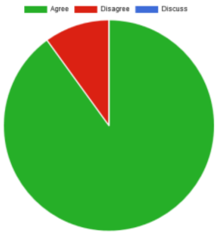
Sample Results (1/2) – 美國牛進口



Claim :支持美國牛輸入台灣

	標題	出處	時間	立場
0	違法進口日本商品大創坦承竄改文件	民視新聞	2018-4-27	disagree
1	美豬牛再被點名經濟部盼TIFA管道溝通	中央社即時新聞	2018-3-31	disagree
2	創科局愚者千慮有一得	on.cc東網	2018-05-11	disagree
3	美爆第5例狂牛症美牛繼續進口嗎？	udn 聯合新聞網	2017-7-19	disagree
4	美促台解禁美牛豬關切稻米、米酒	中時電子報	2018-3-31	disagree
5	台灣「美」其林百味美牛餐廳首批27間名單曝光	udn 聯合新聞網	2018-3-6	disagree
6	他放棄400萬年薪攻破南洋女人心	Cheers快樂工作人雜誌	2018-4-19	disagree
7	AIT主席莫健：敦促台灣解決美牛豬問題	udn 聯合新聞網	2017-12-16	agree
8	年收近百萬贏9成上班族？ 專家這樣說	民視新聞	2018-05-08	disagree
9	因狂牛症被禁14年，台灣將開放日本等國牛肉進口	The News Lens 關鍵評論網	2017-7-17	disagree

Sample Results (2/2) – 美國牛進口



Claim :反對美國牛輸入台灣

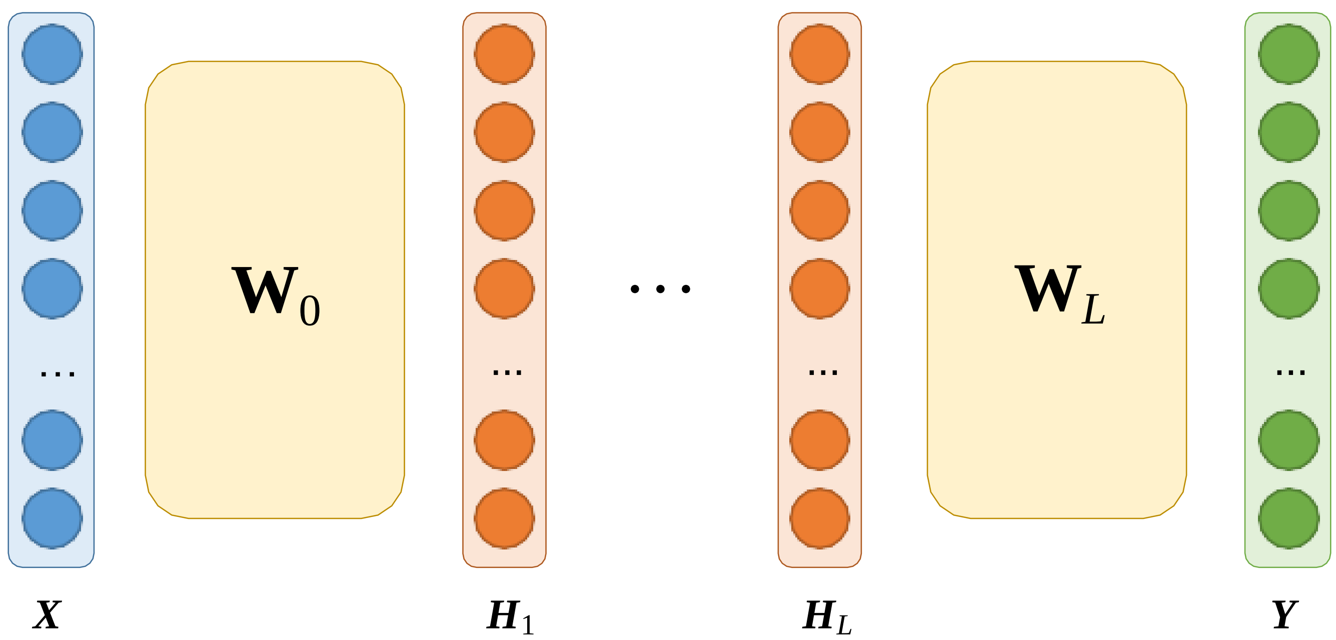
	標題	出處	時間	立場
0	違法進口日本商品大創坦承竄改文件	民視新聞	2018-4-27	agree
1	美豬牛再被點名經濟部盼TIFA管道溝通	中央社即時新聞	2018-3-31	agree
2	創科局愚者千慮有一得	on.cc東網	2018-05-11	agree
3	美爆第5例狂牛症美牛繼續進口嗎？	udn 聯合新聞網	2017-7-19	agree
4	美促台解禁美牛豬關切稻米、米酒	中時電子報	2018-3-31	agree
5	台灣「美」其林百味美牛餐廳首批27間名單曝光	udn 聯合新聞網	2018-3-6	agree
6	他放棄400萬年薪攻破南洋女人心	Cheers快樂工作人雜誌	2018-4-19	agree
7	AIT主席莫健：敦促台灣解決美牛豬問題	udn 聯合新聞網	2017-12-16	disagree
8	年收近百萬贏9成上班族？ 專家這樣說	民視新聞	2018-05-08	agree
9	因狂牛症被禁14年，台灣將開放日本等國牛肉進口	The News Lens 關鍵評論網	2017-7-17	agree

Fast Neural Network Training via Sparse Activation Functions

Introduction

Deep neural networks have shown their powerful fitting ability among numerous applications.

In particular, let's consider a deep *fully-connected* neural network with L hidden layers.



Computational Bottlenecks (1/2)

- We have to learn the weight matrices $W_0, \cdots, W_L$ during the training process.
- Given an activation function σ (e.g., $\max(x, 0)$),

$$H_{k+1}(j) = \sigma\left(\sum_{i=1}^{h_k} W_k(i,j) H_k(i)\right) \quad (j = 1, \cdots, h_{k+1})$$

- Objective (for a loss function $\mathcal{L}$):

$$\min_{W_0,\cdots,W_L} \mathcal{L}(\hat{Y}, Y) \equiv \min_{W_0,\cdots,W_L} \mathcal{L}(H_{L+1}, Y)$$

Computational Bottlenecks (2/2)

- For the k-th weight matrix, we have to determine its $h_k \times h_{k+1}$ entries (let $H_0 := X$ and $H_{L+1} := Y$).

- To calculate gradient, consider computational costs at the k-th layer:

1. Forward pass: fix all W and calculate H.
   $\Rightarrow \Theta(h_k h_{k+1}) = \Theta(\text{size}(W_k))$
2. Backward update (*back-propagation*): compute gradients from the loss:
   $\Rightarrow \Theta(h_k h_{k+1}) = \Theta(\text{size}(W_k))$

Weight Matrices Sparsification (1/2)

- Based on the analyses, we can see the size of W's affects the overall training efficiency.
- The entries of each W is quadratic to the number of hidden units h on each layer.
  - For example, if each layer has $1,000$ hidden units, then each W will contain 10^6 entries.

Research Questions:

$\Rightarrow$ Can we use a sparse W?

$\Rightarrow$ How many *non-zero* entries are required for W?

Weight Matrices Sparsification (2/2)

- Consider parameters of the k-th layer W_k, fixing all the others. Add a *small ℓ_1 penalty* to the objective function:

$$\min_{W_k \in \mathbb{R}^{h_k \times h_{k+1}}} \mathcal{L}(\hat{Y}(W_k), Y) + \lambda \|\text{vec}(W_k)\|_1$$

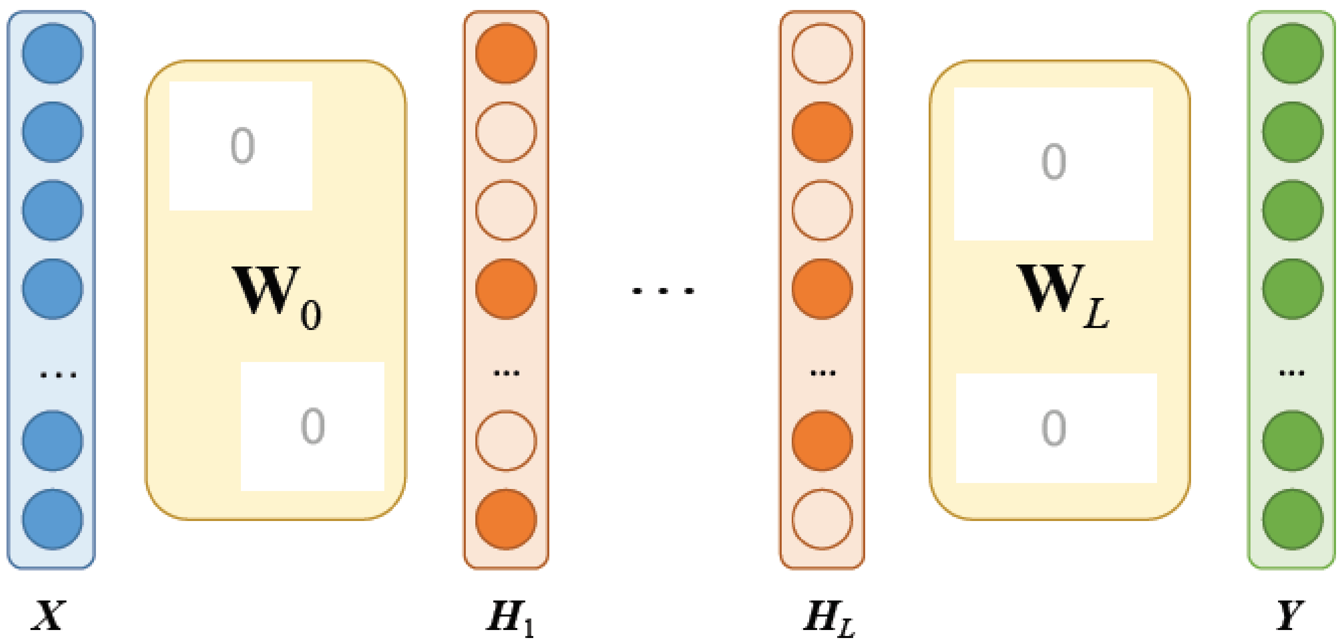
.

- **An upper bound**: For any $\lambda > 0$, the stationary point of the above minimization problem satisfies:

$$nnz(W_k) \leq \left| \{(i,j) \mid \nabla_{\hat{Y}_{ij}} L(\hat{Y}, Y) \neq 0\} \right|$$

Sparse Activation Function (1/2)

- To make W and H sparse, we introduce Top-Q as the sparse activation function: "only preserve the Q *largest* $H_k(j)$ among all j for each layer k, and set the remaining 0"



Sparse Activation Function (2/2)

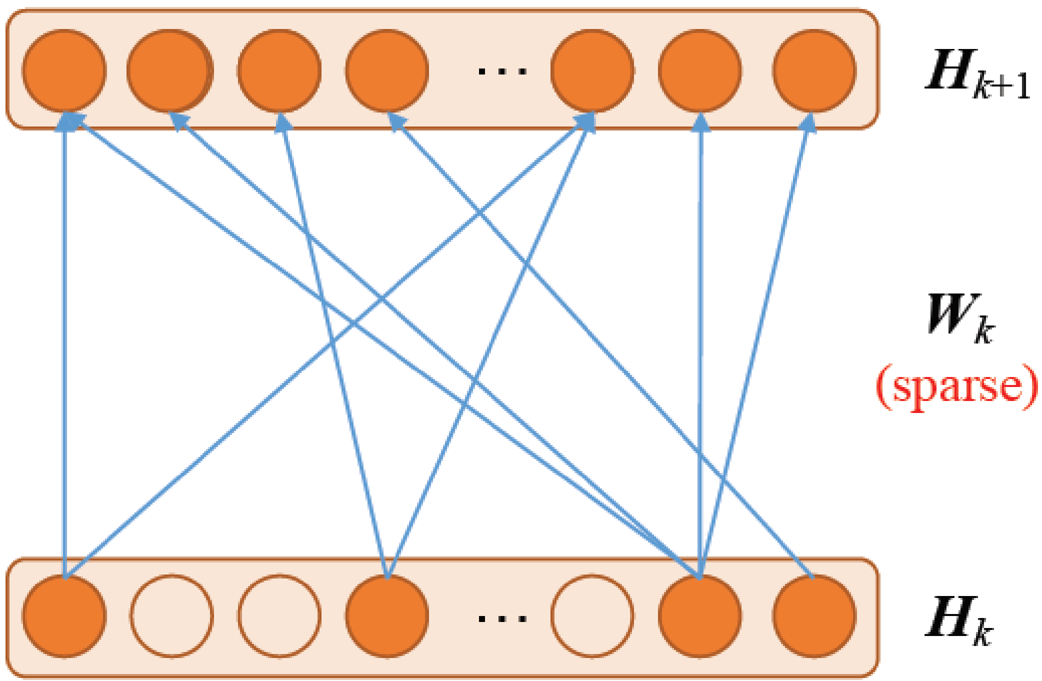
- Empirically, Top-Q can achieve similar model performance (e.g., prediction accuracy on `mnist` $\approx$ 99%) as the original dense activation.
- Now, the computational cost at k-th layer becomes

$$\Theta\left(\underbrace{\text{nnz}(H_k)\text{nnz}(W_k^{\text{row}})}_{\text{forward}} + \underbrace{\text{nnz}(W_k^{\text{col}})\text{nnz}(H_{k+1})}_{\text{backward}}\right).$$

- For instance, if $h = 1000$, $Q = 50$, then we have $\frac{50}{1000} = \frac{1}{20}$ density on each H. Suppose that $\text{nnz}(W) = \frac{1}{10}\text{size}(W)$, then the training will be $20 \times 10 = 200$ times faster.

Sprase Activation Function (3/3)

$$\Theta\left(\underbrace{Q \cdot \mathsf{nnz}(W_k^{\mathsf{row}})}_{\mathsf{forward}} + \underbrace{\mathsf{nnz}(W_k^{\mathsf{col}}) \cdot Q}_{\mathsf{backward}}\right)$$



Thank You!!