On the Quality of Neural Machine Translation of Public Health Information in Hong Kong (English-Chinese Translation)

SIU Sai Cheong (School of Translation, Hang Seng Management College)
SIU Sai Yan (School of Public Health, The University of Hong Kong)

Outline

- Background
- Public Health Information
- Machine Translation of Public Health Information
- New Possibilities Offered by Neural Machine Translation
- Evaluation
- Issues and Suggestions

Background

1. Timely and accurate translation of public health materials is of great importance
2. The use of automatic translation tools could be useful, especially after the rise of neural machine translation systems in recent years
3. However, there have been few studies of its application to the public health domain

This Project

1. Studying neural machine translation of public health materials in Hong Kong
2. Considering three online machine translation engines: Google Translate, Microsoft Translator and Sogou Translator

Public Health Information

Public Health Information

1. Public health departments play a pivotal role in disseminating accurate and comprehensible health promotion messages to the public (Regidor et al., 2007)
2. Public health information is available in different formats including pamphlets, posters, press releases, web pages, TV announcements, video clips and even mobile applications.

Sources of Public Health Information in Hong Kong

1. The Department of Health: Health education, surveillance and epidemiology
2. The Hospital Authority: Management of public hospitals
3. Private medical organizations and hospitals

Machine Translation of Public Health Information

Multilingual Public Health Materials

1. The production of quality cross-lingual public health materials could be costly and time-consuming.
2. The global public health research community has an increasing interest in automatic translation of public health information.

Relevant Research

1. Kirchhoff et al. (2011) investigated the feasibility of translating health promotion materials with a freely available statistical machine translation website. The study only analyzed the translations of documents from English to Spanish.
2. Turner et al. (2015) conducted follow-up research to examine the performance of Google Translate in translating English health promotion documents into Chinese. There were problems such as errors of word sense and word order.

Traditional Approaches to Machine Translation

01	Rule-based Machine Translation	● I go to school by bus. ● 我／去／學校／乘／公車／。 ● 我乘公車去學校。
02	Example-based Machine Translation	● I go to school by bus. ● I go to school by taxi. 我乘計程車去學校。 ● 我乘公車去學校。
03	Statistical Machine Translation	● Translation Model *P*(“I”|“我”) x *P*(“go”|“去”) x ... ● Language Model: *P*(“我乘公車去學校”)

New Possibilities Offered by Neural Machine Translation

New Approach: Neural Machine Translation

1. Deep Learning: the use of multiple layers of artificial neurons
2. Automatic feature engineering: Converting the input into an internal representation that captures the features of the input by using deep neural networks
3. Areas of Application: Speech recognition, handwriting recognition, image processing, and machine translation

Neural Machine Translation

Neural machine translation often adopts the encoder-decoder architecture with recurrent neural networks (RNN) to model the translation process. The bidirectional RNN encoder which consists of a forward RNN and a backward RNN reads a source sentence $\mathbf{x} = x_1, x_2, ..., x_{T_x}$ and transforms it into word annotations of the entire source sentence $\mathbf{h} = h_1, h_2, ..., h_{T_x}$. The decoder uses the annotations to emit a target sentence $\mathbf{y} = y_1, y_2, ..., y_{T_y}$ in a word-by-word manner.

In the training phase, given a parallel sentence $(\mathbf{x}, \mathbf{y})$, NMT models the conditional probability as follows,

$$P(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{T_y} P(y_i|\mathbf{y}_{<\mathbf{i}}, \mathbf{x}) \quad (1)$$

where y_i is the target word emitted by the decoder at step i and $\mathbf{y}_{<\mathbf{i}} = y_1, y_2, ..., y_{i-1}$. The conditional probability $P(y_i|\mathbf{y}_{<\mathbf{i}}, \mathbf{x})$ is computed as

$$P(y_i|\mathbf{y}_{<\mathbf{i}}, \mathbf{x}) = softmax(f(s_i, y_{i-1}, c_i)) \quad (2)$$

where $f(\cdot)$ is a non-linear function and s_i is the hidden state of the decoder at step i:

$$s_i = g(s_{i-1}, y_{i-1}, c_i) \quad (3)$$

where $g(\cdot)$ is a non-linear function. Here we adopt Gated Recurrent Unit (Cho et al., 2014) as the recurrent unit for the encoder and decoder. c_i is the context vector, computed as a weighted sum of the annotations $\mathbf{h}$:

$$c_i = \sum_{j=1}^{T_x} \alpha_{t,j} h_j \quad (4)$$

where h_j is the annotation of source word x_j and its weight $\alpha_{t,j}$ is computed by the attention model.

We train the attention-based NMT model by maximizing the log-likelihood:

$$C(\theta) = \sum_{n=1}^{N} \sum_{i=1}^{T_y} \log P(y_i^n|\mathbf{y}_{<i}^n, \mathbf{x}^n) \quad (5)$$

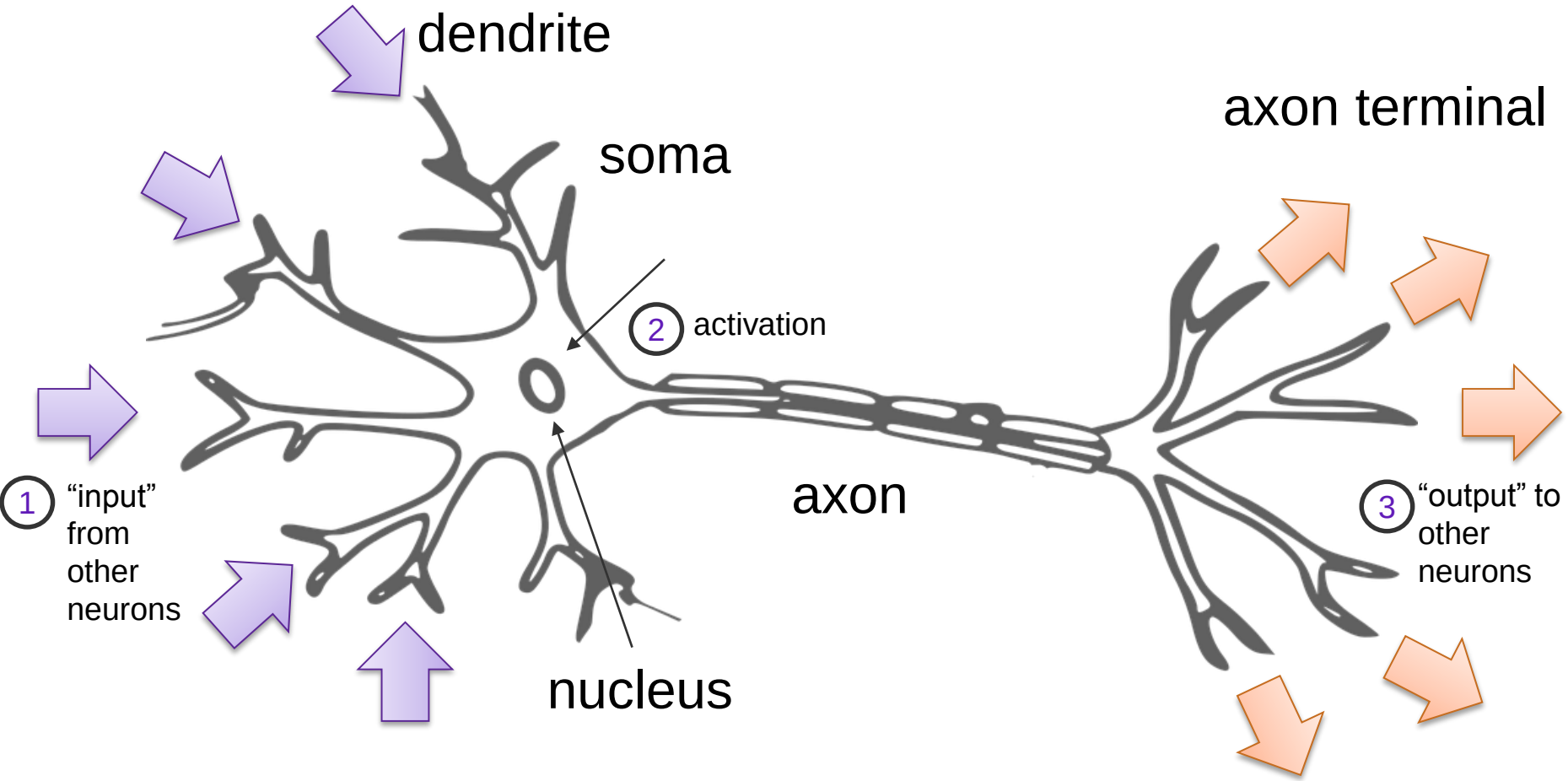
given the training data with N bilingual sentences (Cho, 2015).

In the testing phase, given a source sentence $\mathbf{x}$, we use beam search strategy to search a target sentence $\hat{\mathbf{y}}$ that approximately maximizes the conditional probability $P(\mathbf{y}|\mathbf{x})$

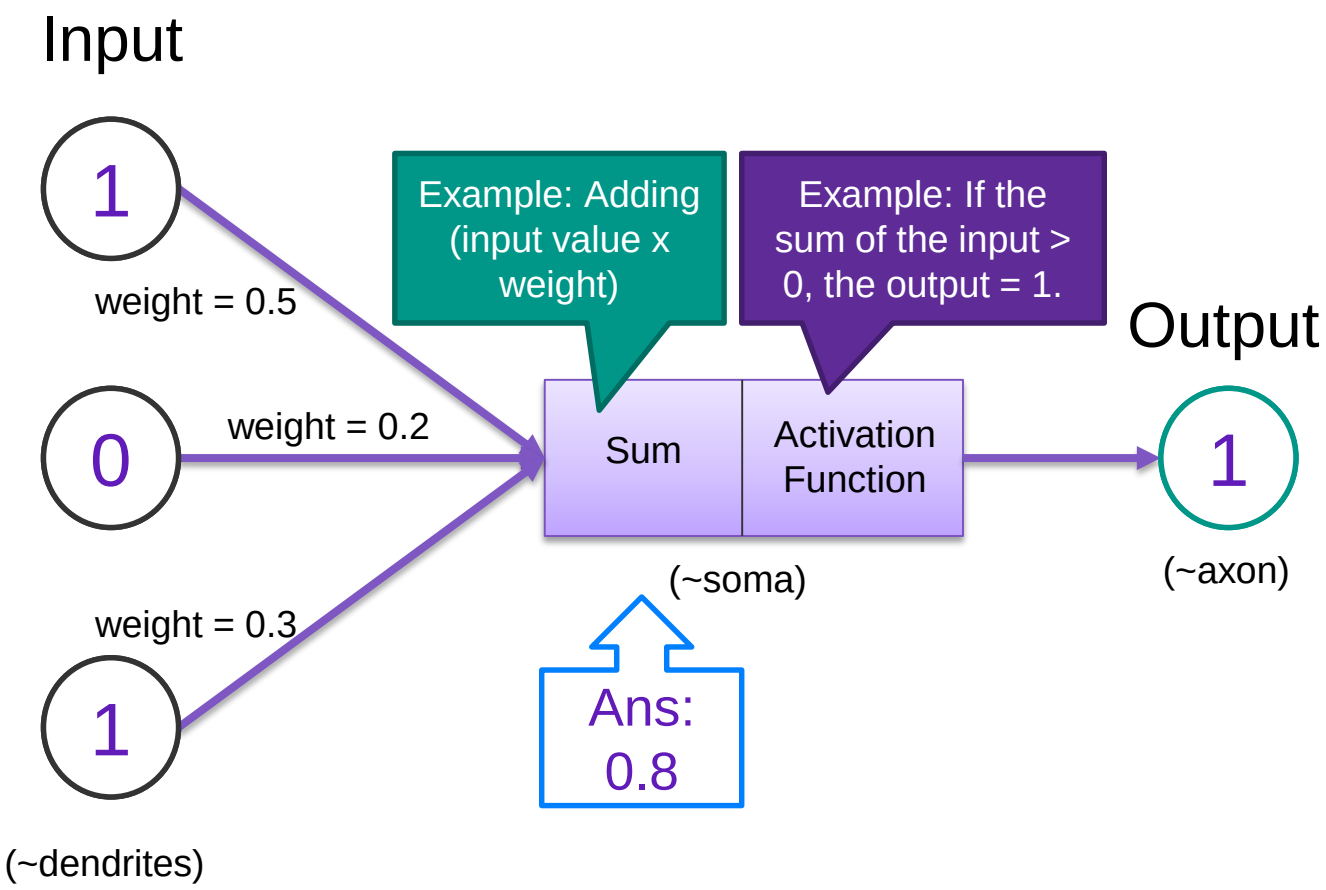
$$\hat{\mathbf{y}} = \underset{\mathbf{y}}{\operatorname{argmax}}\, P(\mathbf{y}|\mathbf{x}) \quad (6)$$

(http://aclweb.org/anthology/D17-1149)

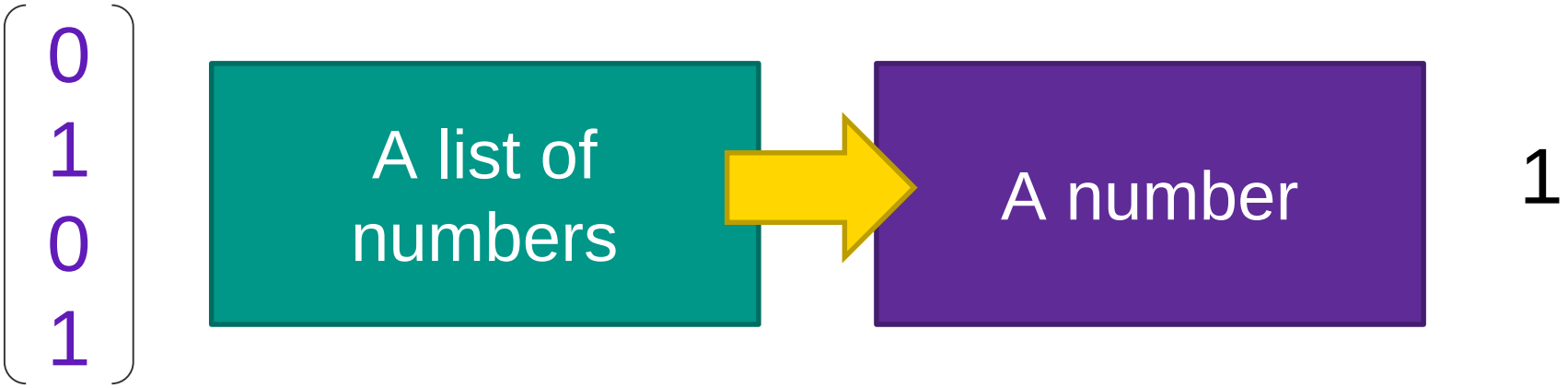
Biological Neuron



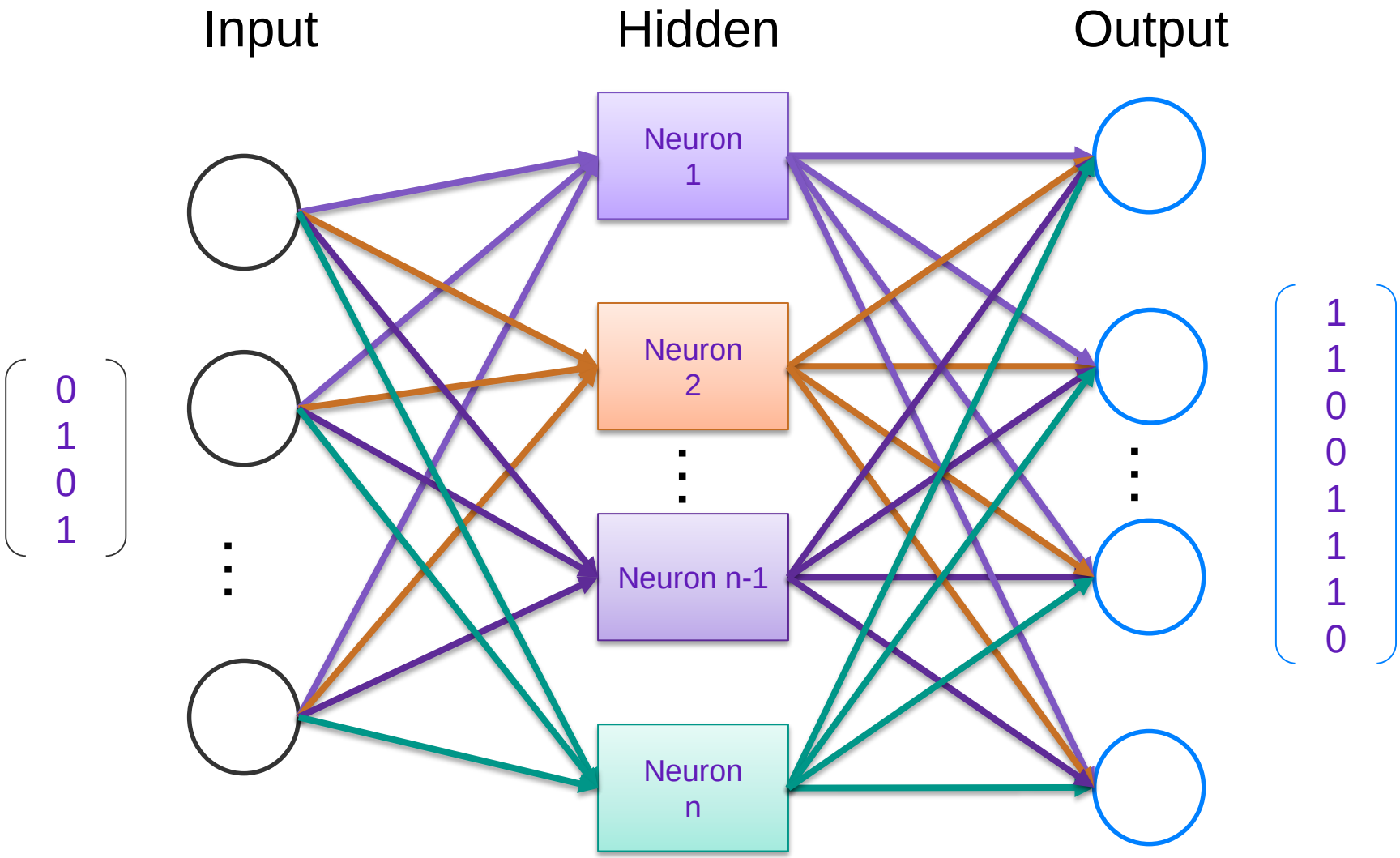
Artificial Neuron



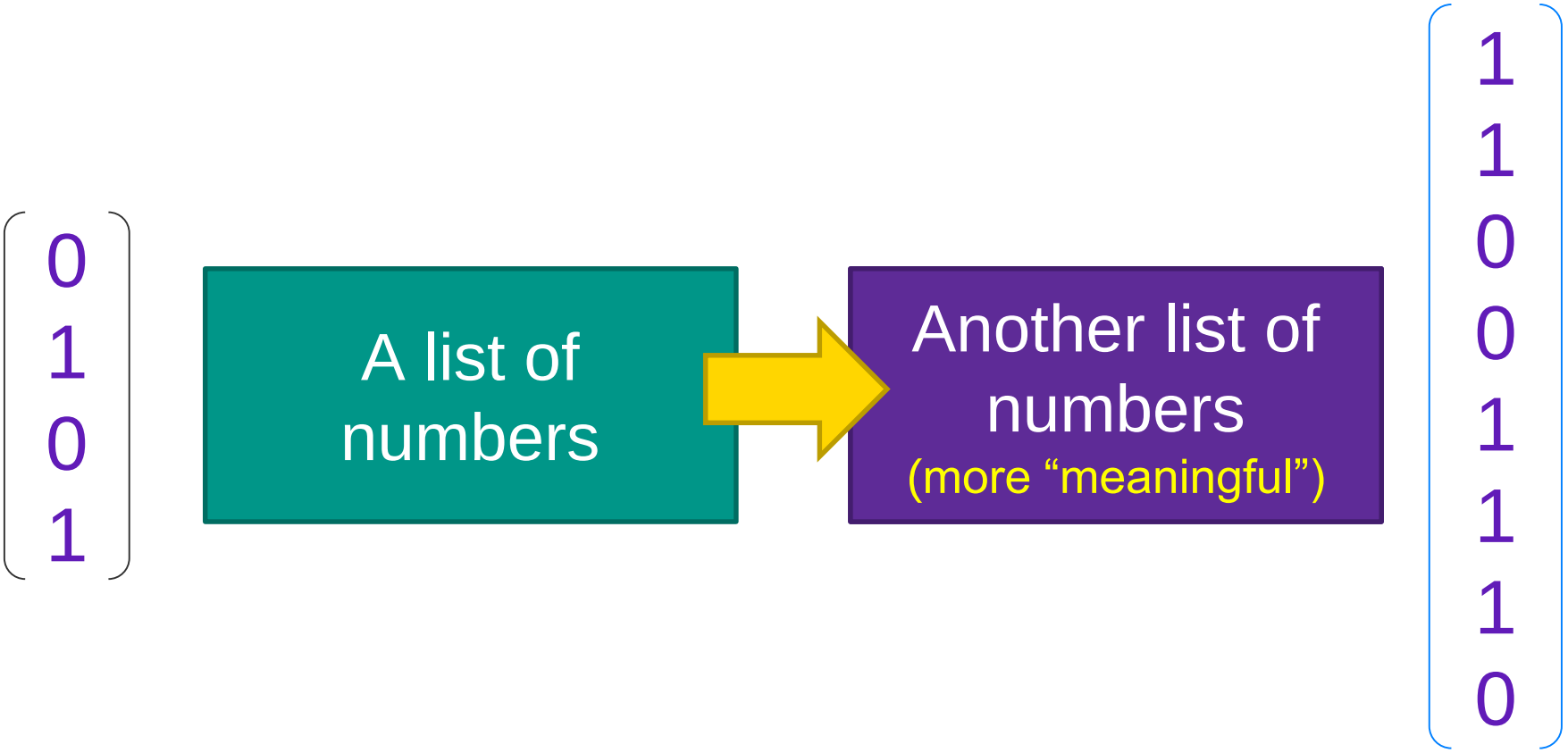
Artificial Neuron



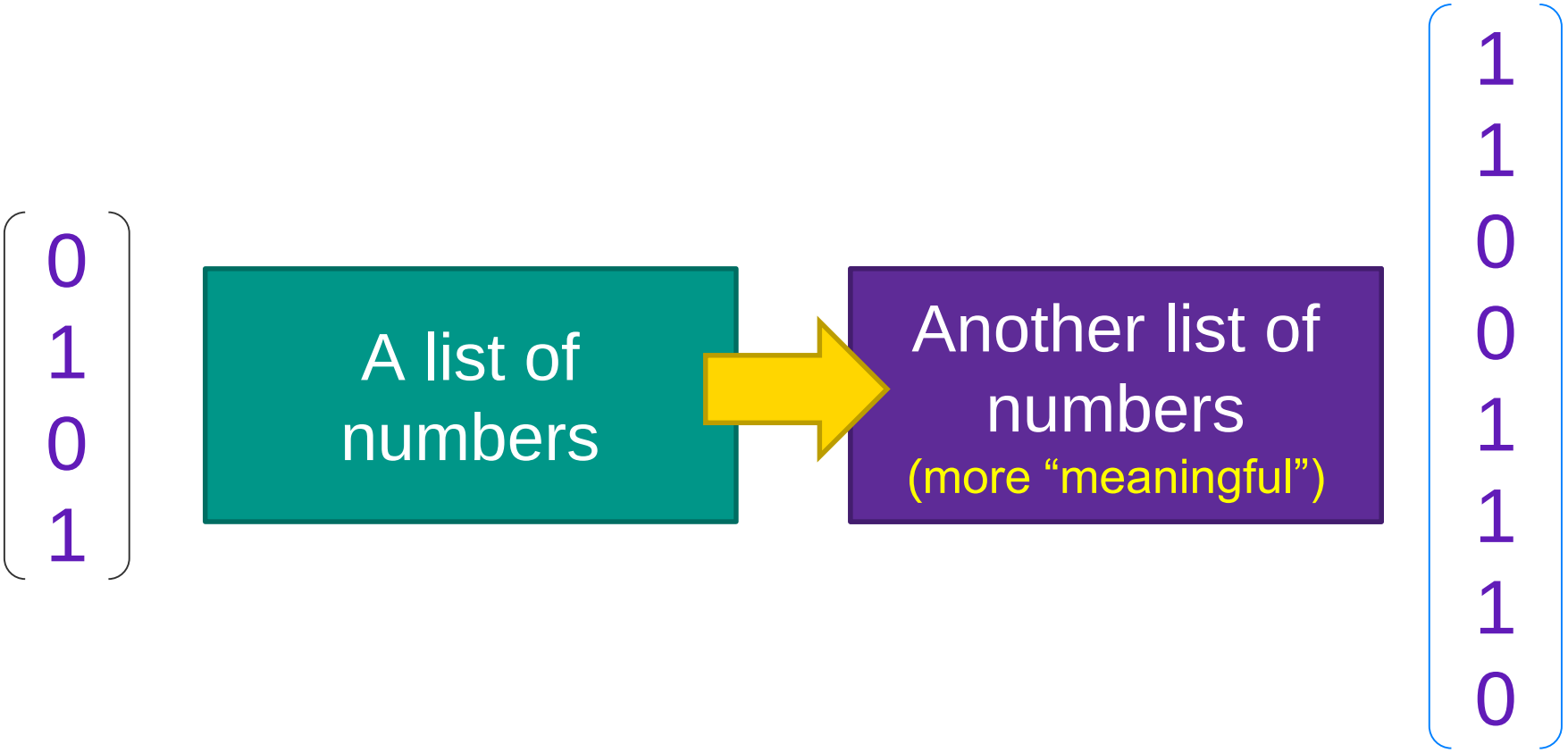
Artificial Neural Network: Many Neurons



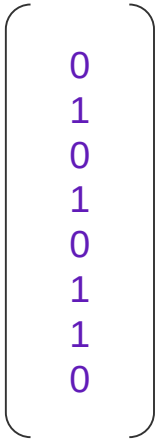
Artificial Neural Network

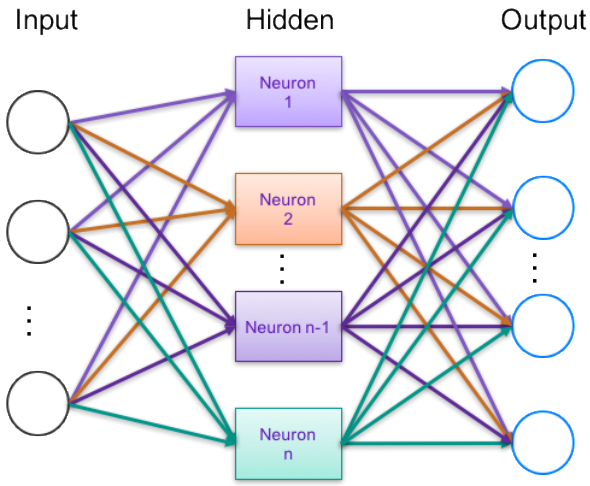


Artificial Neural Network



Artificial Neural Network and Image Processing





Probability	that the picture is a …
0.55	cat
0.300	tiger
.10	dog
0.05	man

Multi-class Classification

Neural Machine Translation

1 Data Collection

Bilingual documents are collected for the training of a neural machine translation system

2 Word Embedding

Each token in the source text is mapped on to a sequence of numbers

3 Encoding

The source text was encoded and converted to a sequence of vectors, an internal representation of the text

4 Decoding

The internal representation is decoded and the most probable tokens in the target language are identified

NMT as a Prediction Problem: Given X, what is the next word?

Given the following:

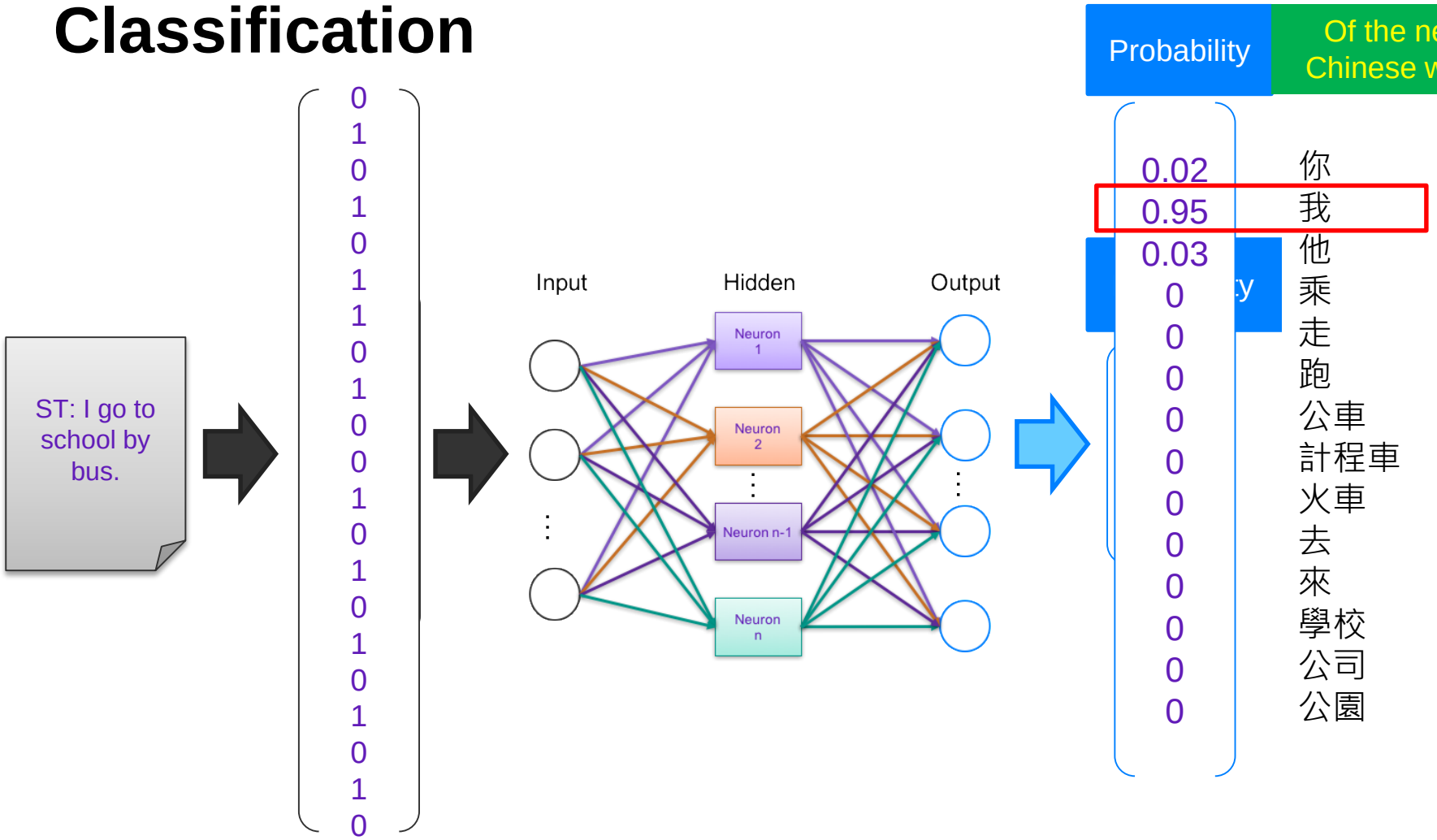
ST: I / go / to / school / by / bus / . /

TT: 我 / 乘

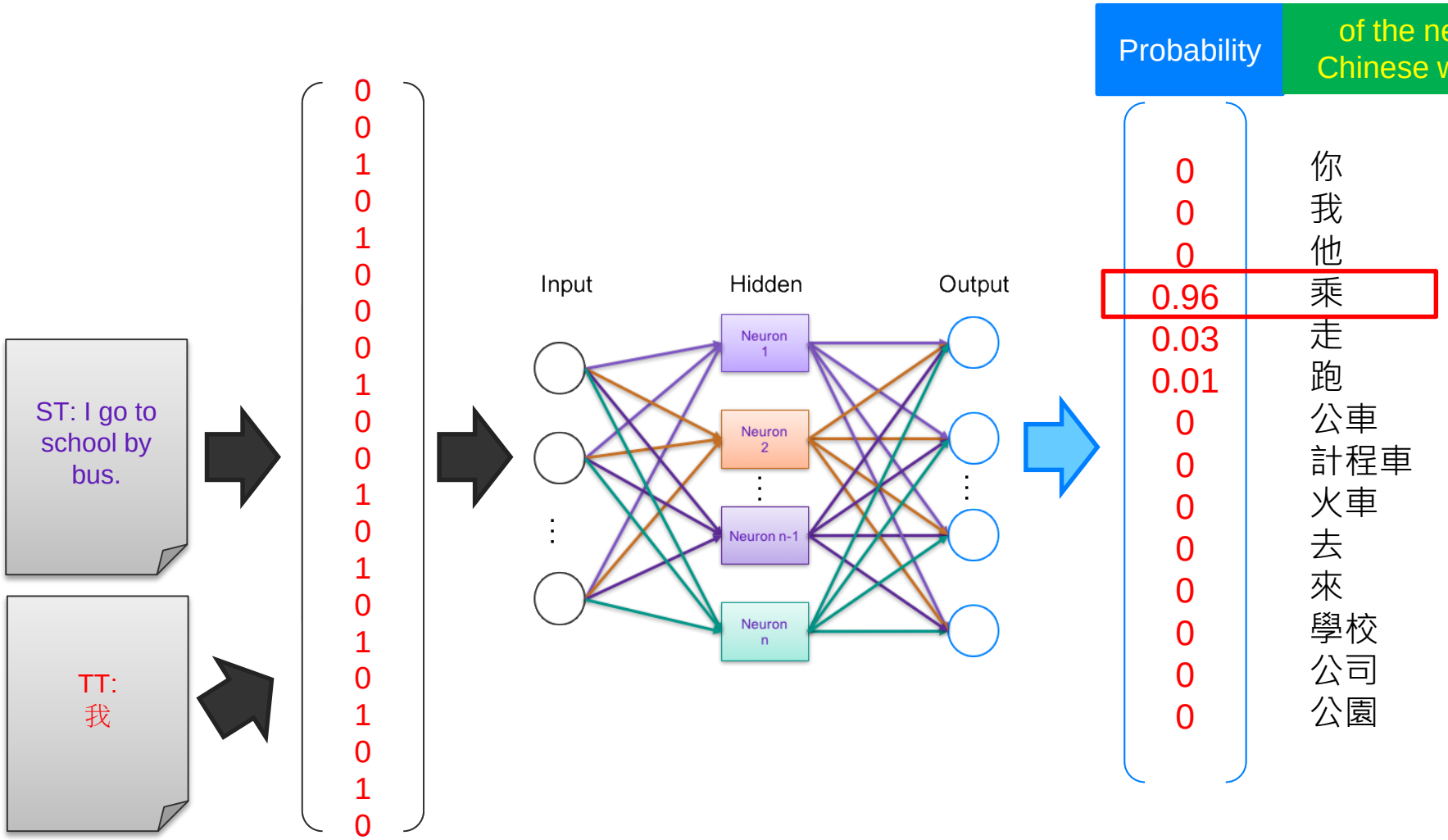
Which of the following is most likely to be the next word?

公車
火車
學校
咖啡
電腦
翻譯

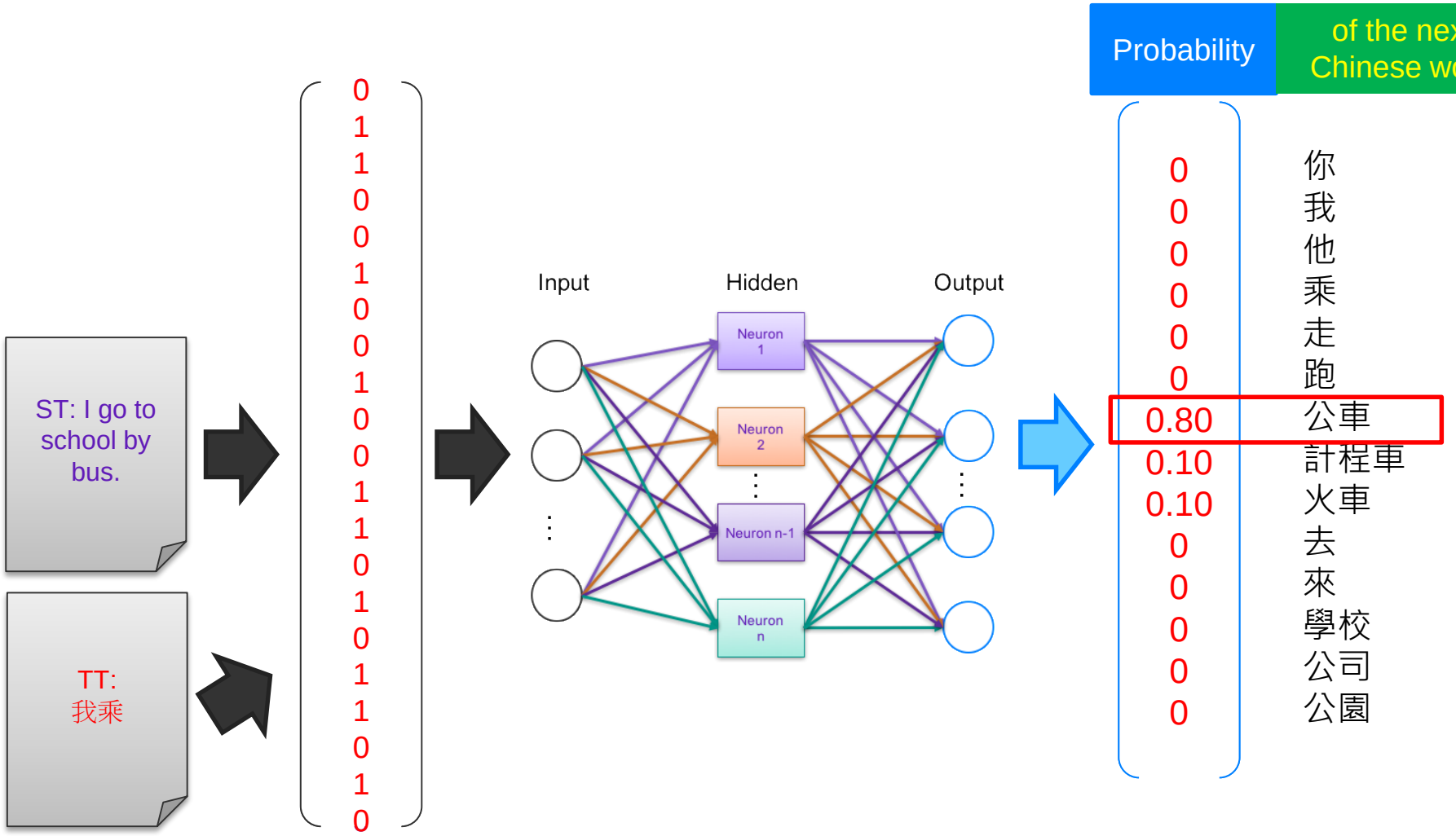
Neural Machine Translation as Multi-class Classification



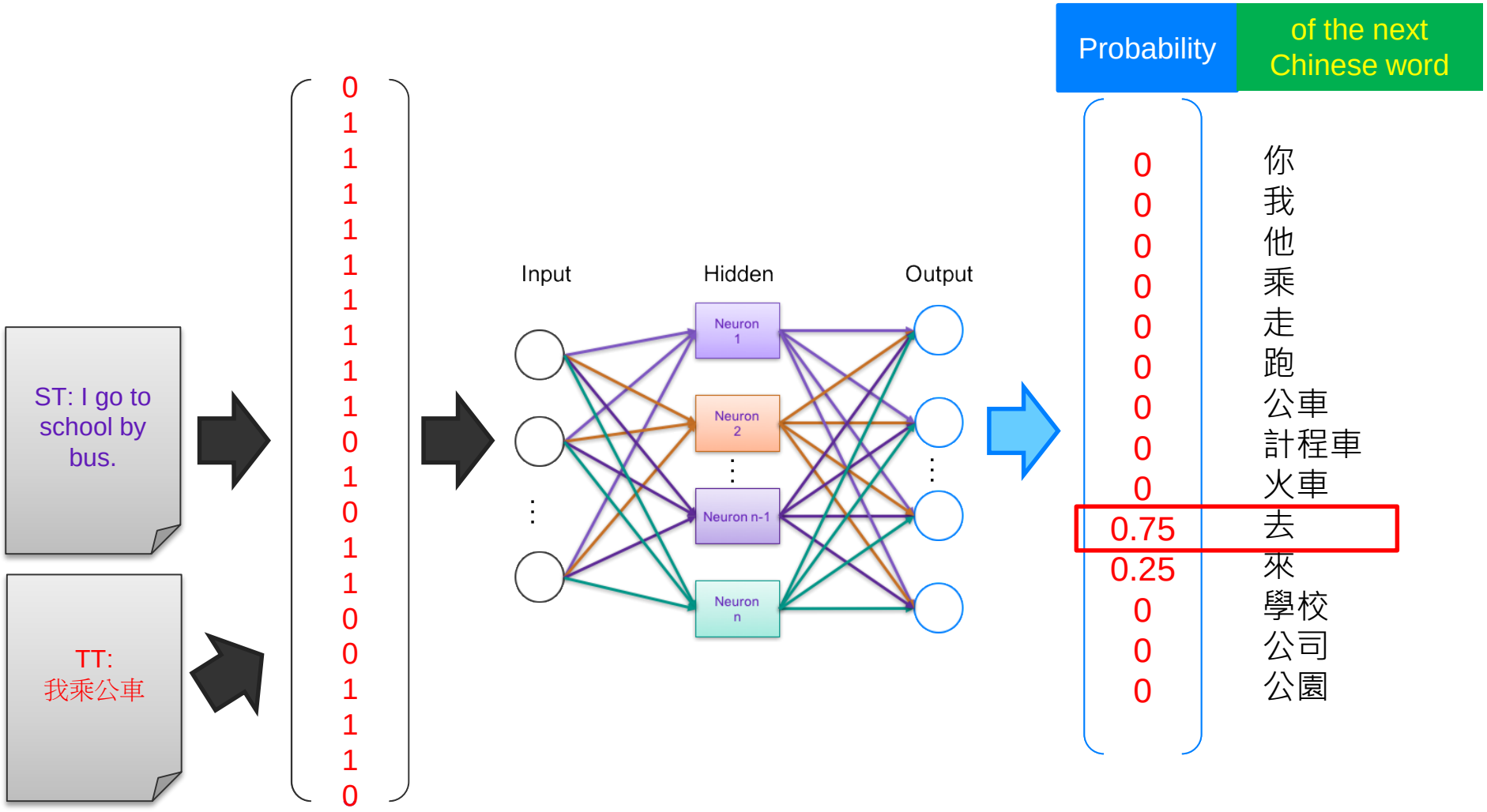
Neural Machine Translation



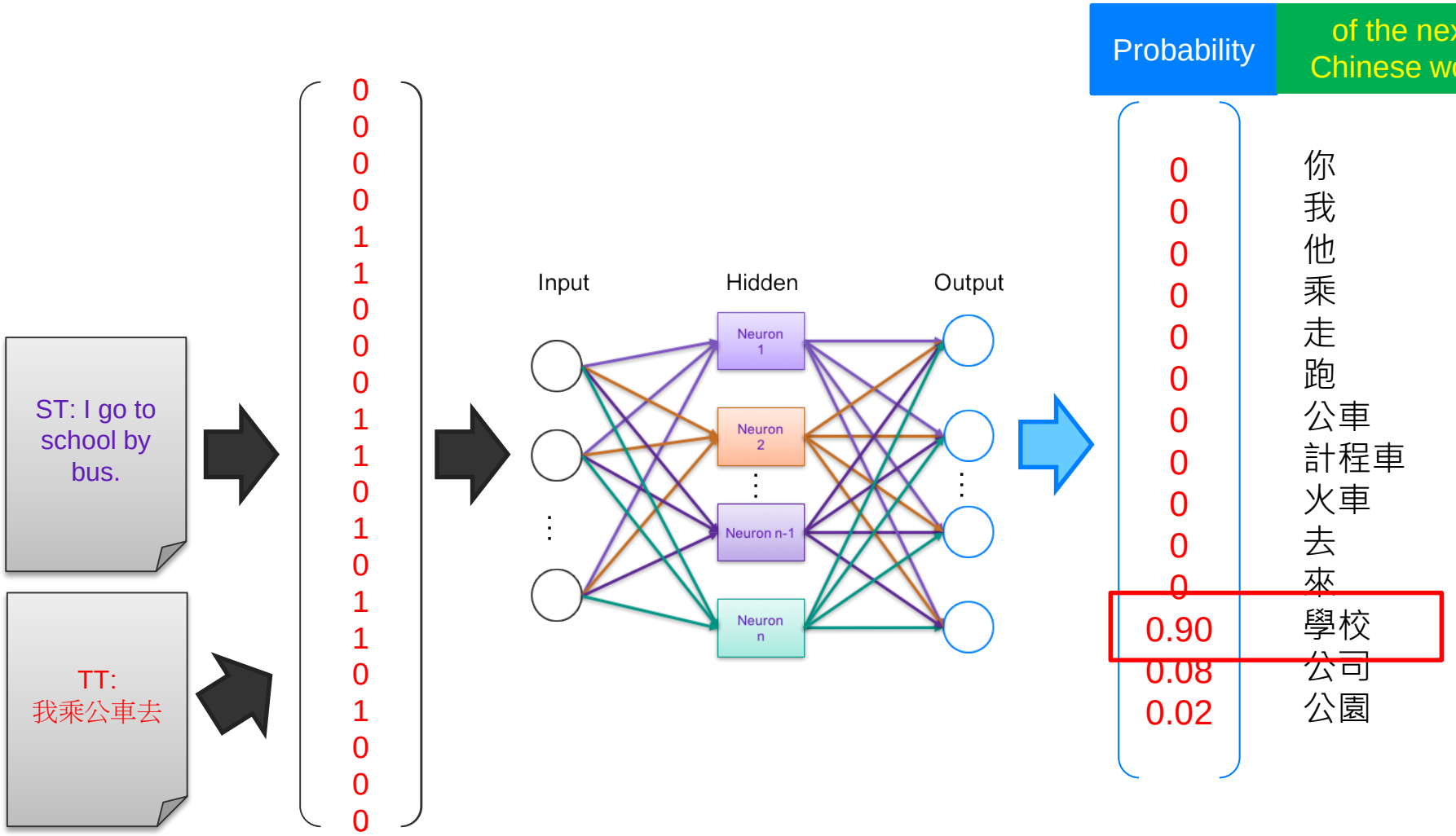
Neural Machine Translation



Neural Machine Translation



Neural Machine Translation



Types of Neural Machine Translation

1. Encoder-decoder (Recurrent Neural Networks using Long-short Term Memory Cells or Gated Recurrent Units) with Attention (Bahdanau, Cho, & Bengio, 2014; Luong, Pham, & Manning, 2015; Wu et al., 2016)
2. Convolutional Neural Networks (Gehring et al., 2017)
3. Self-attentional Transformers (Vaswani et al., 2017)

Examples (since 2016)

1. Google Neural Machine Translation
2. Microsoft Translator: AI Chinese-English machine translation system (news translation) that can translate with the same accuracy as a human
3. Tencent: Conference Interpreting (Simultaneous transcription and translation)
4. Sogou: WMT 2017 News Translation Task (1st in Chinese-English Translation and 3rd in English-Chinese Translation)

Evaluation

Press Releases

1. Bilingual press releases published by the Department of Health (see http://www.dh.gov.hk/english/press/press.html and http://www.dh.gov.hk/tc_chi/press/press.html)
2. Topics covered: updates on cases of infectious diseases, food poisoning clusters, preventive measures against diseases, follow-up investigations in response to detection of virus, identification of problematic cosmetic, medical or pharmaceutical products or devices, and health promotion activities

Source Texts for Evaluation

1. English press releases published by the Department of Health between January and March 2018 (20 press releases; 9,472 tokens)
2. Topics covered: Middle East Respiratory Syndrome (MERS), dengue fever, oral health, food poisoning, slimming products, influenza B infection, precautions against cold weather, Legionnaires' disease, undeclared Western drug ingredients in powder prescribed by a Chinese medicine practitioner, avian influenza, and acute gastroenteritis

Reference Translations for Comparison

The corresponding press releases in Chinese

Machine Translation Systems Evaluated

1. Google Translate: Supporting over 100 languages and offers text translation (neural machine translation for languages including Chinese), speech translation, image translation, and offline translation (Google, n.d.)
2. Microsoft Translator: Providing text translation for more than 60 languages (neural machine translation for 21 languages including Chinese), speech translation for 10 languages, and image translation (Microsoft Translator, 2017)

Machine Translation Systems Evaluated

3. Sogou Translator: The first commercial translation system that is fully based on neural machine translation in Mainland China and supports some 60 languages (Sogou, 2018b)

Automatic Evaluation

The use of Bilingual Evaluation Understudy (BLEU) metric (Papineni, Roukos, Ward, & Zhu, 2002), a measure of n-gram precision with a value between 0 and 1, comparing a machine translation output with a reference:

$$BLEU = \min(1, e^{1-\frac{r}{c}}) \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

Automatic Evaluation

We computed the following BLEU scores:

(1) Sentence-based BLEU scores for individual sentences in the target language using the "add-1" smoothing technique (i.e. adding 1 to the number of matched n-grams and total number of n-grams for $n \geq 2$; see Lin & Och (2004), as cited in "Smoothing 2" in Chen & Cherry (2014)) and

(2) Overall BLEU scores of all the translated sentences with the smoothing method that combines length-scaled pseudo-counts and match interpolation (see "Smoothing 7" in Chen & Cherry (2014)).

Automatic Evaluation

- Summary of sentence-based BLEU scores: Sogou Translator attained the highest average sentence-based BLEU score

	Google	Microsoft	Sogou
1st Qu.	0.0514	0.0426	0.0538
Median	0.0729	0.0613	0.0800
Mean	**0.0879**	**0.0762**	**0.1008**
3rd Qu.	0.1150	0.0931	0.1261
Max	0.3976	0.3593	0.4518

Automatic Evaluation

- Distribution of sentence-based BLEU scores: Sogou had a higher percentage of sentences with a score greater than 0.30

Sentence BLEU	Google (%)	Microsoft (%)	Sogou (%)
<0.05	24.27	35.73	20.22
0.05≤x<0.10	44.27	42.7	44.27
0.10≤x<0.15	19.55	12.81	19.33
0.15≤x<0.20	7.87	4.04	6.97
0.20≤x<0.25	1.35	1.8	3.15
0.25≤x<0.30	1.35	2.02	2.92
0.30≤x<0.35	**0.22**	**0**	**0.9**
0.35≤x<0.40	**1.12**	**0.9**	**1.8**
0.40≤x<0.45	**0**	**0**	**0**
≥0.45	**0**	**0**	**0.45**

- Overall BLEU scores: Sogou attained the highest overall BLEU score

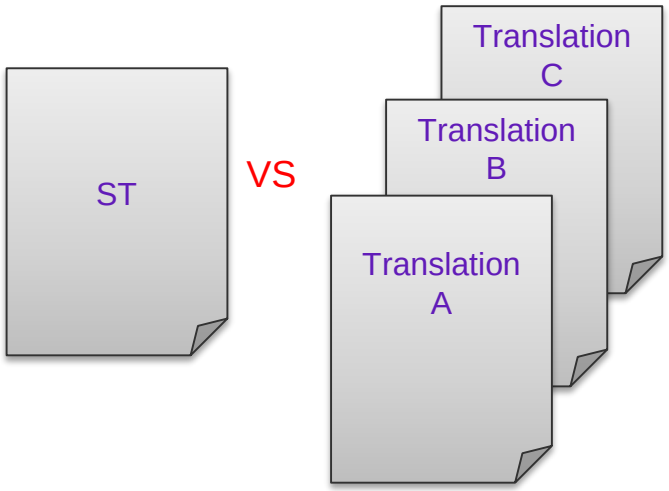
	Google	Microsoft	Sogou
Overall BLEU with smoothing	0.2430	0.2319	**0.2470**

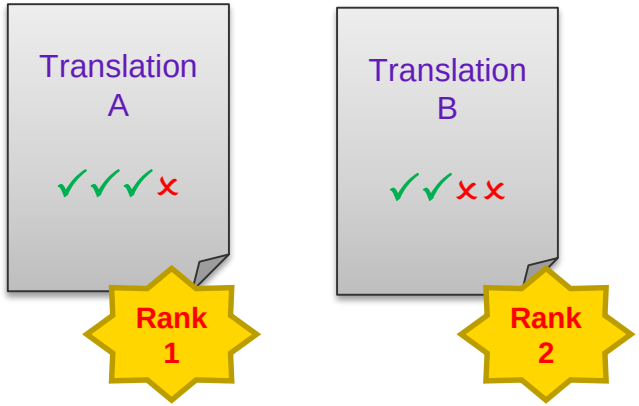
Human Evaluation

We conducted human evaluation by using 100 sentences randomly drawn from the dataset. We used the method of translation ranking (see, for example, Toral & Way (2018)), which has been a common way of conducting human assessment of automatic translation system in machine translation competitions in recent years.

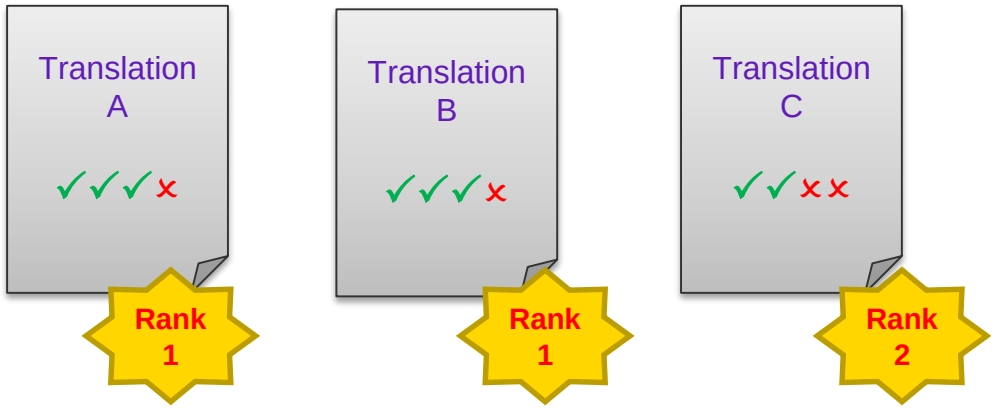
Human Evaluation

- Preservation of information?
- Number of errors?
- Naturalness?





OR



Result of Human Ranking

Sogou Translator ranked first in terms of the proportion of translations given rank 1 (Google Translate and Microsoft Translator ranked second and third respectively). As for the aggregate proportion of translations given ranks 2 or 3, Microsoft Translator ranked first, followed by Google Translate and Sogou Translator.

	Google (%)	Microsoft (%)	Sogou (%)
Rank 1	38	22	**55**
Rank 2	35	**46**	35
Rank 3	27	**32**	10
Total	100	100	100

Translation Issues and Suggestions

Readable outputs

The three systems sometimes provided readable outputs with little post-processing, especially translations with a BLEU score equal to or higher than 0.25.

- **Source Text**: Poultry should be cooked thoroughly.
- **Google Translate**: 家禽應徹底煮熟。

- **Source Text**: Investigations are ongoing.
- **Microsoft Translator**: 目前正在進行調查。

- **Source Text**: Read carefully the label instructions first;
- **Sogou Translator**: 先仔細閱讀標籤說明；

Issue 1: Abbreviations

- **Source Text:** Update on latest **MERS** situation in Oman
- **Reference:** 阿曼中東呼吸綜合症最新情況

- **Google Translate:** 更新阿曼最新的**MERS**情況
- **Microsoft Translator:** 阿曼**最新匯率情況(X)**的最新情況
- **Sogou Translator:** 阿曼**市場匯率(X)**最新情況

Issue 2: Proper nouns

- **Source Text:** People who have purchased the products may submit them to the **Drug Office** of the DH at Room 1856, **Wu Chung House**, 213 Queen's Road East, Wan Chai, during office hours for disposal.
- **Reference:** 已購買上述產品的市民可於辦公時間內將有關產品交予灣仔皇后大道東二一三號**胡忠大廈**一八五六室衞生署**藥物辦公室**銷毀。

- **Google Translate:** 購買產品的人士可於辦公時間將其送交灣仔皇后大道東213號胡忠大廈1856室衛生署**藥物辦事處(X)**處理。
- **Microsoft Translator:** 購買該等產品的人士, 可于辦公時間內, 在灣仔皇后大道東213號**五湧樓(X)**1856室向**弶鍍 ' p 禁毒處(X)**提交。
- **Sogou Translator:** 市民如購買該等藥物，可于辦公時間內，將該等藥物交回灣仔皇后大道東213號胡忠大廈1856室衛生署**毒品調查科(X)**處理。

Issue 3: Word sense

- **Source Text:** He **attended** the Accident and Emergency Department (AED) of Alice Ho Miu Ling Nethersole Hospital yesterday (March 7) and was **admitted** today.
- **Reference:** 他於昨日（三月七日）**前往**雅麗氏何妙齡那打素醫院急症室**求診**，並於今日**入院**。

- **Google Translate:** 他於昨日（三月七日）**出席(X)**雅麗氏何妙齡那打素醫院急症室，並於今日**受理(X)**。
- **Microsoft Translator:** 他昨日 (3月7日) **出席(X)愛麗絲(X)**何妙齡那打素醫院急症室 (AED), 並於今日入院。
- **Sogou Translator:** 他昨日(三月七日)前往雅麗氏何妙齡那打素醫院急症室求診，並於今日入院。

Issue 4: Numbers

- **Source Text:** According to the latest information, 2 144 cases have been reported to the WHO (with at least 750 deaths), including **1 927 in 10 Middle East countries comprising 1 769 in the Kingdom of Saudi Arabia, 86 in the United Arab Emirates**, 28 in Jordan, 19 in Qatar, 11 in Oman, six in Iran, four in Kuwait, two in Lebanon, and one each in Yemen and Bahrain.
- **Reference:** 根據最新資料，世衛至今獲通報2 144宗個案，至少750人死亡。當中10個中東國家確診共1927宗，**其中沙特阿拉伯王國有1 769宗，阿拉伯聯合酋長國86宗，約旦28宗**，卡塔爾19宗，阿曼11宗，伊朗六宗，科威特四宗，黎巴嫩兩宗，以及也門和巴林各一宗。

- **Google Translate:** 根據最新資料，世衛組織報告了2 144例（至少750人死亡），其中包括10個中東國家的1 927例，其中沙特阿拉伯王國1 769例，阿拉伯聯合酋長國86例，**阿拉伯聯合酋長國28例，約旦(X)**，卡塔爾19人，阿曼11人，伊朗6人，科威特4人，黎巴嫩2人，也門和巴林各1人。
- **Microsoft Translator:** 根據最新資料, 已向衛生組織報告了 2 144 起案件 (至少有750人死亡), **包括927個中東國家的 1 10 個(X)**, 沙烏地阿拉伯王國 1 769 個, 阿拉伯聯合大公國86個, 28 個在約旦, 19 個在卡塔爾。, **11 在阿曼, 六在伊朗(X)**, 四人在科威特, 兩個在黎巴嫩, 一個在葉門和巴林。
- **Sogou Translator :** 根據最新資料，向世衛組織報告了2 144例病例(至少750人死亡)，其中10個中東國家1 927例，其中沙烏地阿拉伯王國1 769例，阿拉伯聯合大公國86例，約旦28例，卡塔爾19例，阿曼11例，伊朗6例，科威特4例，黎巴嫩2例 ，葉門和巴林各1例。

Issue 5: Sentence structure and word order

- **Source Text:** The second case is a 6-year-old girl **with good past health**.
- **Reference:** 第二宗個案涉及一名過往健康良好的六歲女童。

- **Google Translate:** 第二種情況是一名身體健康的(X)6歲女孩。
- **Microsoft Translator:**第二個案件是一個6歲的女孩與良好的過去健康(X)。
- **Sogou Translator:** 第二個病例是一名6歲女孩，過去健康情況良好。

Issue 6: Punctuation marks

- **Source Text:** Do not try to use **salt, vinegar, wine and wasabi** to kill bacteria as they are not effective; and (Department of Health, 2018o)
- **Reference:** 不要嘗試使用**鹽、醋、酒及日本芥末**殺菌，因為均沒有殺菌效用；及(Department of Health, 2018p)

- **Google Translate:** 不要嘗試用**鹽，醋，酒和芥末(X)**來殺死細菌，因為它們沒有效果;和
- **Microsoft Translator:** 不要試圖使用**鹽, 醋, 葡萄酒和芥末(X)**殺死細菌, 因為它們是不有效的;和
- **Sogou Translator:** 不要嘗試使用**鹽、醋、酒和芥末**來殺滅細菌，因為它們沒有效果；和

Issue 7: Omission or incomplete translation

- **Source Text:** As of yesterday (March 8), 11 cases had been confirmed this year, all of which were imported cases. The cases were imported from **the Philippines (five), Cambodia (two), India (one), Indonesia (one), Singapore (one) and Thailand (one)**.
- **Reference:** 截至昨日（三月八日），今年共錄得11宗個案，均為外地傳入，分別從**菲律賓（五宗）、柬埔寨（兩宗）、印度（一宗）、印尼（一宗）、新加坡（一宗）和泰國（一宗）傳入**。

- **Google Translate:** 截至昨日（3月8日），今年已確認11起，均為外來案件。**這些病例是從菲律賓（五），柬埔寨（二），印度（一），印度尼西亞（一），新加坡（一）和泰國（一）進口的(X)**。
- **Microsoft Translator:** 截至昨日 (3月8日), 今年已有11宗個案獲證實, 全部均為進口個案。**案件從菲律賓 (五)、柬埔寨 (二)、印度 (一)、印尼 (一)、新加坡 (一) 和泰國 (一) 進口(X)**。
- **Sogou Translator:** 截至昨日(三月八日)，本年共證實十一宗個案，全部為外地傳入個案。 **這些病例來自菲律賓(5例)、柬埔寨(2例)、印度(1例)、印尼(1例)、新加坡(1例)和泰國(1例)**。

Issue 8: Problematic addition

- **Source Text:** The CHP has also stepped up surveillance for possible human cases. **Letters will be sent** to doctors and hospitals today urging them to report any suspected cases.
- **Reference:** 中心已就可能出現的人類感染個案加強監測，今日**亦會發信**予醫生及醫院呼籲他們呈報可疑個案。

- **Google Translate:** 衛生防護中心也加強了對可能的人類病例的監測。信件將在今天發送給醫生和醫院，敦促他們報告任何疑似病例。
- **Microsoft Translator:** 此外, 生防護中心亦加強監察可能發生的人類個案。**警方(X)**今日 **(星期五)(X)** 將信件送交醫生及醫院, 促請他們舉報懷疑個案。
- **Sogou Translator:** 生防護中心亦加強監察可能出現的人類個案。 **警方(X)**今日會發信給醫生及醫院，促請他們舉報懷疑個案。

Issue 9: Wrong characters

- **Source Text:** **DH** holds World Oral Health Day publicity event (with photo)
- **Reference:** 衞生署舉辦世界口腔健康日推廣活動（附圖）

- **Google Translate:** 衞生署舉行世界口腔健康日宣傳活動（附圖）
- **Microsoft Translator:** 弼鍍 ' p (X)舉行世界口腔健康日宣傳活動 (帶照片)
- **Sogou Translator:** 生署(X)舉辦世界口腔健康日宣傳活動(附圖)

Issue 10: Collocation

- **Source Text:** They should **seek medical advice** and avoid school till 48 hours after **the fever has subsided**.
- **Reference:** 他們應求診及避免上學，直至退燒後四十八小時。

- **Google Translate:** 他們應該尋求醫療建議，並在發燒平息(X)後48小時內避免學校。
- **Microsoft Translator:** 在發燒消退48小時後, 他們應該尋求醫療建議, 避免上學。
- **Sogou Translator:** 他們應尋求醫療諮詢，避免上學，直到發燒消退後48小時。

Concluding Remarks

Application of NMT technology

- Google Translate, Microsoft Translator and Sogou Translator were sometimes able to generate readable and publishable results
- Sogou Translator tended to get higher scores in our automatic and manual tests
- Common translation issues: pre-editing and post-editing

Future work

- Expanding our testing datasets
- Deriving pre-editing and post-editing guidelines
- Considering other forms of public health information
- Increasing the number of machine translation engines to be evaluated

Deep Learning Research and Application Centre

Hang Seng Management College

Deep Learning Research and Application Centre

- Funding of HK$6.8 million (~US$890,000)
- Hardware: Graphics processing unit (GPU)-accelerated hardware for much faster training and decoding
- Software: PyTorch, Tensorflow, Keras, Transformer…
- Projects involving deep neural networks

Projects

- Machine translation (e.g., automatic translation of listing documents; recent funding: HK$1.4 million (~US$178,000))
- Text mining, text summarization and information retrieval (e.g., clinical data or medical articles)
- Classification and image analysis (e.g., more accurate classification of medical images for computer-aided medical diagnosis)
- Prediction (e.g., epidemic outbreaks and stock prices)

References

Please refer to the proceedings.

Thank you

scsiu@hsmc.edu.hk
saicheong.com

More about Artificial Neural Networks

Neural networks and back-propagation explained in a simple way (https://medium.com/datathings/neural-networks-and-backpropagation-explained-in-a-simple-way-f540a3611f5e)

More about Artificial Neural Networks

Back-Propagation is very simple. Who made it Complicated? (https://medium.com/@14prakash/back-propagation-is-very-simple-who-made-it-complicated-97b794c97e5c)

More about BLEU

$$BLEU = \min(1, e^{1-\frac{r}{c}}) \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

Given T = MT output and T_R = reference translation,

$$Precision = \frac{C(T \cap T_R)}{|T|}$$

Example

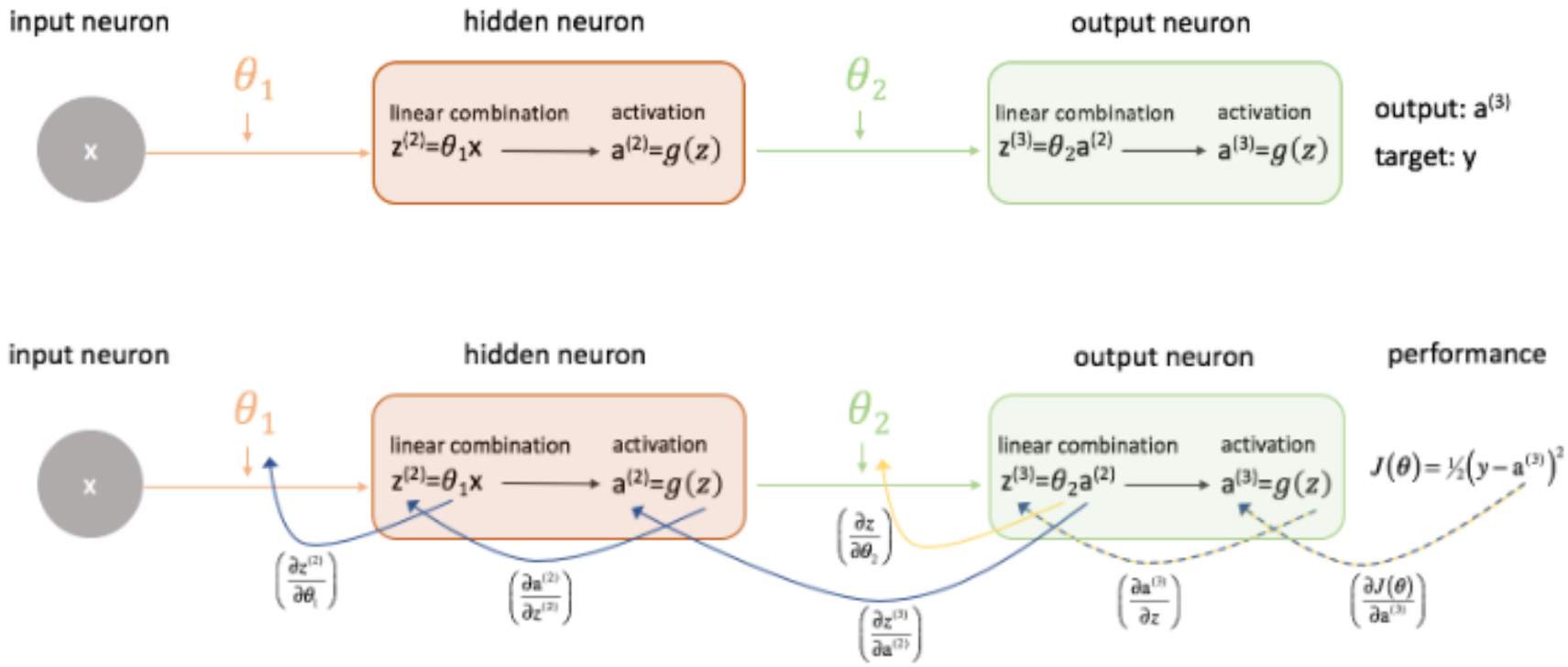
T = "Chinese officials responsibility of airport safety"

T_R = "Chinese officials are responsible for airport security"

$$Precision = \frac{3}{6} = 0.5$$

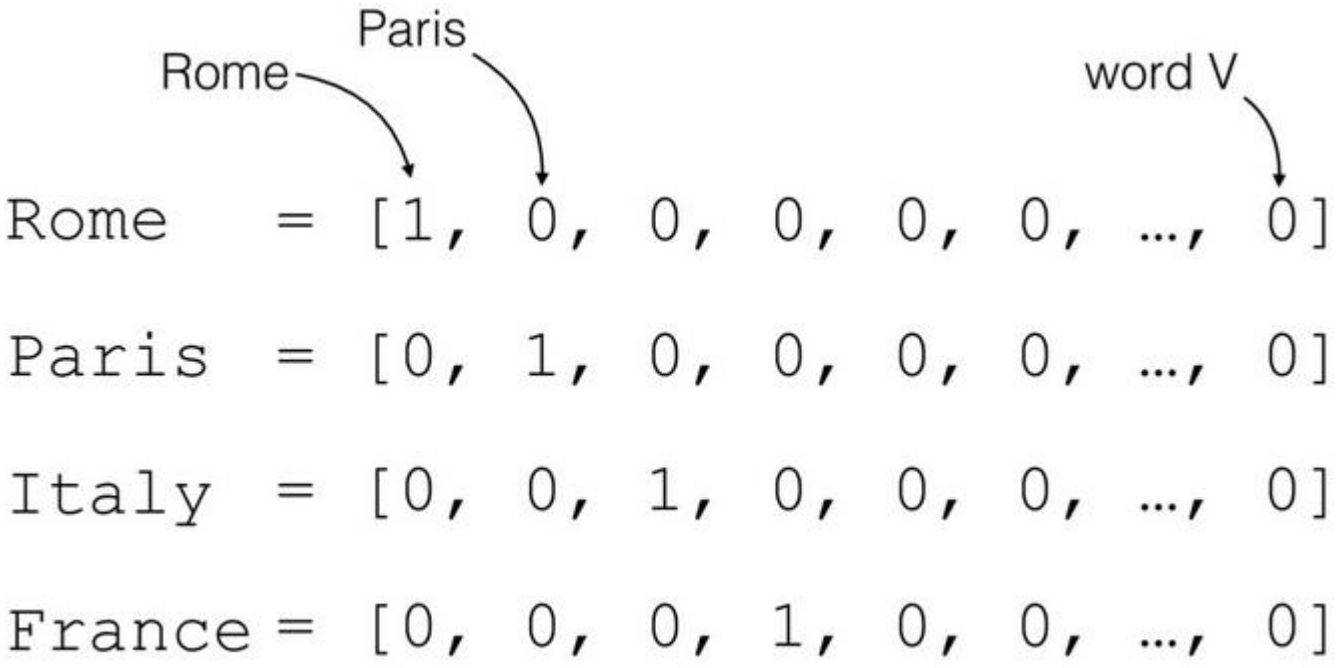
For more information, visit
https://www.aclweb.org/anthology/P02-1040.pdf

Training and Backpropagation



For more information, visit
https://www.jeremyjordan.me/neural-networks-training/

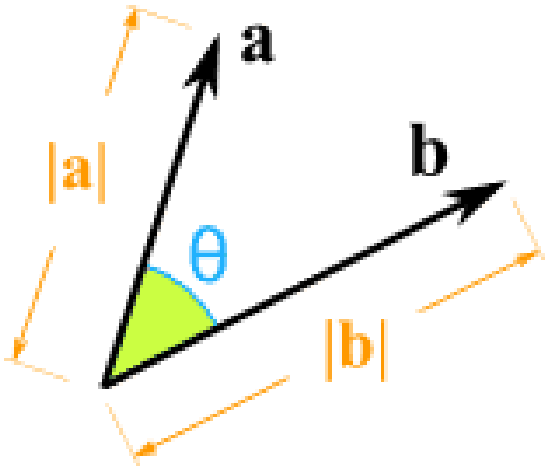
Word Embedding (1): One-hot Vector

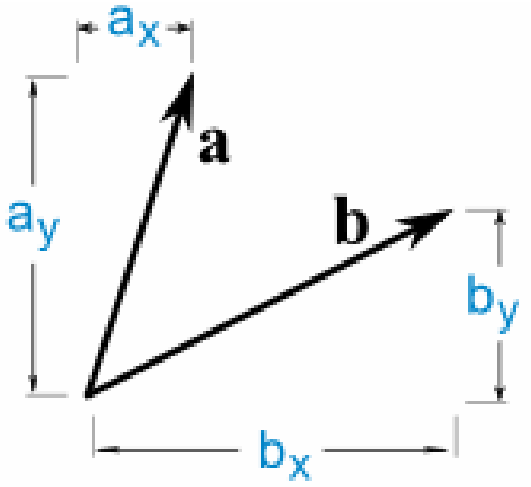


For more information, visit
https://medium.com/@athif.shaffy/one-hot-encoding-of-text-b69124bef0a7

Problem

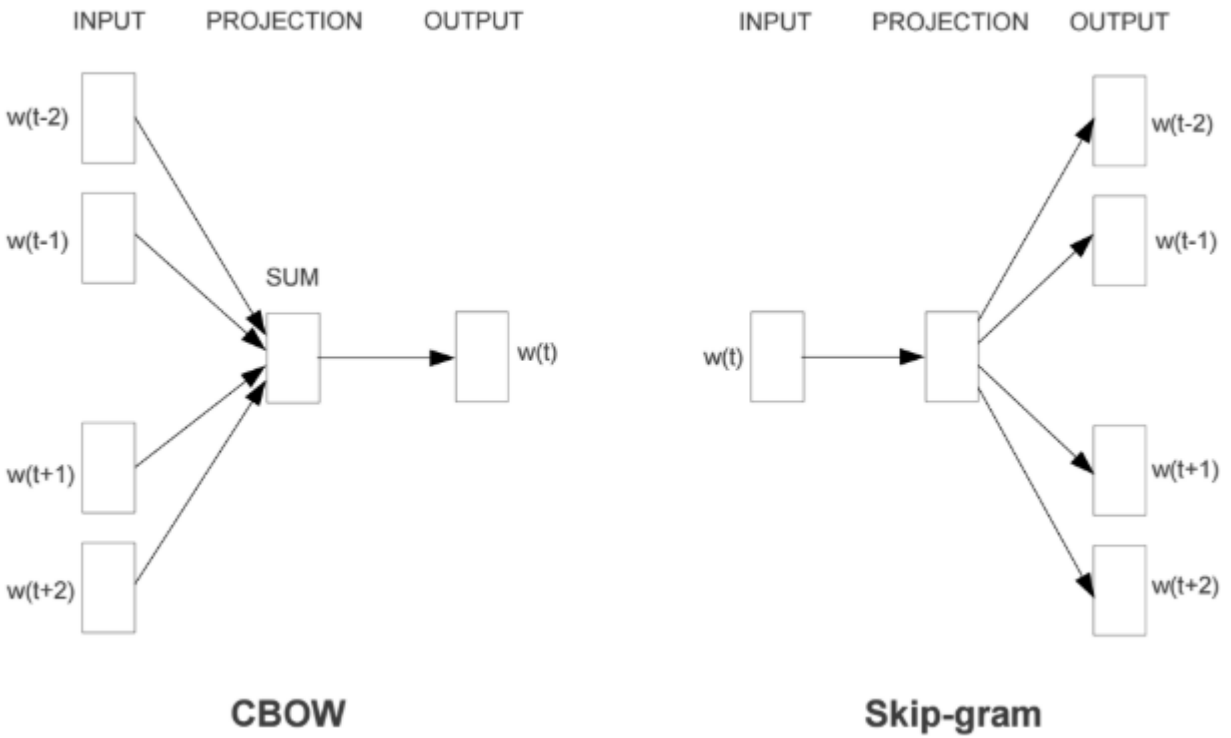
$$a \cdot b = a_x \times b_x + a_y \times b_y = |a||b|\cos(\theta)$$





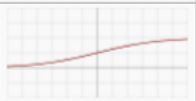
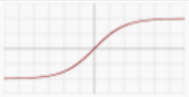
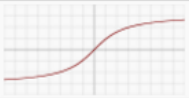
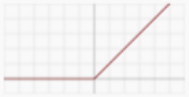
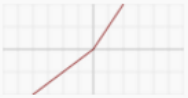
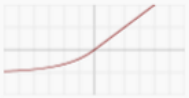
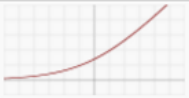
For more information, visit
https://www.mathsisfun.com/algebra/vectors-dot-product.html

Word Embedding (2): Continuous Bag of Words and Skip-gram



For more information, visit
https://mubaris.com/2017/12/14/word2vec/

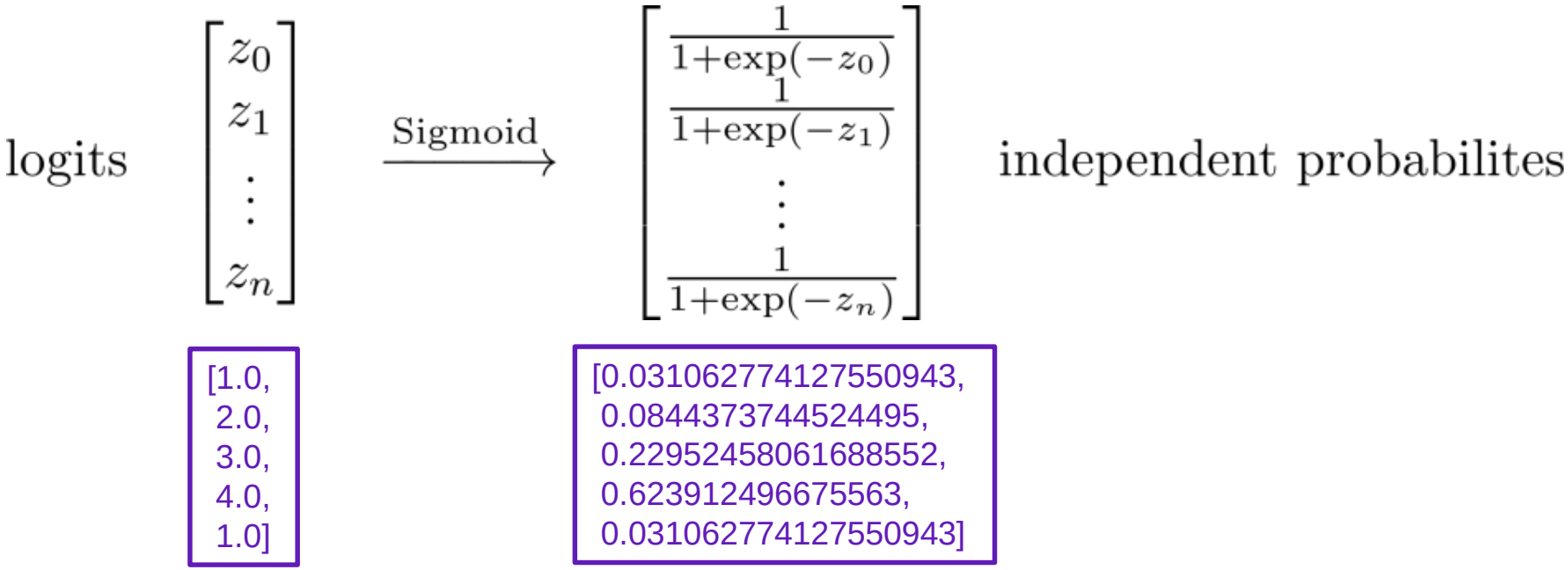
Activation functions

Name	Plot	Equation	Derivative
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$
Logistic (a.k.a Soft step)		$f(x) = \frac{1}{1+e^{-x}}$	$f'(x) = f(x)(1-f(x))$
TanH		$f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1$	$f'(x) = 1 - f(x)^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2+1}$
Rectified Linear Unit (ReLU)		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Parameteric Rectified Linear Unit (PReLU) [2]		$f(x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Exponential Linear Unit (ELU) [3]		$f(x) = \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} f(x) + \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
SoftPlus		$f(x) = \log_e(1+e^x)$	$f'(x) = \frac{1}{1+e^{-x}}$

For more information, visit
https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

Softmax for Multi-class Classification (1)

The softmax function is a generalization of the logistic function that "squashes" a K-dimensional vector **z** of arbitrary real values to a K-dimensional vector of real values in the range [0, 1] that add up to 1.



For more information, visit
https://www.depends-on-the-definition.com/guide-to-multi-label-classification-with-neural-networks/

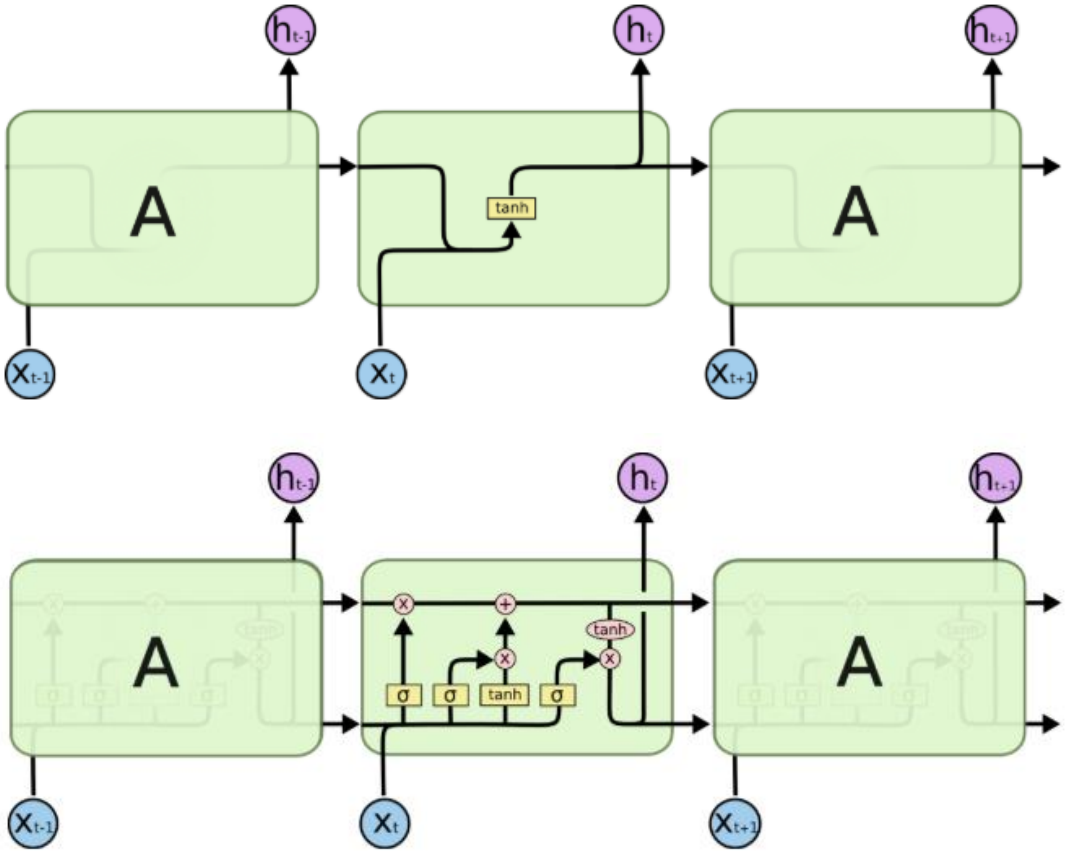
Softmax for Multi-class Classification (2)

Softmax function takes an N-dimensional vector of real numbers and transforms it into a vector of real number in range (0,1) which add up to 1.

$$
\begin{aligned}
p_j &= \frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}} \\
&= \frac{Ce^{a_i}}{C\sum_{k=1}^{N} e^{a_k}} \\
&= \frac{e^{a_i+\log(C)}}{\sum_{k=1}^{N} e^{a_k+\log(C)}}
\end{aligned}
$$

For more information, visit
https://deepnotes.io/softmax-crossentropy

Standard RNN VS Long Short-term Memory



For more information, visit
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

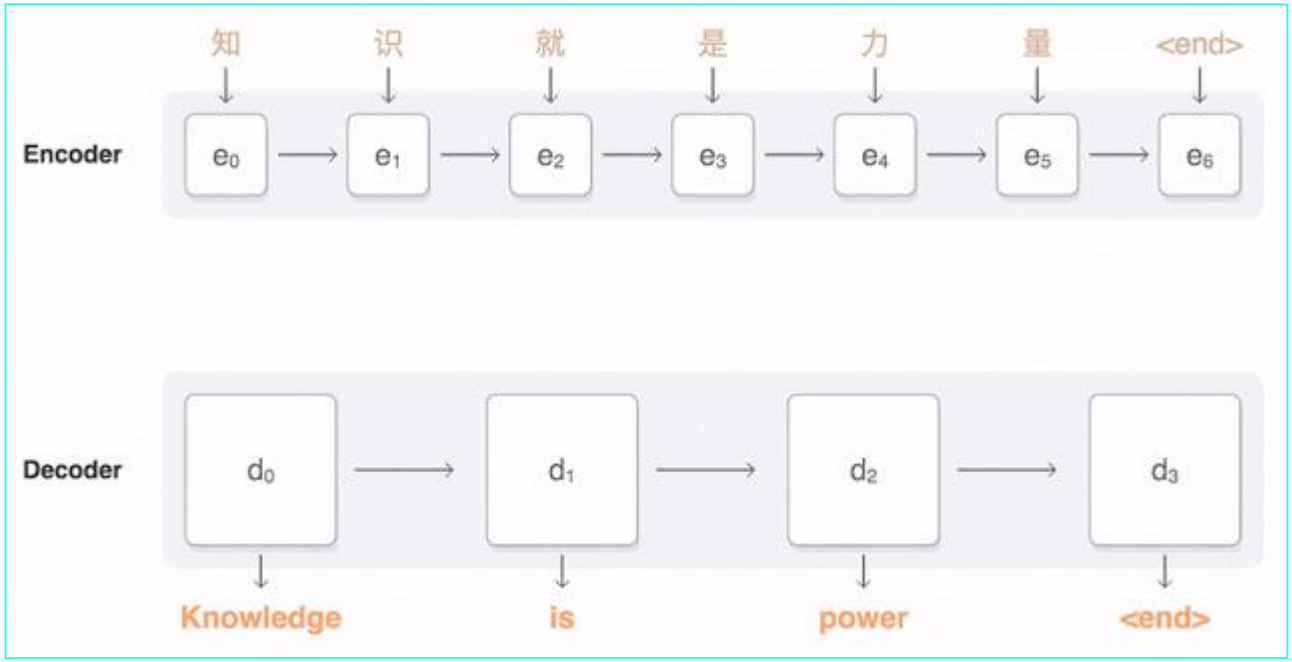
Attention Mechanism

Given **[I]** / [go] / [to] / [school] / [by] / [bus] / [.], what is the next word?
我
[I] / [go] / [to] / [school] / **[by]** / [bus] / [.] / [我] ⇒ ?
乘
[I] / [go] / [to] / [school] / [by] / **[bus]** / [.] / [我] / [乘] ⇒ ?
公車
[I] / **[go]** / **[to]** / [school] / [by] / [bus] / [.] / [我] / [乘] / [公車] ⇒ ?
去
[I] / [go] / [to] / **[school]** / [by] / [bus] / [.] / [我] / [乘] / [公車] / [去] ⇒ ?
學校
[I] / [go] / [to] / [school] / [by] / [bus] / **[.]** / [我] / [乘] / [公車] / [去] / [學校] / ⇒ ?
。

Translation: 我乘公車去學校。

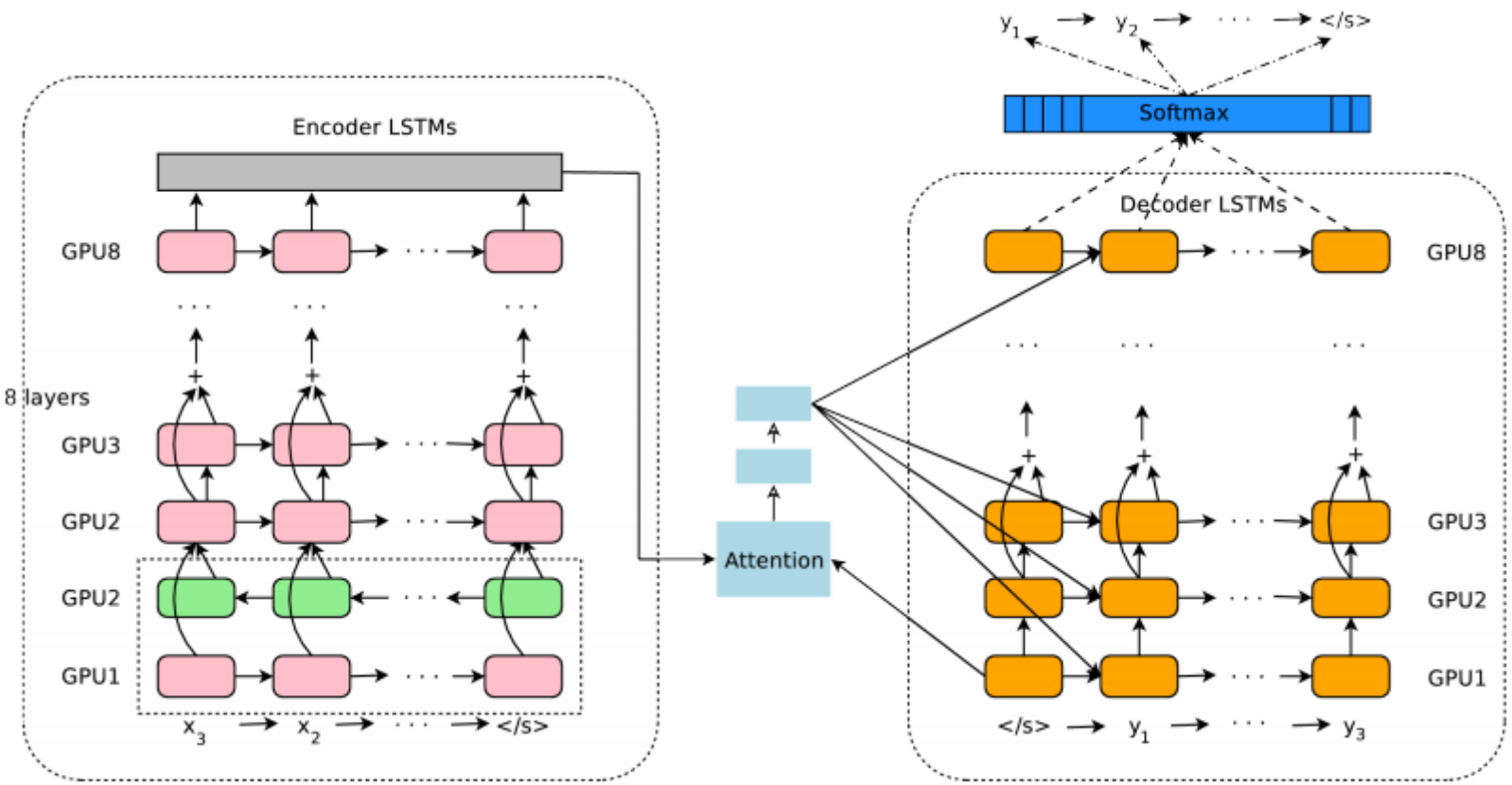
Note: [ ] refers to internal representation

RNN Encoder-decoder with Attention



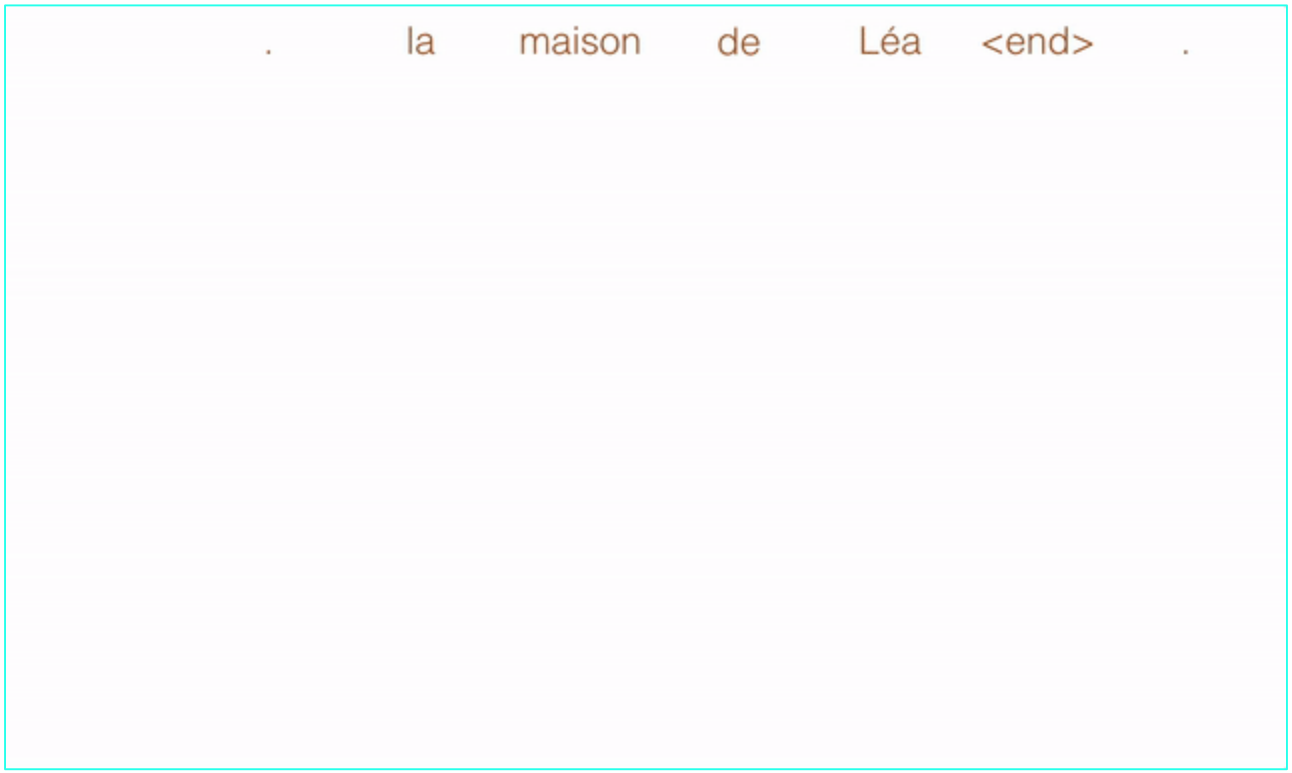
For more information, visit https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html

RNN Encoder-decoder with Attention



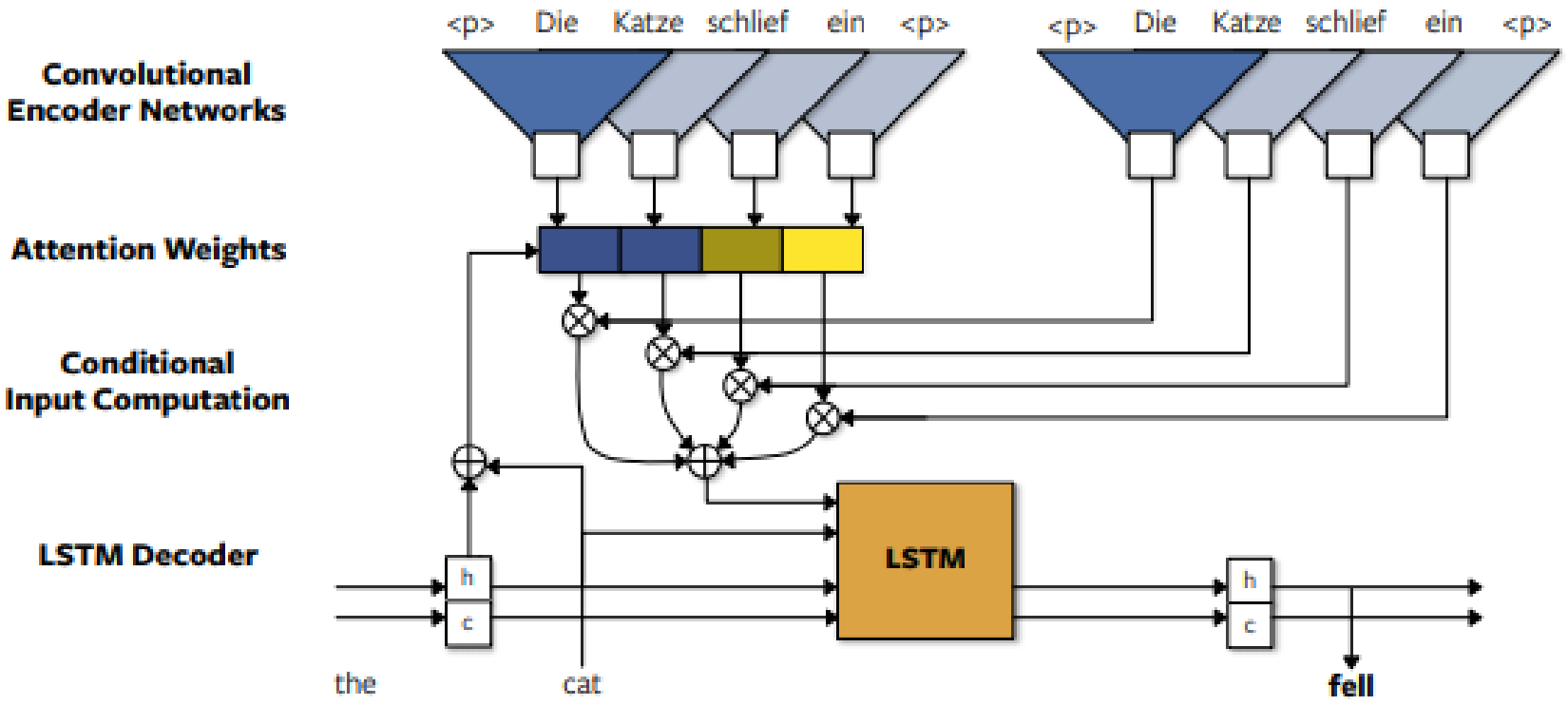
For more information, visit
https://arxiv.org/pdf/1609.08144.pdf

Convolutional NMT



For more information, visit
https://code.facebook.com/posts/1978007565818999/a-novel-approach-to-neural-machine-translation/

Convolutional NMT

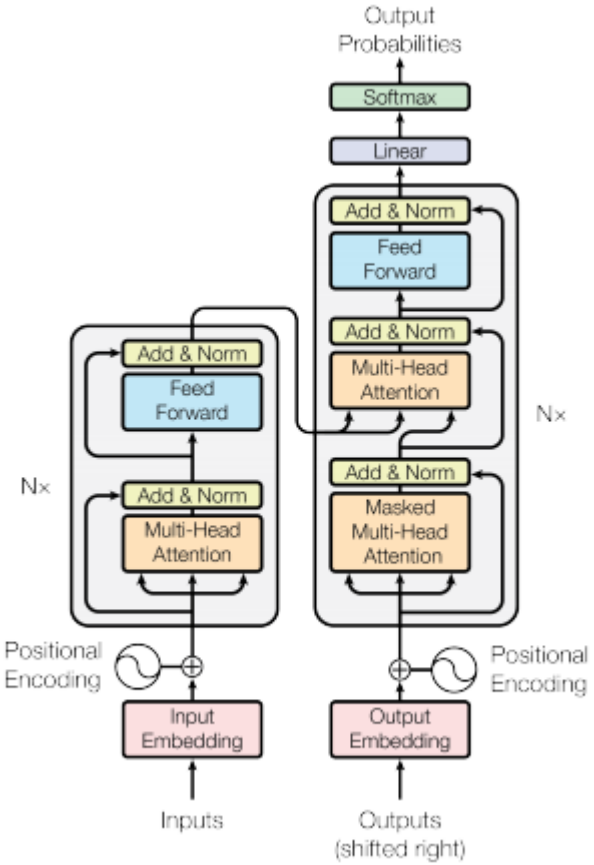


For more information, visit
https://arxiv.org/pdf/1611.02344.pdf

Self-attentional Transformer

For more information, visit
https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html

Self-attentional Transformer



For more information, visit
https://arxiv.org/pdf/1706.03762.pdf

Statistical Machine Translation

$$P(T|S) = \frac{P(S|T) \times P(T)}{P(S)}$$

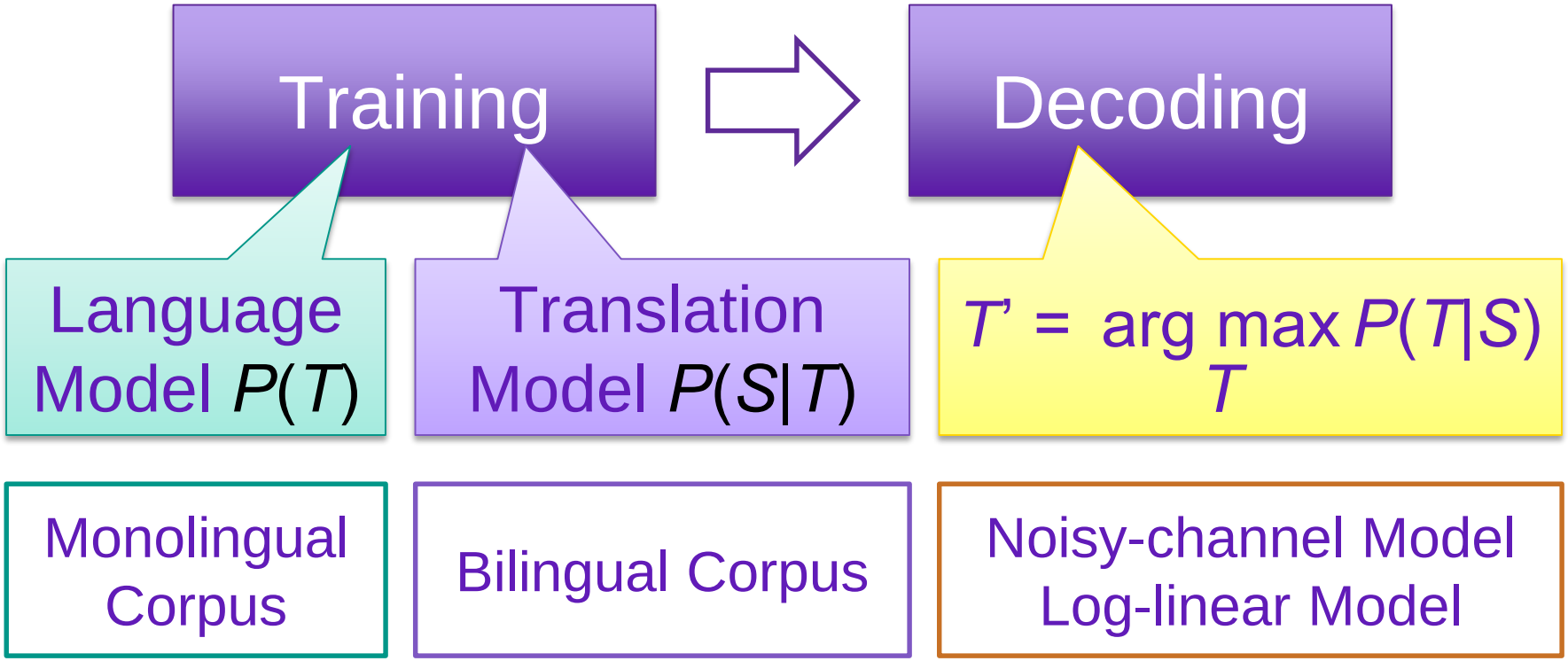
$$T' = \arg\max_{T} P(T|S)$$

$$= \arg\max_{T} P(S|T) \times P(T)$$

Translation Model

Language Model

Statistical Machine Translation



Monolingual (in TL)

Bilingual

Data

Do you love dogs ?
Elephants loves her .
Give me fish .
They love elephants .
I go swimming .
We love dogs .

我喜歡貓。➔ I love cats.
大象喜歡我。➔ Elephants love me.
他喜歡狗。➔ He loves dogs.
他們喜歡魚。➔ They love fish.
我們喜歡猴子。➔ We love monkeys.
你喜歡豬。➔ You love pigs.

Tables

I	0.04
you	0.04
dogs	0.08
fish	0.04
elephants	0.08
love	0.12
loves	0.04
we	0.04
they	0.04

我	I	0.50
我	me	0.50
你	you	1.00
他	he	1.00
貓	cats	1.00
狗	dogs	1.00
魚	fish	1.00
豬	pigs	1.00
猴子	monkeys	1.00
大象	elephants	1.00
喜歡	love	0.83
喜歡	loves	0.17
我們	we	1.00
他們	they	1.00
。	.	1.00

User's Input

狗喜歡大象。

Dogs love elephants.
Dogs elephants love.
Love dogs elephants.
Love elephants dogs.
Elephants love dogs.
Elephants dogs love.
Dogs loves elephants.
Dogs elephants loves.
Loves dogs elephants.
Loves elephants dogs.
Elephants loves dogs.
Elephants dogs loves.

MT Output

Dogs love elephants.

我	I	0.50
我	me	0.50
你	you	1.00
他	he	1.00
貓	cats	1.00
狗	dogs	1.00
魚	fish	1.00
豬	pigs	1.00
猴子	monkeys	1.00
大象	elephants	1.00
喜歡	love	0.83
喜歡	loves	0.17
我們	we	1.00
他們	they	1.00
。	.	1.00

狗	dogs	1.00
大象	elephants	1.00
喜歡	love	0.83
喜歡	loves	0.17
。	.	1.00

Bilingual Table

dogs	0.08
elephants	0.08
love	0.12
loves	0.04

Monolingual Table

$P(T)$

- $= P(t_1 t_2 \ldots t_n)$
- $= P(t_1) \prod_{i=2}^{n} P(t_i | t_1 t_2 \ldots t_{i-1})$

N-gram Model

- We consider *n* words at a time (i.e., the word *w* and *n*-1 words in the context history *h*).

$$
\begin{aligned}
& P(W) \\
= \; & P(w_1 w_2 \ldots w_n w_{n+1}) \\
= \; & \prod_{i=1}^{n+1} P(w_i | w_{i-n+1} w_{i-n+2} \ldots w_{i-1}) \\
= \; & \prod_{i=1}^{n+1} \frac{C(w_{i-n+1} w_{i-n+2} \ldots w_{i-1} w_i)}{C(w_{i-n+1} w_{i-n+2} \ldots w_{i-1})}
\end{aligned}
$$

n = 1

$$P(w_i) = \frac{C(w_i)}{|\text{corpus}|}$$

n > 1

$$P(w_i|h) = \frac{C(h+w_i)}{C(h)},$$

where $h = w_{i-n+1} w_{i-n+2} \ldots w_{i-1}$

where $n=|W|$, "<BOS>"$=w_{-n+2}, w_{-n+3}, \ldots, w_0$ and "<EOS>"$=w_{n+1}$

Unigram Model (n=1)

$$
\begin{aligned}
& P(W) \\
= \; & P(w_1 w_2 \ldots w_n w_{n+1}) \\
= \; & P(w_1) \times P(w_2) \times \ldots \times P(w_n) \times P(w_{n+1}) \\
= \; & \prod_{i=1}^{n+1} P(w_i)
\end{aligned}
$$

where $n=|W|$ and w_{n+1}="<EOS>"

Bigram Model (n=2)

$$
\begin{aligned}
& P(W) \\
= \; & P(w_1 w_2 \ldots w_n w_{n+1}) \\
= \; & P(w_1|w_0) \times P(w_2|w_1) \times P(w_3|w_2) \times \ldots \times P(w_n|w_{n-1}) \times P(w_{n+1}|w_n) \\
= \; & \prod_{i=1}^{n+1} P(w_i|w_{i-1})
\end{aligned}
$$

where n=|W|, w_0="<BOS>" and w_{n+1}="<EOS>"

Trigram Model (n=3)

$$
\begin{aligned}
& P(W) \\
= \; & P(w_1 w_2 \ldots w_n w_{n+1}) \\
= \; & P(w_1 | w_{-1} w_0) \times P(w_2 | w_0 w_1) \times P(w_3 | w_1 w_2) \times \ldots \times P(w_n | w_{n-2} w_{n-1}) \times P(w_{n+1} | w_{n-1} w_n) \\
= \; & \prod_{i=1}^{n+1} P(w_i | w_{i-2} w_{i-1})
\end{aligned}
$$

where n=|W|, w_{-1}=w_0="<BOS>" and w_{n+1}="<EOS>"