翻譯學院
SCHOOL OF TRANSLATION
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

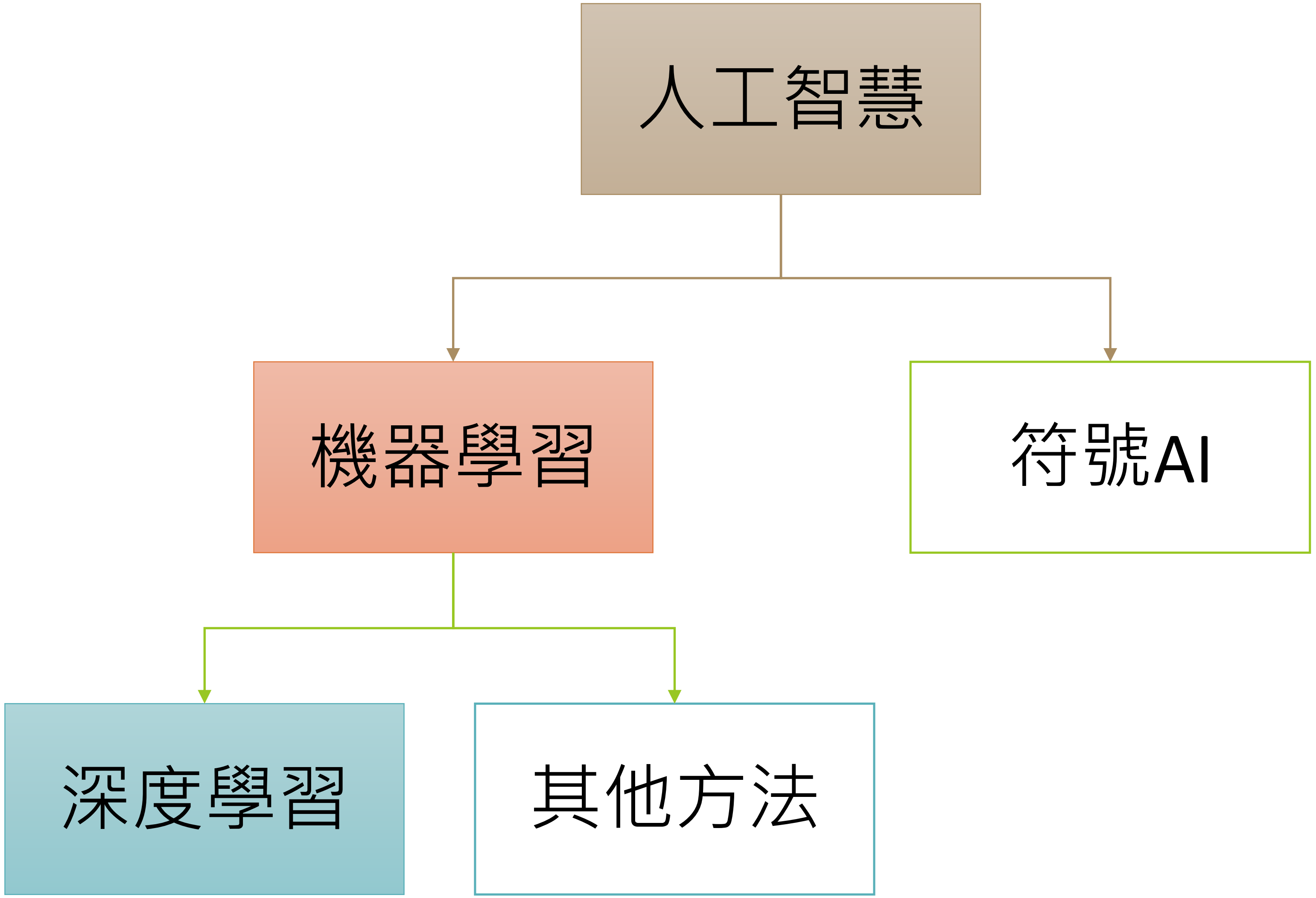
人工神經網絡及語言科技

蕭世昌
翻譯文學碩士（電腦輔助翻譯）課程主任
深度學習研究與應用中心主任

大綱

1. **自我介紹**：關於講者
2. **重要概念**：深度學習與類神經網路
3. **類神經網路在語言科技的應用**: 以機器翻譯為例
4. **實作**：用Python寫一個簡單的類神經網路機器翻譯系統

重要概念


人工智慧
機器學習
符號AI
深度學習
其他方法

深度學習的應用

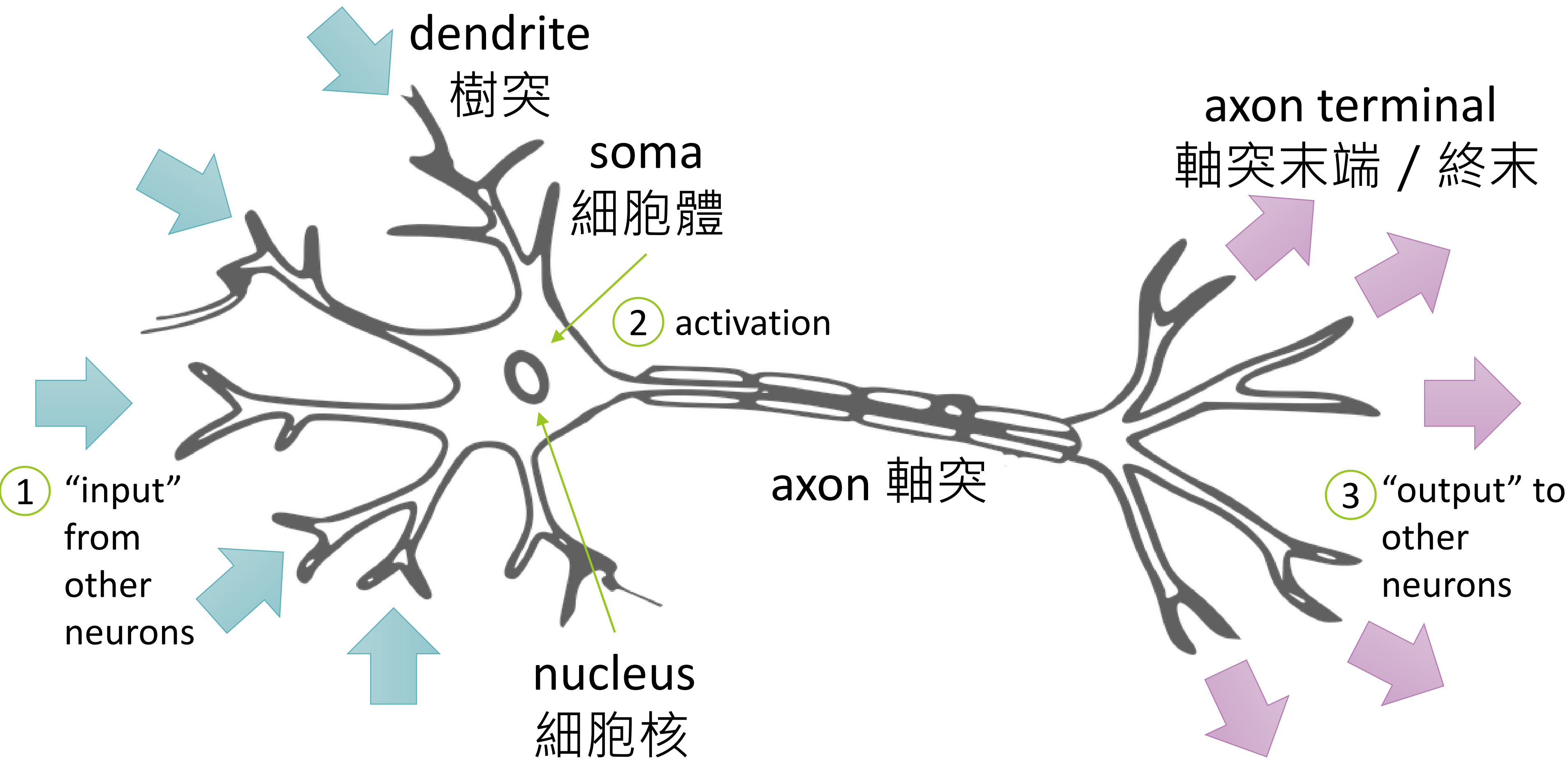
AlphaGo	Speech recognition	Prediction
Image classification	Object detection	Self-driving cars and many more

深度學習

↕

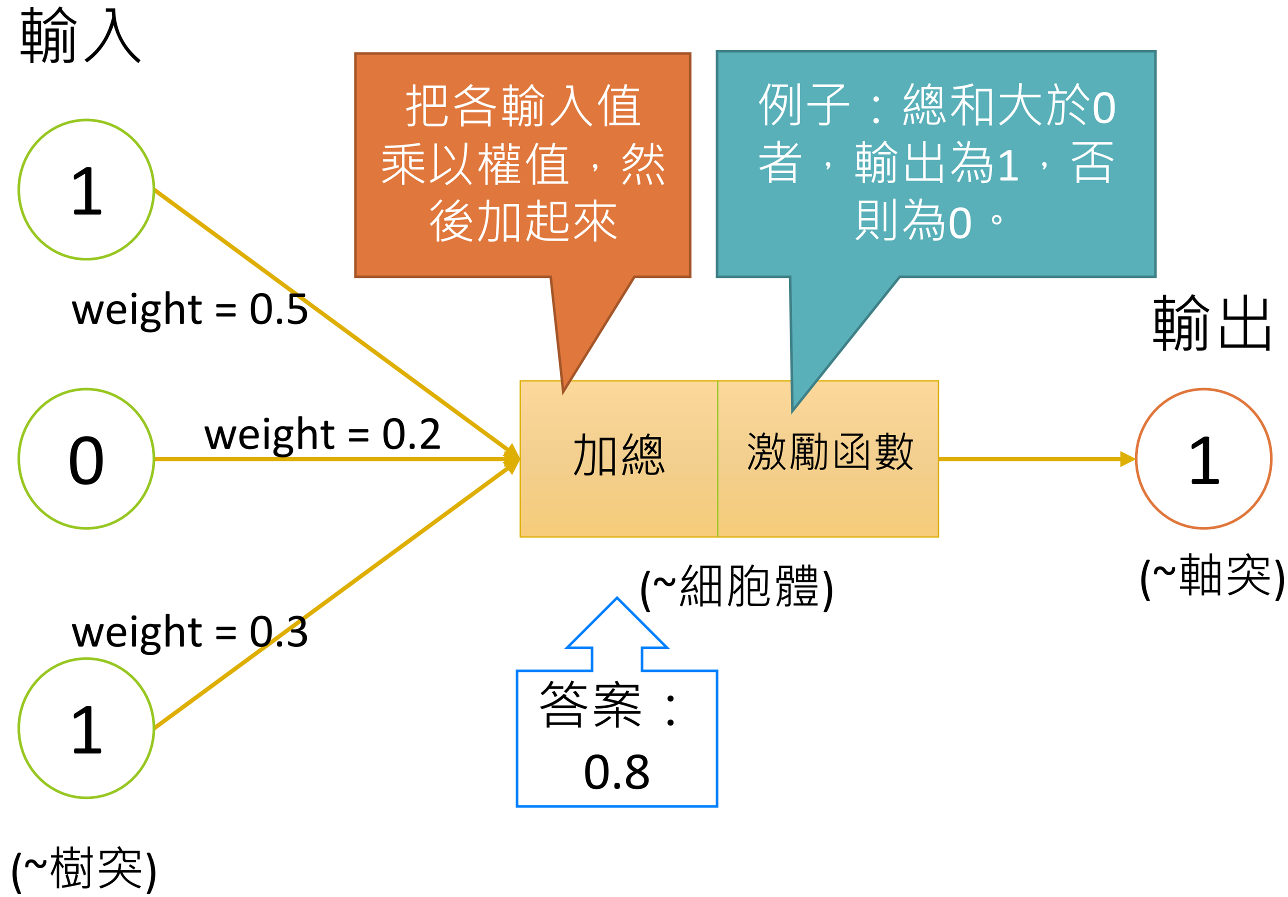
類神經網路

生物神經元



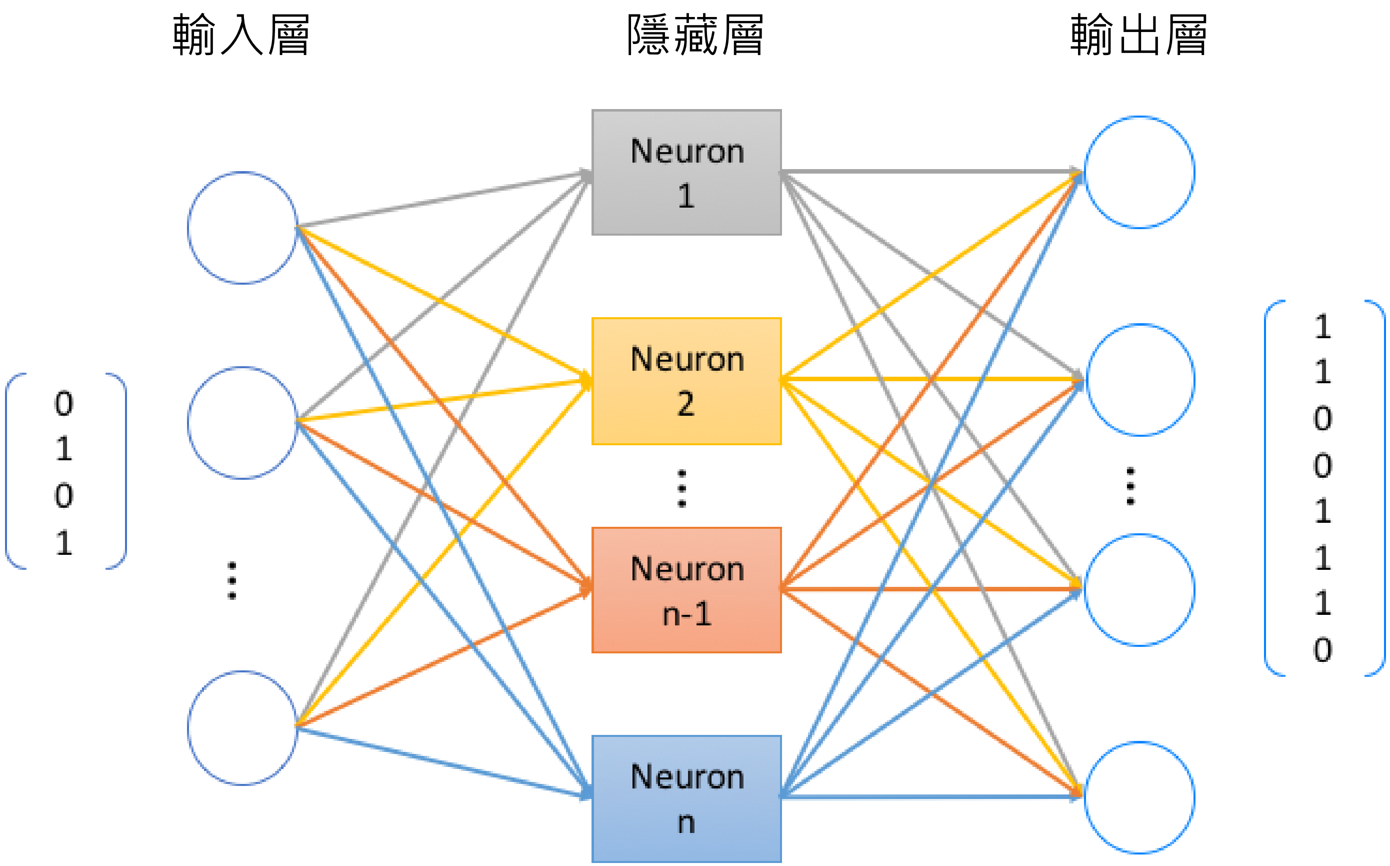
dendrite
樹突
soma
細胞體
2 activation
axon terminal
軸突末端 / 終末
1 “input” from other neurons
axon 軸突
3 “output” to other neurons
nucleus
細胞核

人工神經元

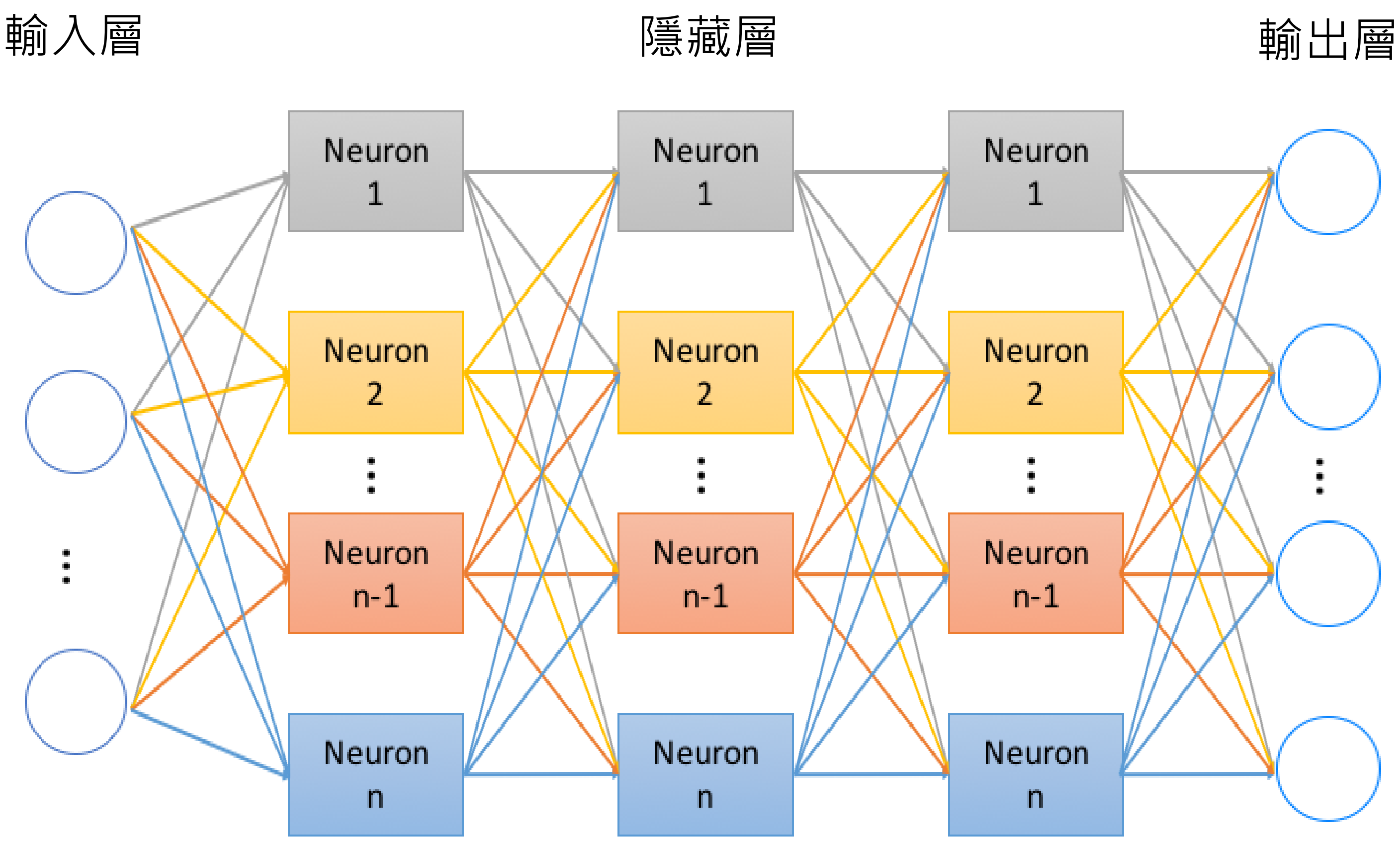


輸入
1
0
1
weight = 0.5
weight = 0.2
weight = 0.3
(~樹突)
把各輸入值乘以權值，然後加起來
例子：總和大於0者，輸出為1，否則為0。
加總
激勵函數
(~細胞體)
答案：
0.8
輸出
1
(~軸突)

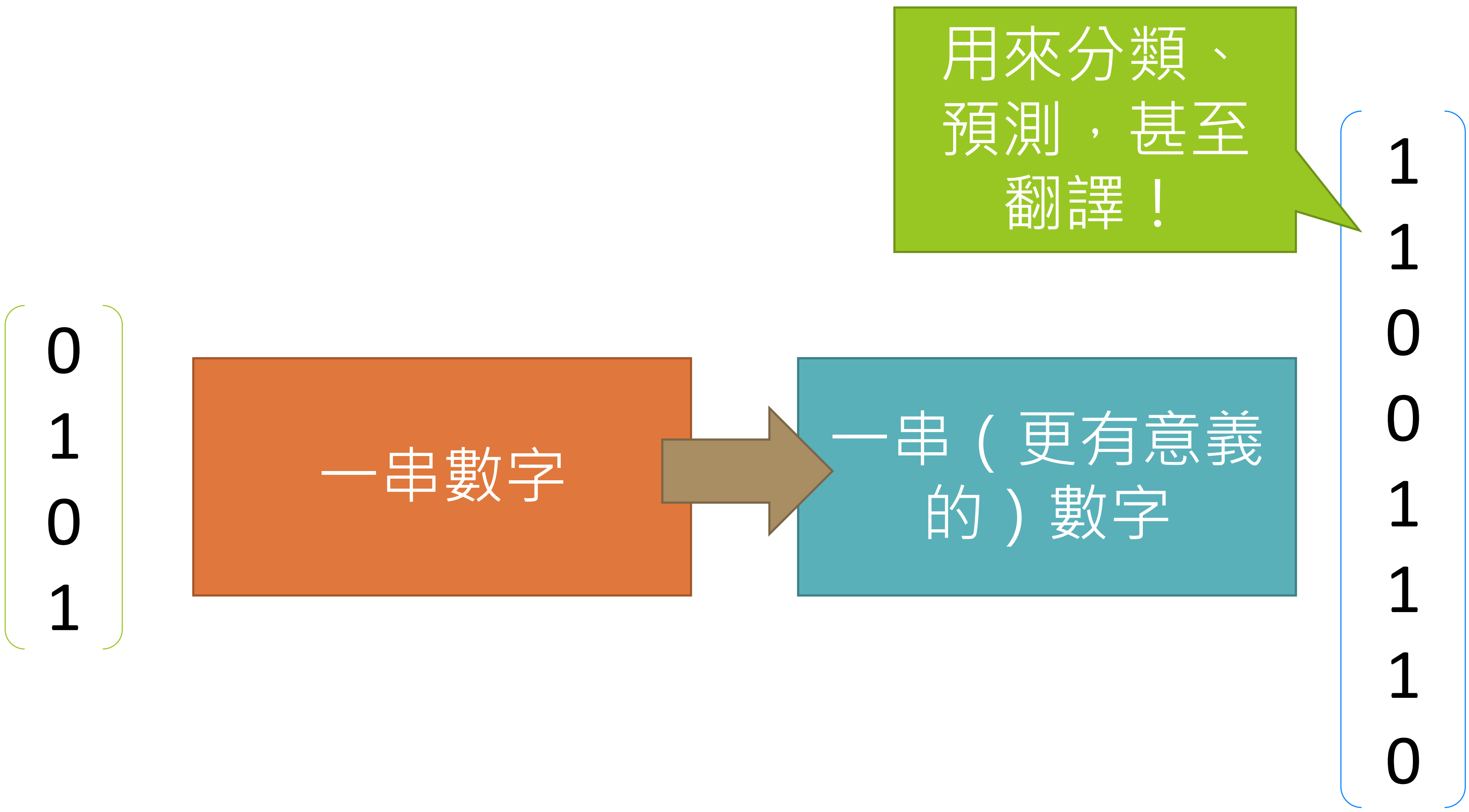
類神經網路：神經元連在一起



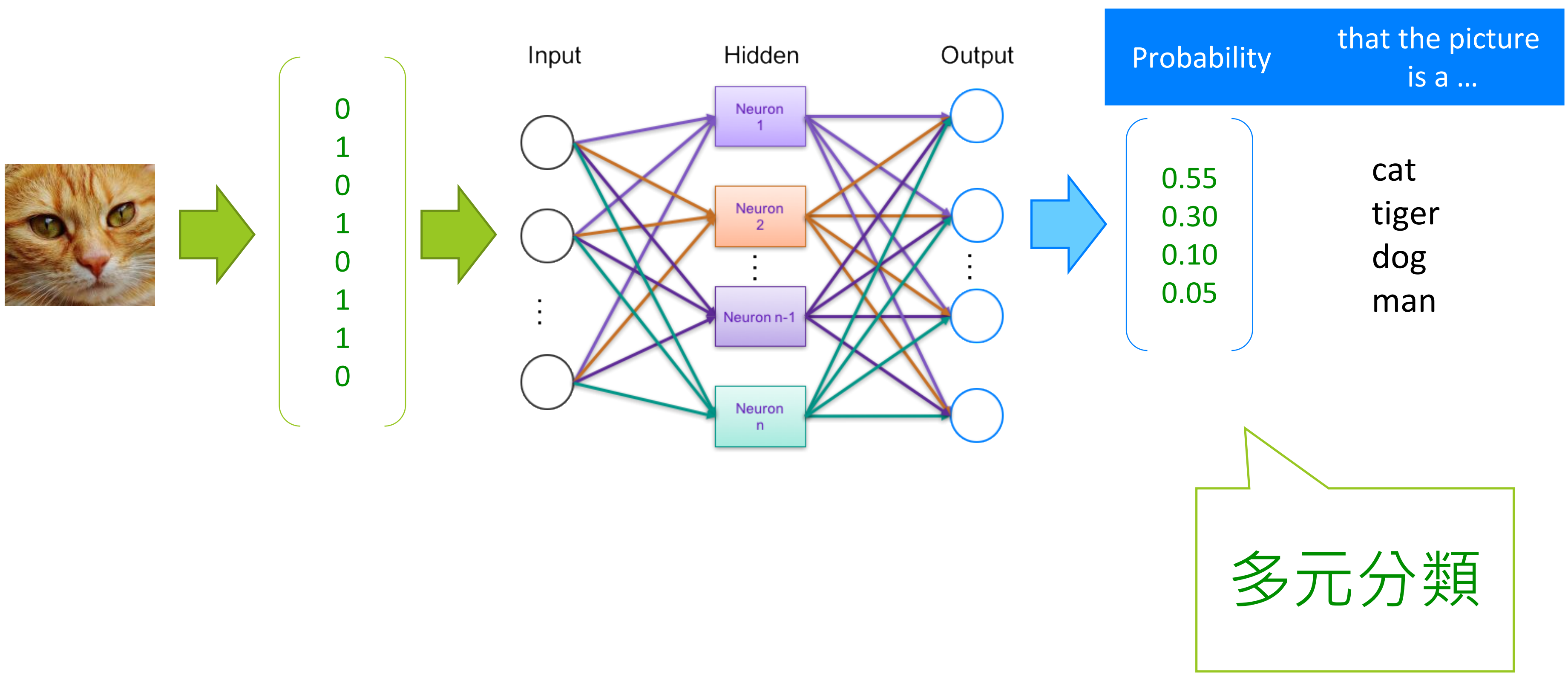
深度學習：Many 隱藏層！



類神經網路



示例：分類



0
1
0
1
0
1
1
0
Input
Hidden
Output
Neuron 1
Neuron 2
Neuron n-1
Neuron n
Probability
that the picture is a ...
0.55
0.30
0.10
0.05
cat
tiger
dog
man
多元分類

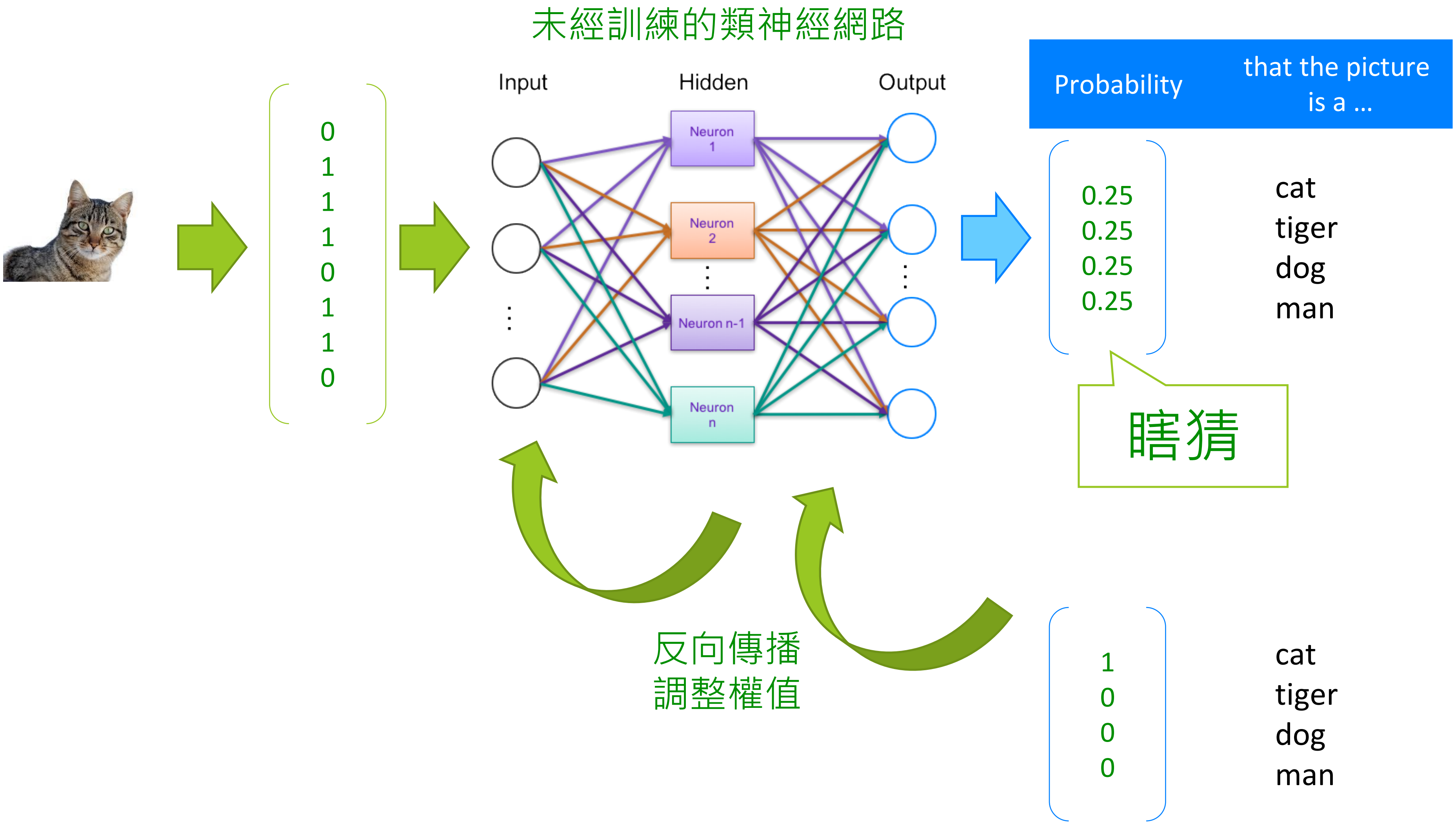
類神經網路怎麼懂得分類？

訓練

Training data

INPUT	**OUTPUT**
	CAT
	DOG
	CAT
	CAT
	DOG

示例：分類



未經訓練的類神經網路
Input
Hidden
Output
Neuron 1
Neuron 2
Neuron n-1
Neuron n
0
1
1
1
0
1
1
0
Probability
that the picture is a ...
0.25
0.25
0.25
0.25
cat
tiger
dog
man
瞎猜
反向傳播
調整權值
1
0
0
0
cat
tiger
dog
man

深度學習的三大重點

數據	**機器**	**算法**

圖形處理器

類神經網路在語言科技的應用

以機器翻譯為例

先看例子，再講理論

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,

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

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

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

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

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

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 $\mathbf{h}$:

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

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:

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

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 $P(\mathbf{y}|\mathbf{x})$

$$\hat{\mathbf{y}} = \operatorname*{argmax}_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}) \tag{6}$$

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

兩個財經翻譯工具

機器翻譯原理

機器翻譯的三大傳統方法

01 Rule-based Machine Translation

- I go to school by train.
- 我／去／學校／乘／火車／。
- 我乘火車去學校。

02 Example-based Machine Translation

- I go to school by train.
- I go to school by <u>taxi</u>. 我乘<u>計程車</u>去學校。
- 我乘火車去學校。

03 Statistical Machine Translation

- Translation Model
  P(“I” | “我”) x *P*(“go” | “去”) x ...
- Language Model:
  P(“我乘火車去學校”)

Automatic Translation 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

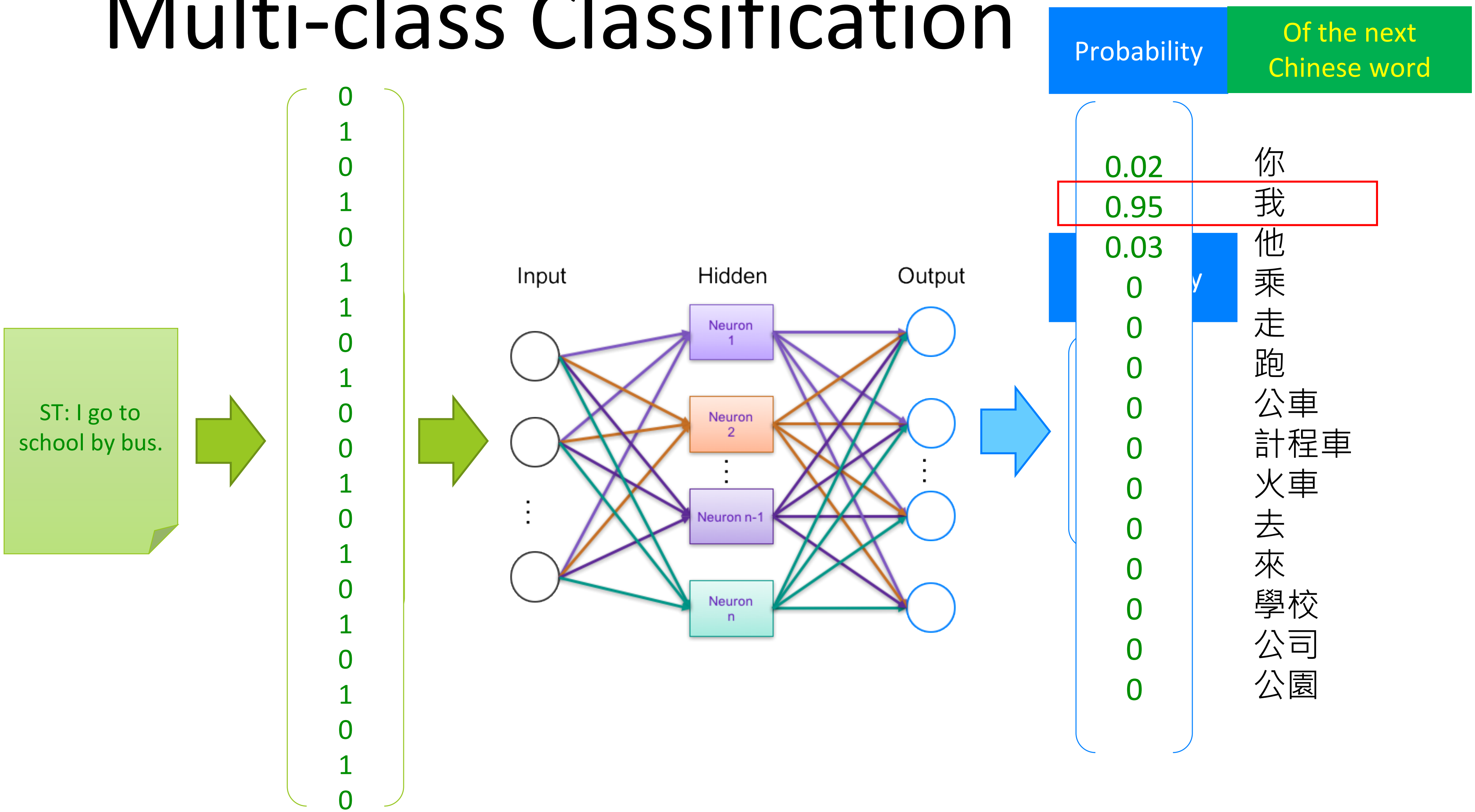
學校

咖啡

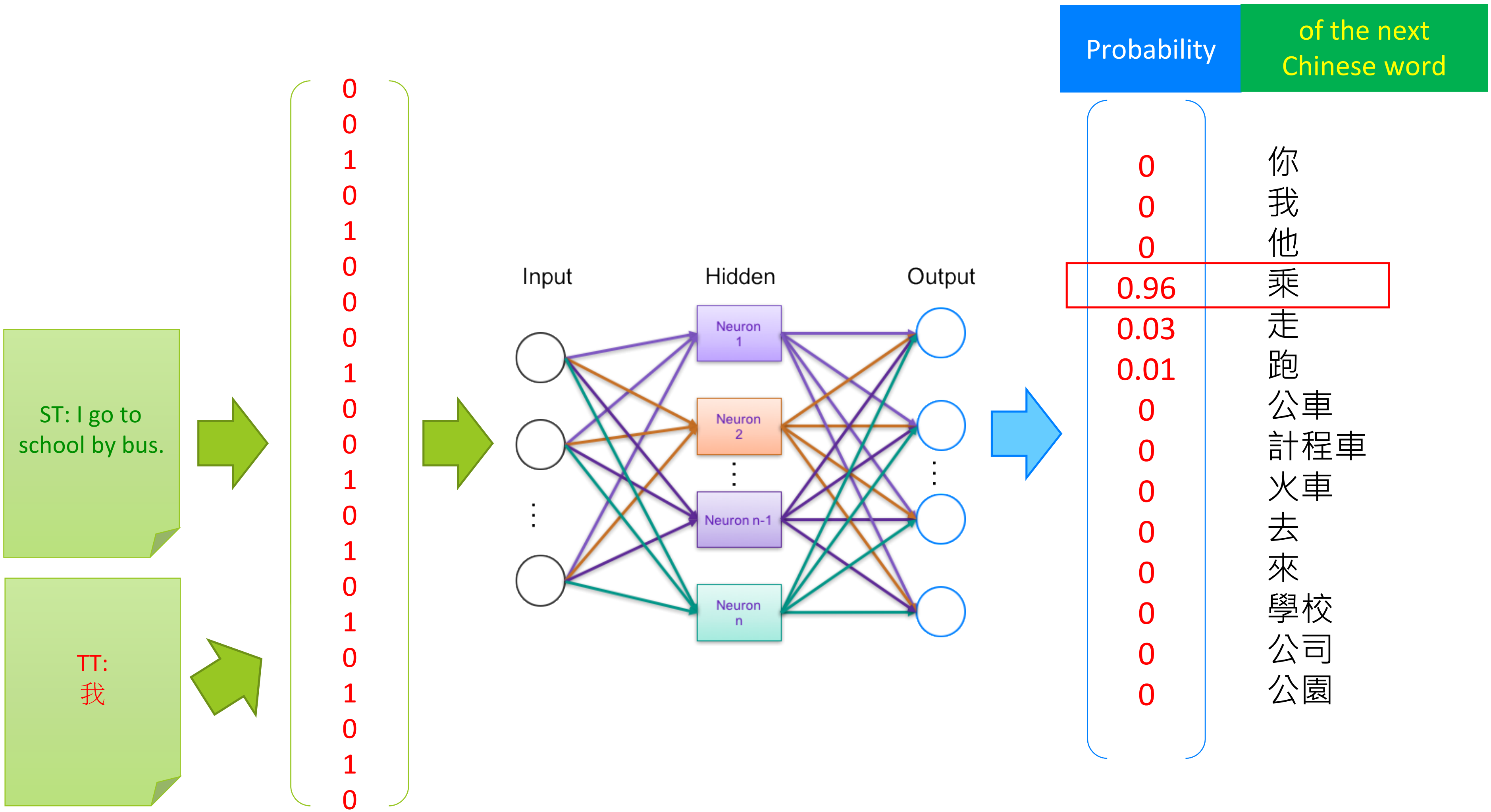
電腦

翻譯

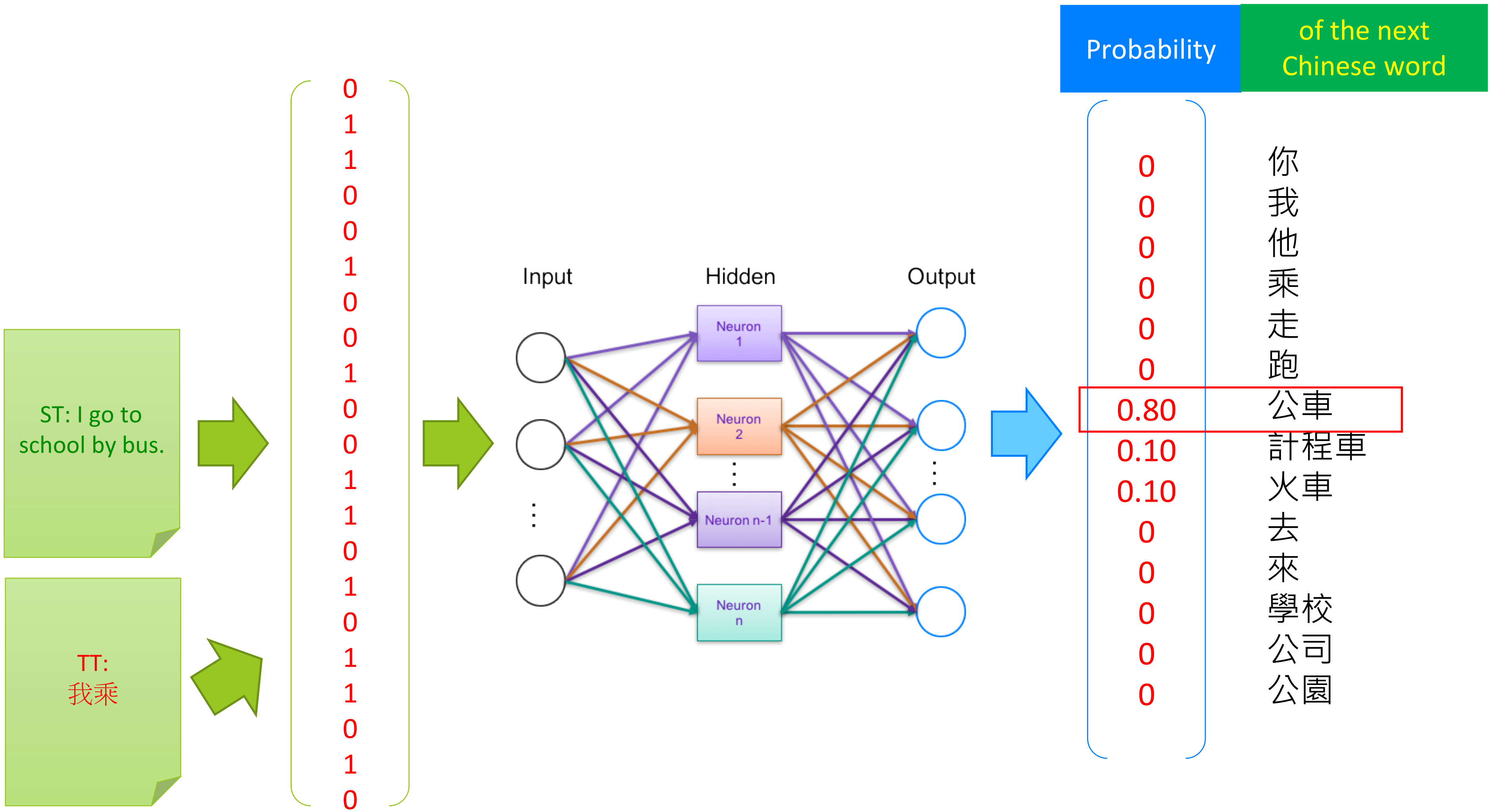
Neural Machine Translation as Multi-class Classification



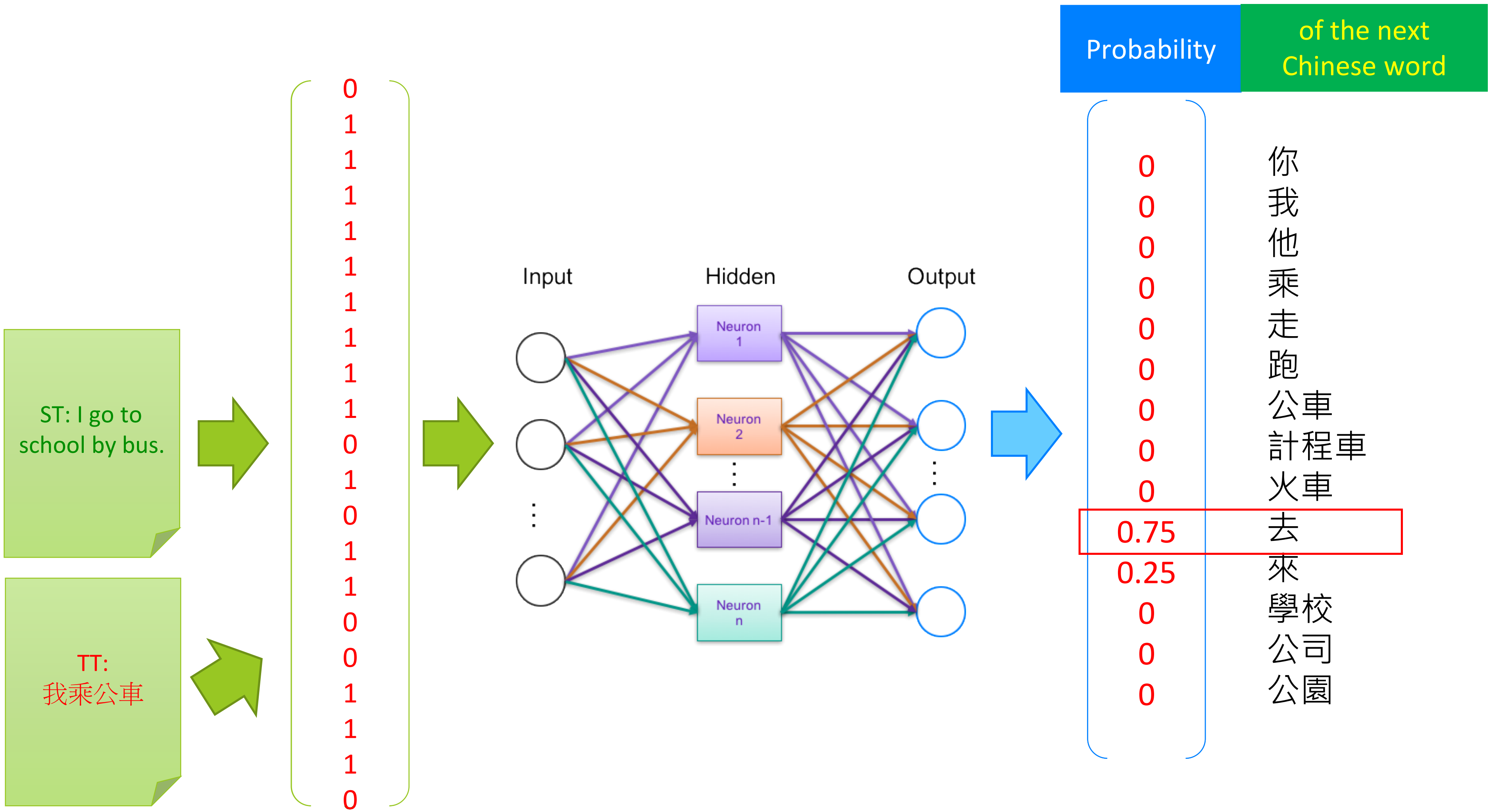
Neural Machine Translation



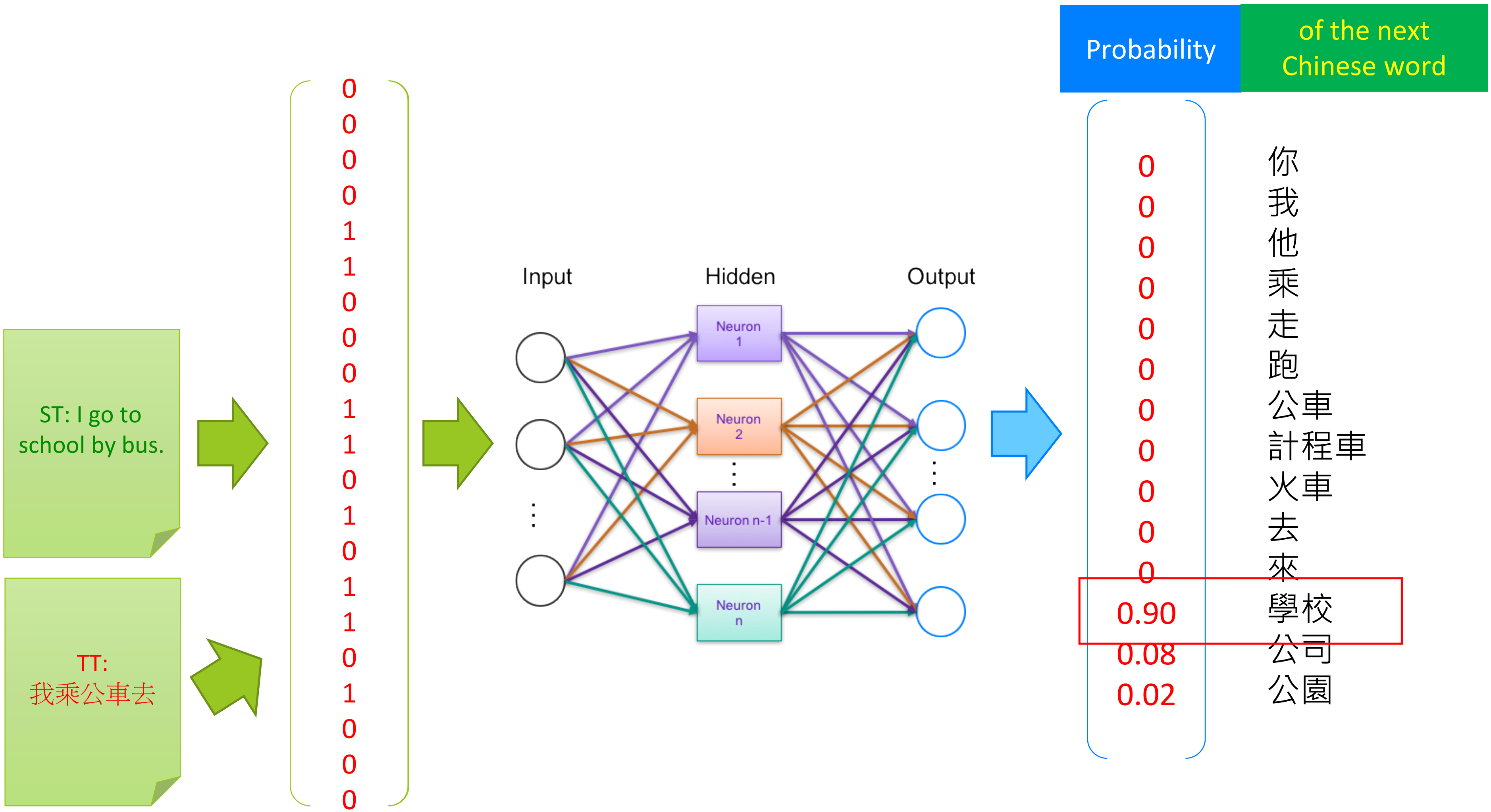
Neural Machine Translation



Neural Machine Translation



Neural Machine Translation



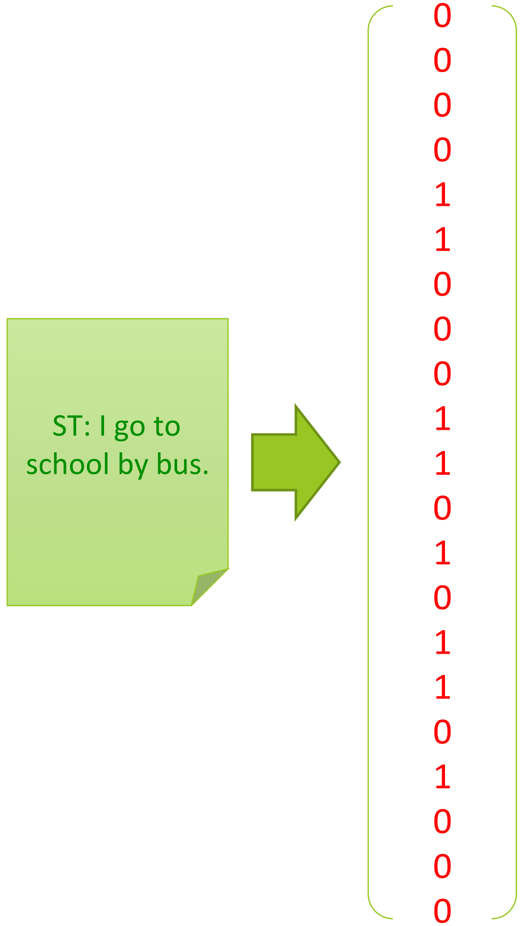
如何訓練？

Training data

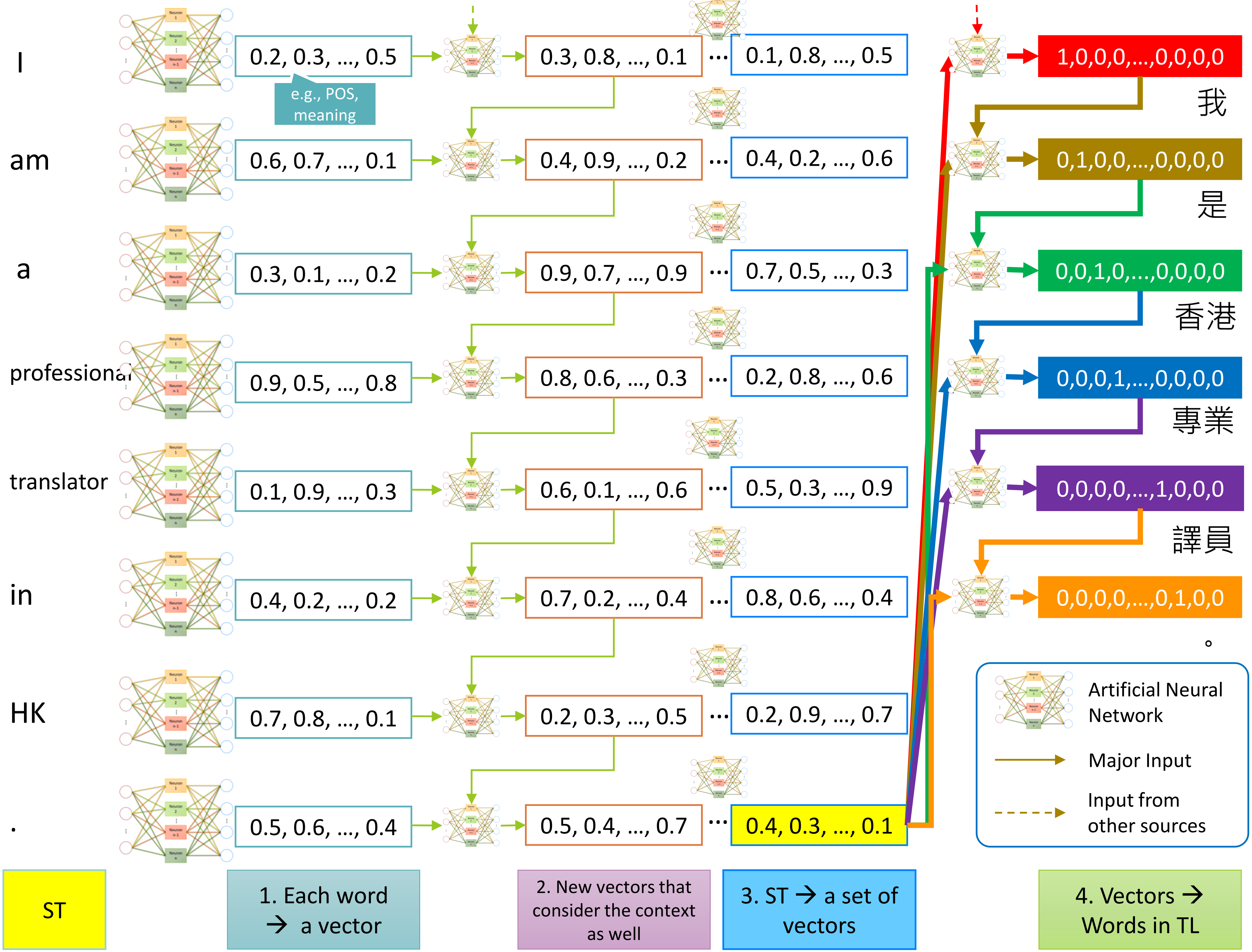
INPUT	OUTPUT
I go to school by bus.	我乘公車上學去。
This is a cup.	這是一個杯子。
How old are you?	你幾歲？
Good morning.	早上好。
Where is he?	他在哪裡？

預測結果		標準答案
0.02	你	0
0.95	我	1
0.03	他	0
0	乘	0
0	走	0
0	跑	0
0	公車	0
0	計程車	0
0	火車	0
0	去	0
0	來	0
0	學校	0
0	公司	0
0	公園	0

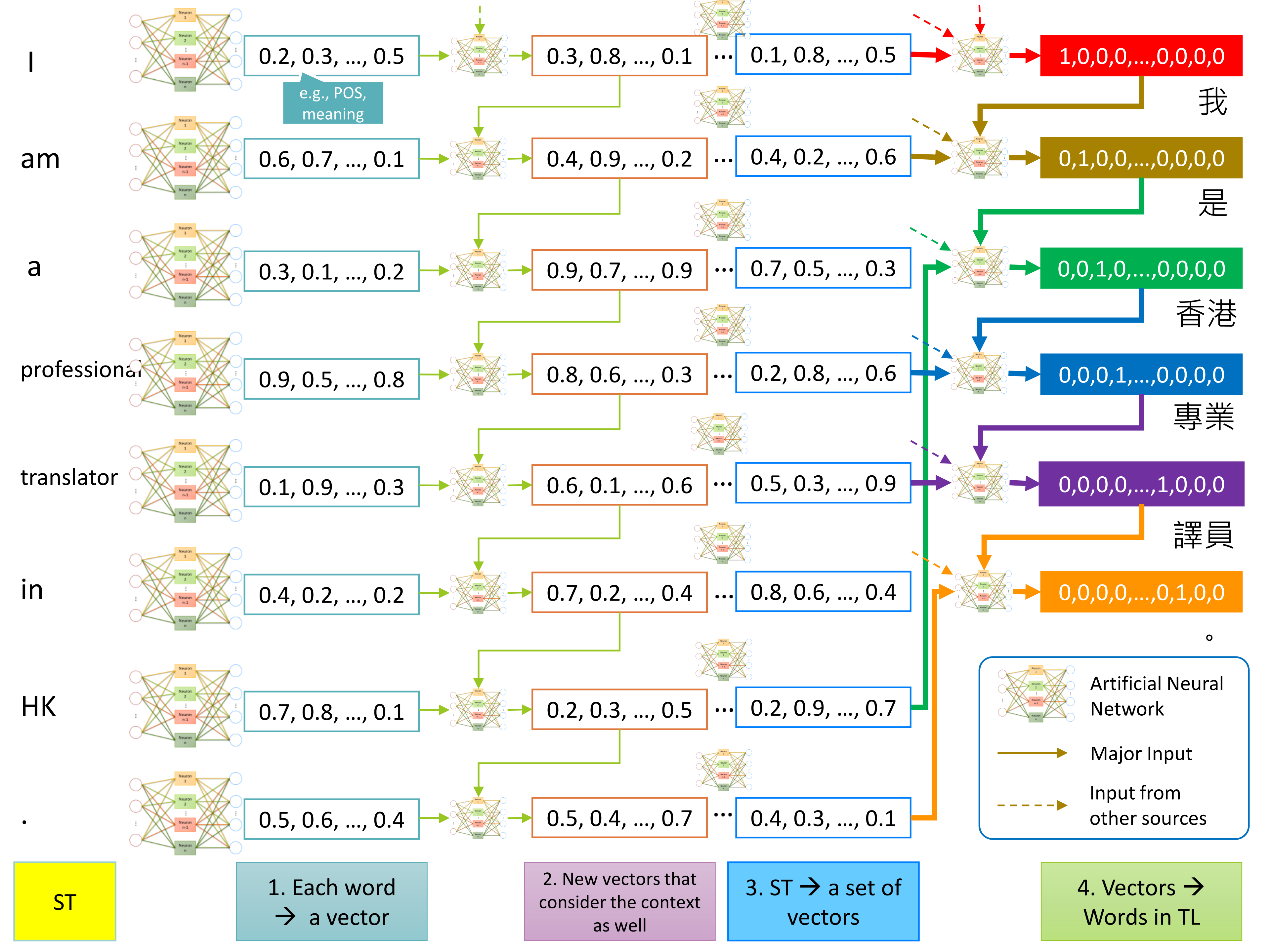
接下來的問題是



How?

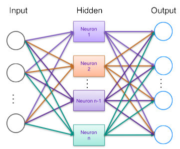


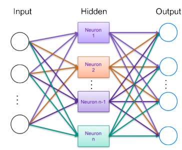
I
am
a
professional
translator
in
HK
.
0.2, 0.3, ..., 0.5
e.g., POS, meaning
0.6, 0.7, ..., 0.1
0.3, 0.1, ..., 0.2
0.9, 0.5, ..., 0.8
0.1, 0.9, ..., 0.3
0.4, 0.2, ..., 0.2
0.7, 0.8, ..., 0.1
0.5, 0.6, ..., 0.4
0.3, 0.8, ..., 0.1
0.4, 0.9, ..., 0.2
0.9, 0.7, ..., 0.9
0.8, 0.6, ..., 0.3
0.6, 0.1, ..., 0.6
0.7, 0.2, ..., 0.4
0.2, 0.3, ..., 0.5
0.5, 0.4, ..., 0.7
0.1, 0.8, ..., 0.5
0.4, 0.2, ..., 0.6
0.7, 0.5, ..., 0.3
0.2, 0.8, ..., 0.6
0.5, 0.3, ..., 0.9
0.8, 0.6, ..., 0.4
0.2, 0.9, ..., 0.7
0.4, 0.3, ..., 0.1
1,0,0,0,...,0,0,0,0
我
0,1,0,0,...,0,0,0,0
是
0,0,1,0,...,0,0,0,0
香港
0,0,0,1,...,0,0,0,0
專業
0,0,0,0,...,1,0,0,0
譯員
0,0,0,0,...,0,1,0,0
。
Artificial Neural Network
Major Input
Input from other sources
ST
1. Each word → a vector
2. New vectors that consider the context as well
3. ST → a set of vectors
4. Vectors → Words in TL



I
am
a
professional
translator
in
HK
.
0.2, 0.3, ..., 0.5
e.g., POS, meaning
0.6, 0.7, ..., 0.1
0.3, 0.1, ..., 0.2
0.9, 0.5, ..., 0.8
0.1, 0.9, ..., 0.3
0.4, 0.2, ..., 0.2
0.7, 0.8, ..., 0.1
0.5, 0.6, ..., 0.4
0.3, 0.8, ..., 0.1
0.4, 0.9, ..., 0.2
0.9, 0.7, ..., 0.9
0.8, 0.6, ..., 0.3
0.6, 0.1, ..., 0.6
0.7, 0.2, ..., 0.4
0.2, 0.3, ..., 0.5
0.5, 0.4, ..., 0.7
0.1, 0.8, ..., 0.5
0.4, 0.2, ..., 0.6
0.7, 0.5, ..., 0.3
0.2, 0.8, ..., 0.6
0.5, 0.3, ..., 0.9
0.8, 0.6, ..., 0.4
0.2, 0.9, ..., 0.7
0.4, 0.3, ..., 0.1
1,0,0,0,...,0,0,0,0
我
0,1,0,0,...,0,0,0,0
是
0,0,1,0,...,0,0,0,0
香港
0,0,0,1,...,0,0,0,0
專業
0,0,0,0,...,1,0,0,0
譯員
0,0,0,0,...,0,1,0,0
。
Artificial Neural Network
Major Input
Input from other sources
ST
1. Each word → a vector
2. New vectors that consider the context as well
3. ST → a set of vectors
4. Vectors → Words in TL

除了機器翻譯之外，
還有其他NLP應用嗎？





文本 → 數字 → （更有意義的）數字

→ 分類

→ 預測

→ 生成序列

情感分析

自動摘要

聊天機器人

寫作批改

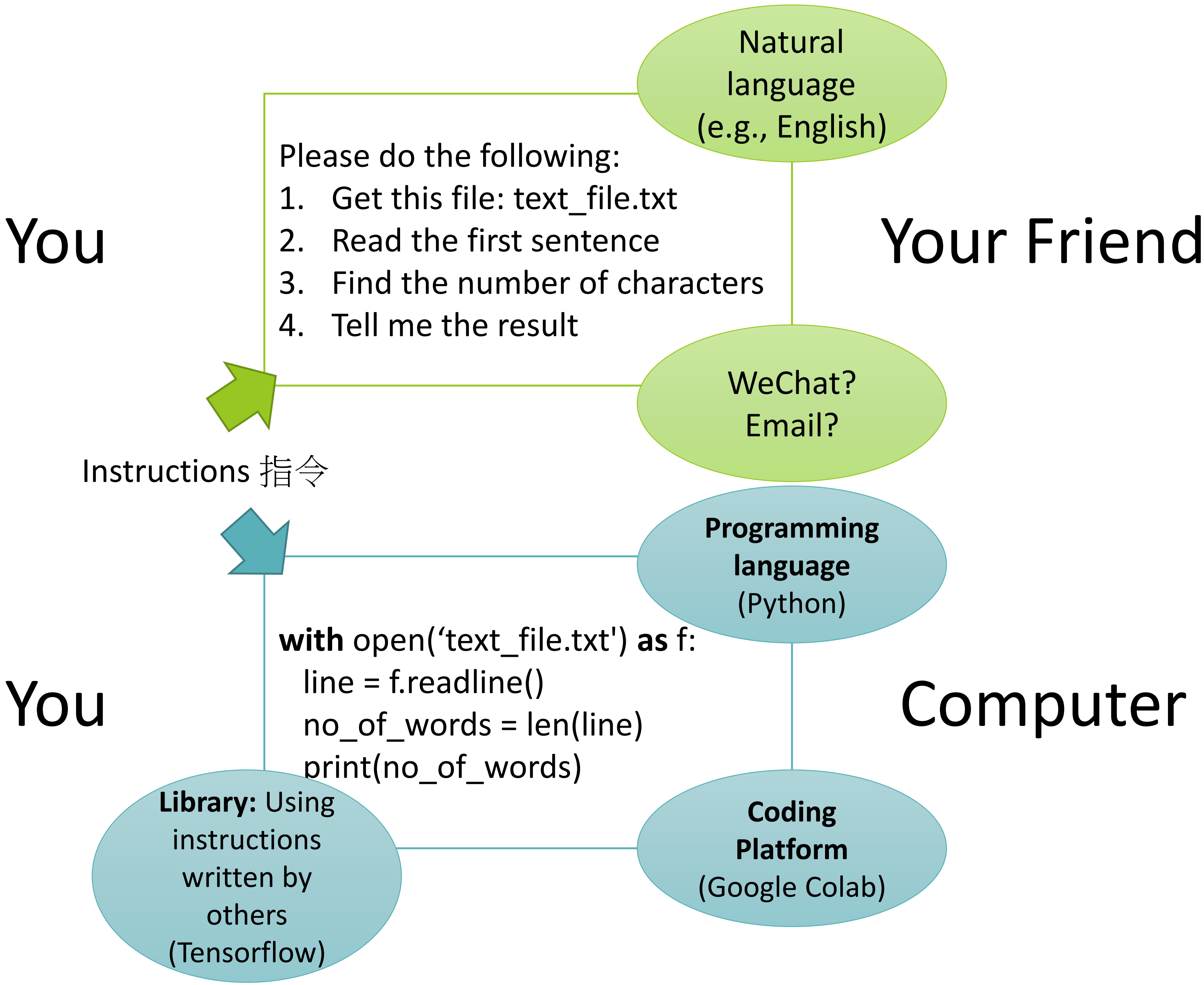
實作

工具

- **程式語言**：Python
- **編程平台**：Google Colab
- **支援工具**：Tensorflow，還有一個預先寫好的模組

寫程式不過是翻譯而已。

自然語言 → 電腦語言



Natural language (e.g., English)
You
Your Friend
Please do the following:
1. Get this file: text_file.txt
2. Read the first sentence
3. Find the number of characters
4. Tell me the result
WeChat? Email?
Instructions 指令
Programming language (Python)
with open(‘text_file.txt') as f:
line = f.readline()
no_of_words = len(line)
print(no_of_words)
You
Computer
Library: Using instructions written by others (Tensorflow)
Coding Platform (Google Colab)

要點

1. 設置 Google Colab Notebook
2. 寫幾行Python代碼試試看
3. 了解流程：建置一個簡單的類神經網路翻譯系統
4. 試試看！

個人Twitter:

ainow6

謝謝！