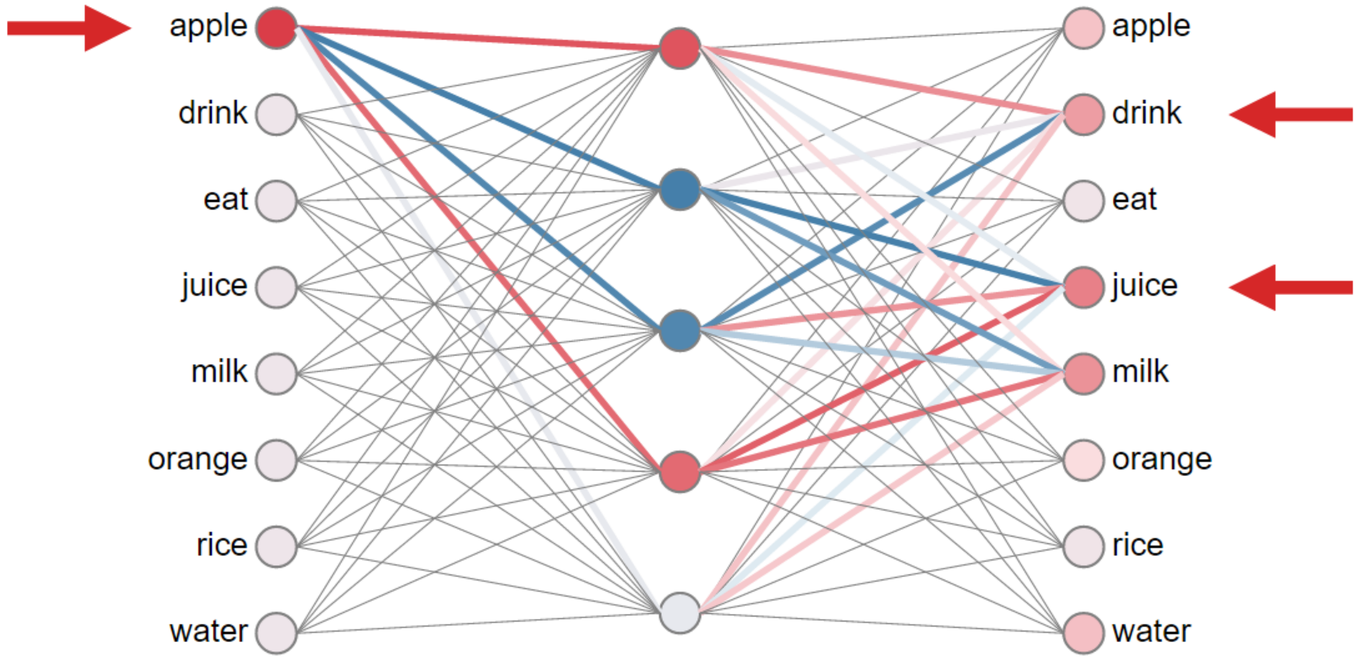
Word Embeddings
Word2Vec Variants

國立臺灣大學 資訊工程學系
陳縕儂 助理教授
http://vivianchen.idv.tw

Word2Vec Skip-Gram Visualization https://ronxin.github.io/wevi/

- Skip-gram training data:
  apple|drink^juice,orange|eat^apple,rice|drink^juice,juice|drink^milk,milk|drink^rice,water|drink^milk,juice|orange^apple,juice|apple^drink,milk|rice^drink,drink|milk^water,drink|water^juice,drink|juice^water



wevi: word embedding visual inspector

Everything you need to know about this tool - Source code

Control Panel

Config:

```
{"hidden_size":5,"random_state":1,"learning_rate":0.2}
```

Training data (context|target):

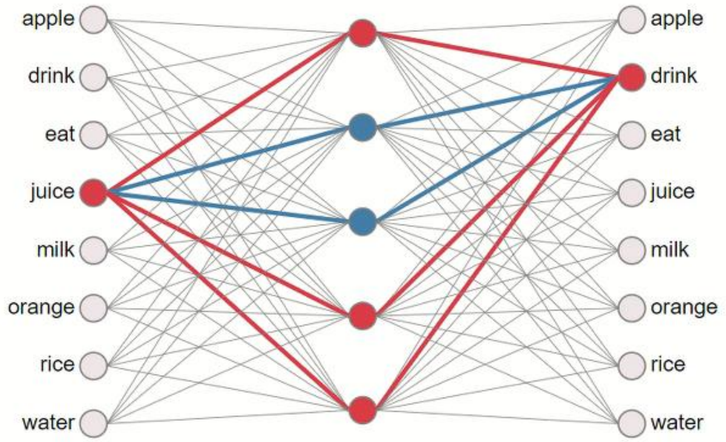
```
apple|drink^juice,orange|eat^apple,rice|drink^juice,juice|drink^milk,milk|drink^rice,water|drink^milk,juice|orange^apple,juice|apple^drink,milk|rice^drink,drink|milk^water,drink|water^juice,drink|juice^water
```

Presets: Fruit and juice (Skip-gram) ▼

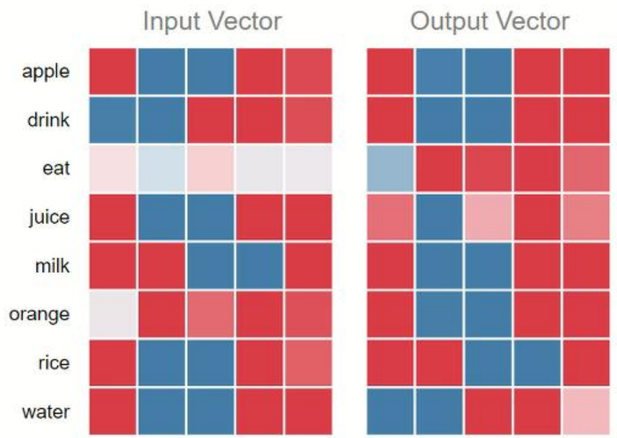
Update and Restart | Update Learning Rate

Next | 20 | 100 | 500 | PCA

Neurons



Weight Matrices



Vectors

Word2Vec Variants

- **Skip-gram**: predicting surrounding words given the target word (Mikolov+, 2013)

better

$$p(w_{t-m}, \cdots w_{t-1}, w_{t+1}, \cdots, w_{t+m} \mid w_t)$$

- **CBOW (continuous bag-of-words)**: predicting the target word given the surrounding words (Mikolov+, 2013)

$$p(w_t \mid w_{t-m}, \cdots w_{t-1}, w_{t+1}, \cdots, w_{t+m})$$

- **LM (Language modeling)**: predicting the next words given the proceeding contexts (Mikolov+, 2013)

first

$$p(w_{t+1} \mid w_t)$$

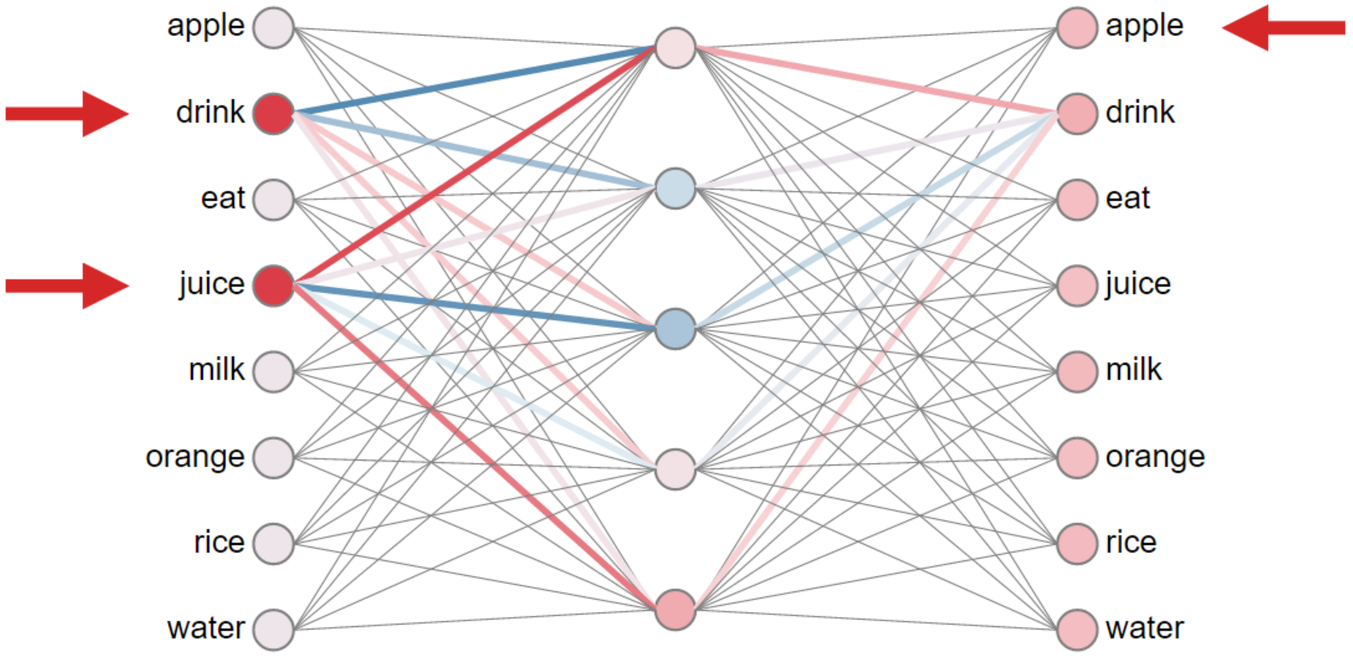
Practice the derivation by yourself!!

Mikolov et al., “Efficient estimation of word representations in vector space,” in *ICLR Workshop*, 2013.
Mikolov et al., “Linguistic regularities in continuous space word representations,” in *NAACL HLT*, 2013.

Word2Vec CBOW

- Goal: predicting the target word given the surrounding words

$$p(w_t \mid w_{t-m}, \cdots w_{t-1}, w_{t+1}, \cdots, w_{t+m})$$



Word2Vec LM

- Goal: predicting the next words given the proceeding contexts

$$p(w_{t+1} \mid w_t)$$