

RDFグラフから抽出した多様な特徴ベクトルによる教師あり学習

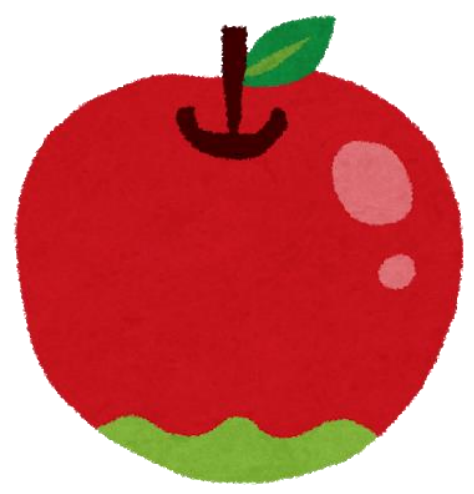
兼岩研究室 南陽太

RDFグラフ

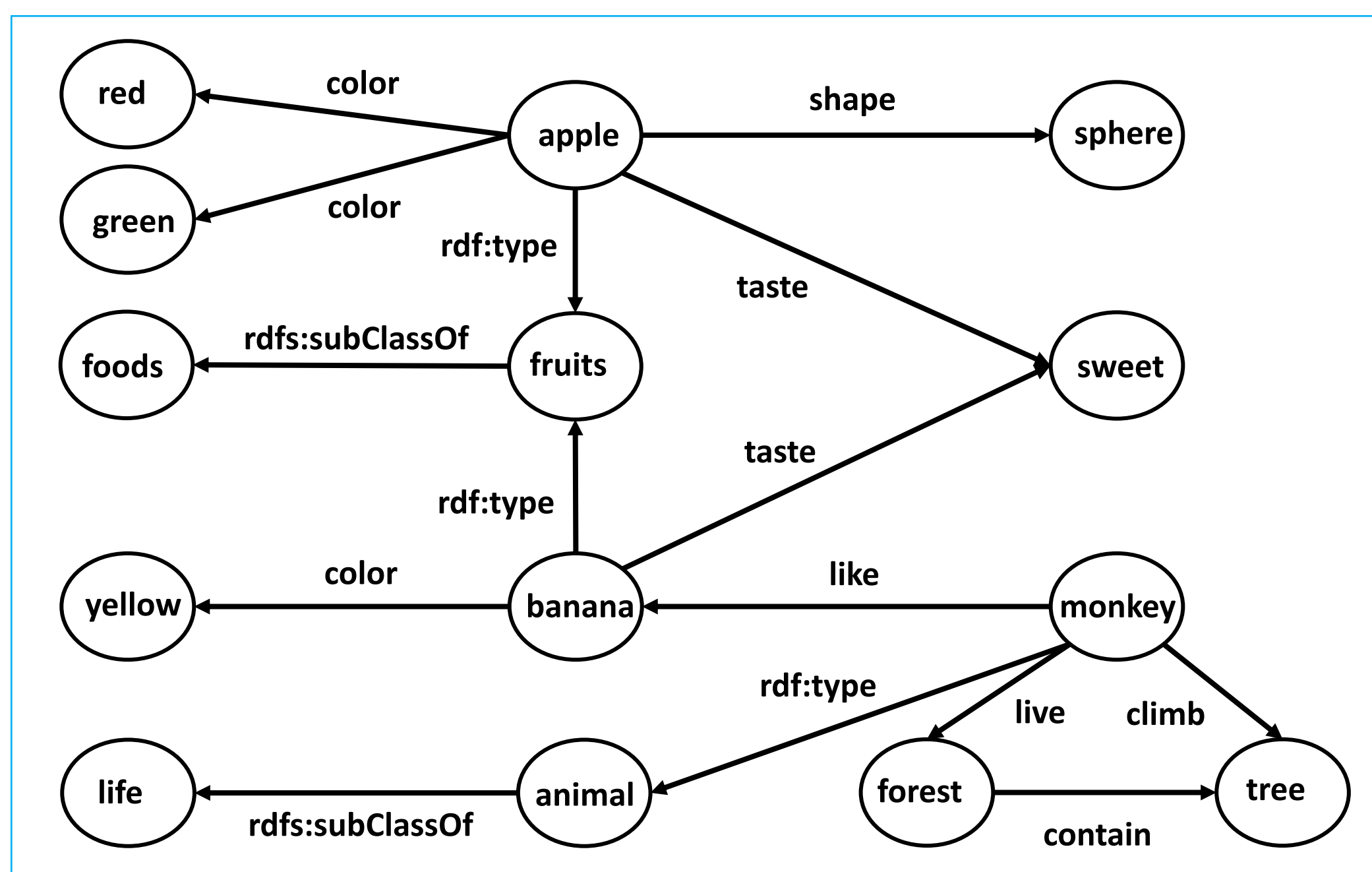
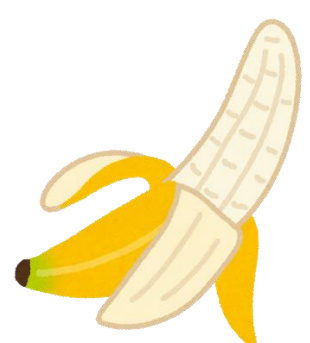
- **RDFトリプル**：主語s, 述語p, 目的語oの三つ組(s,p,o).
- **RDFグラフ**：RDFトリプルの集合として作られるグラフ.

RDFトリプル

```
{(apple,color,red),  
(apple,shape,sphere),  
(apple,rdf:type,fruits),  
(banana,rdf:type,fruits),  
:  
(forest,contain,tree)}
```

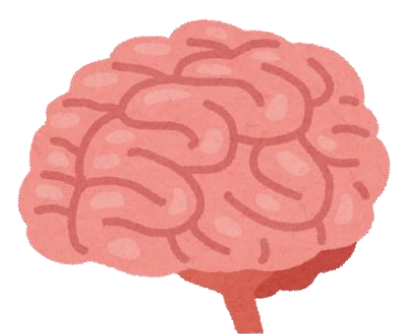


RDFグラフ

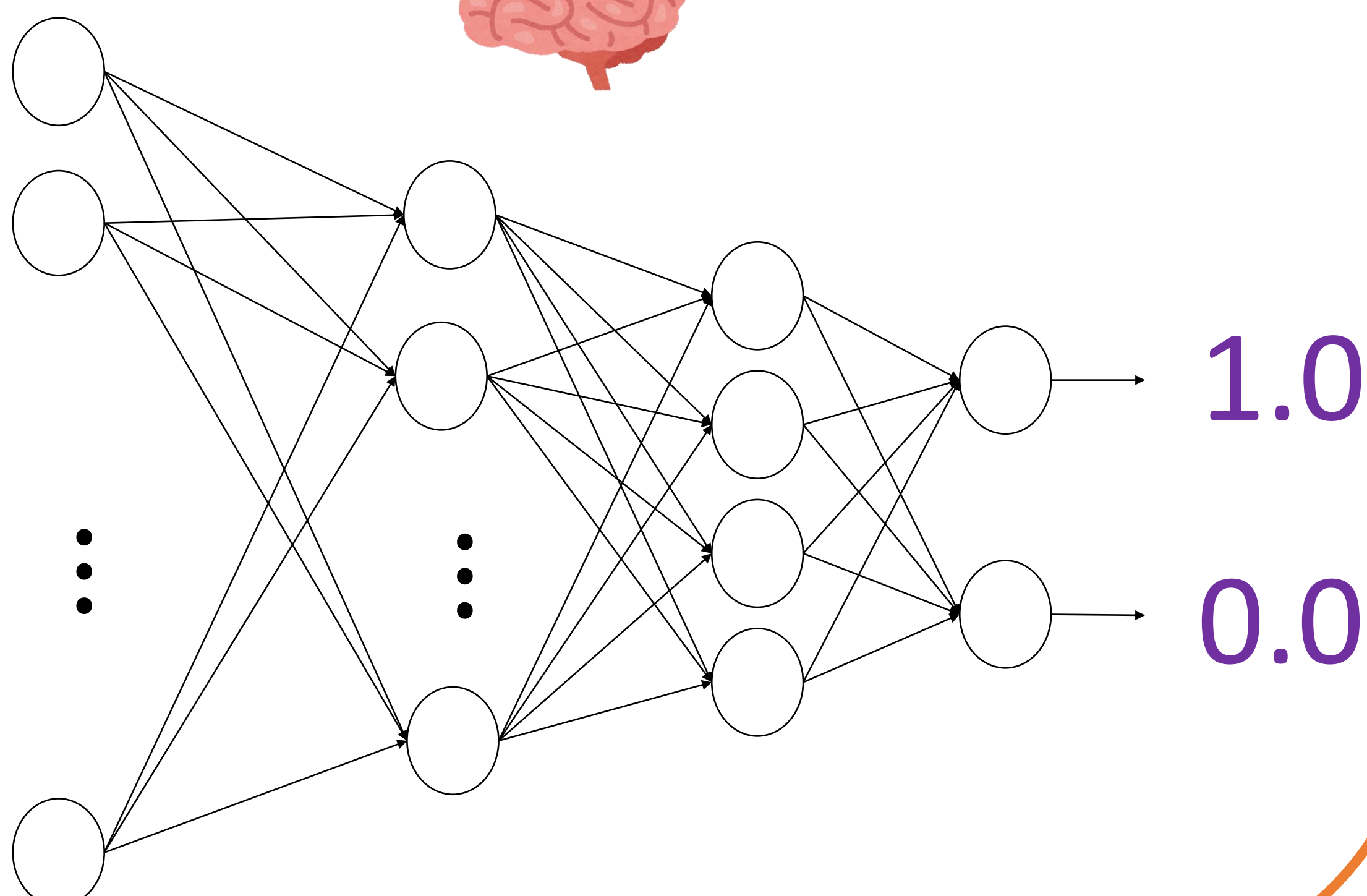


ニューラルネットワーク

- 人の脳内にある神経回路網をモデルとした機械学習の一種.
- 教師データ (ベクトル) とそのクラスを受け取って、誤差逆伝搬法により学習を行う.
- 新たなベクトルを受け取ると、そのクラスが何なのかを答えてくれる.

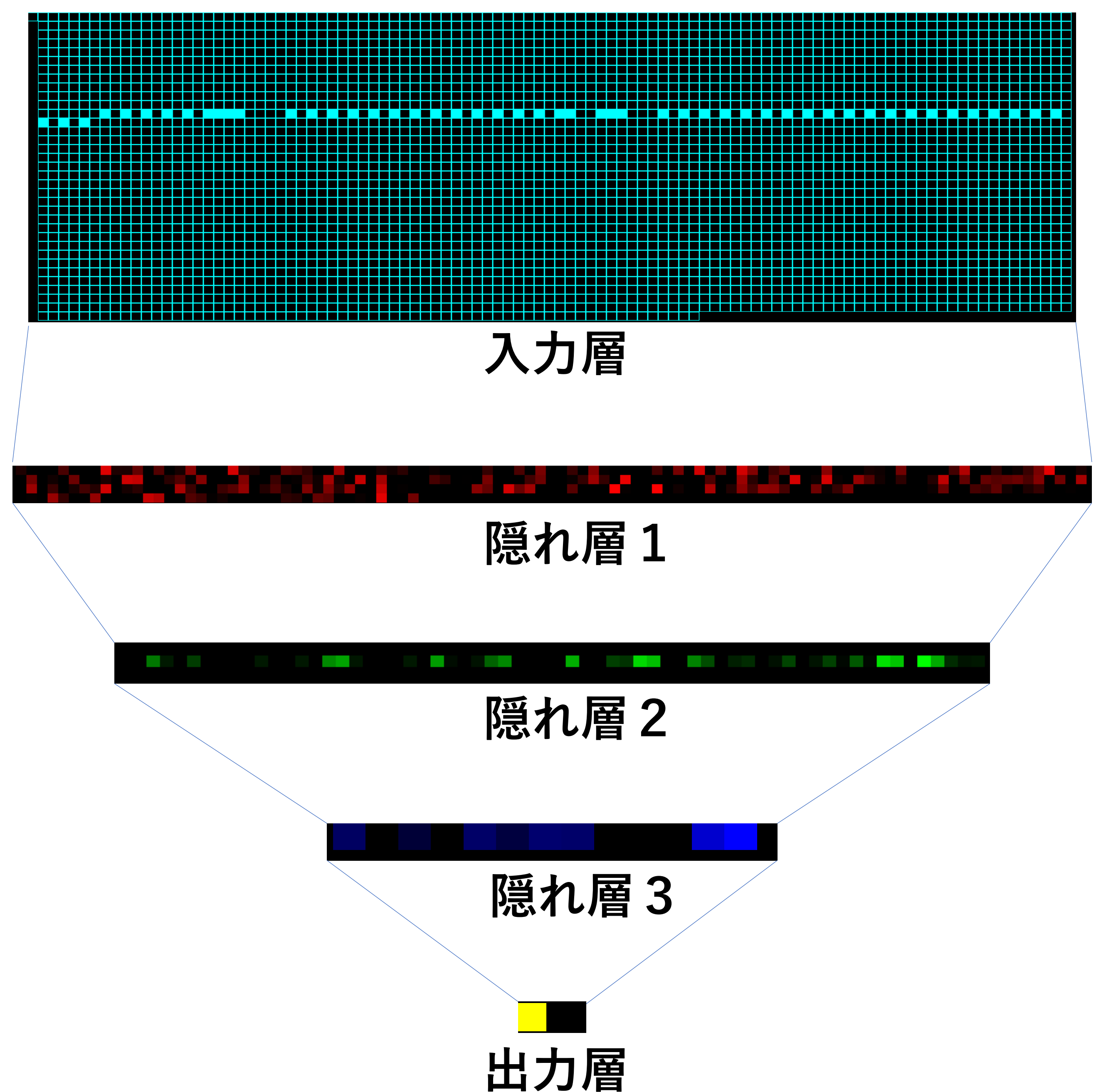


$\begin{bmatrix} 0 \\ 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$



提案手法

- **RDFグラフ**のデータをニューラルネットワークに適用したいが、**RDFグラフ**構造のデータをニューラルネットワークに適用することは不可能.
- そこで、**RDFグラフ**中のリソース (主語s) から**特徴ベクトル**を作成することで、ニューラルネットワークに適用出来るようにすることを提案する.
- 実際には、9つの特徴パターン (述語や目的語など) に注目し、その特徴を持っているなら1, 持っていないなら0とするような**特徴ベクトル**を作成する.



- さらに、特徴の数が大きすぎると**ベクトル**の次元が大きくなり、**ベクトル**の次元が大きくなると学習に時間が掛かり過ぎてしまうため、特徴毎の情報利得率を用いて学習に使用する特徴を事前に選別する.

まとめ

- **RDFグラフ**のリソースをニューラルネットワークで**クラス**分類できるようになった
- 今後は分類モデルの精度を上げるために**特徴ベクトル**の作り方を工夫したい