

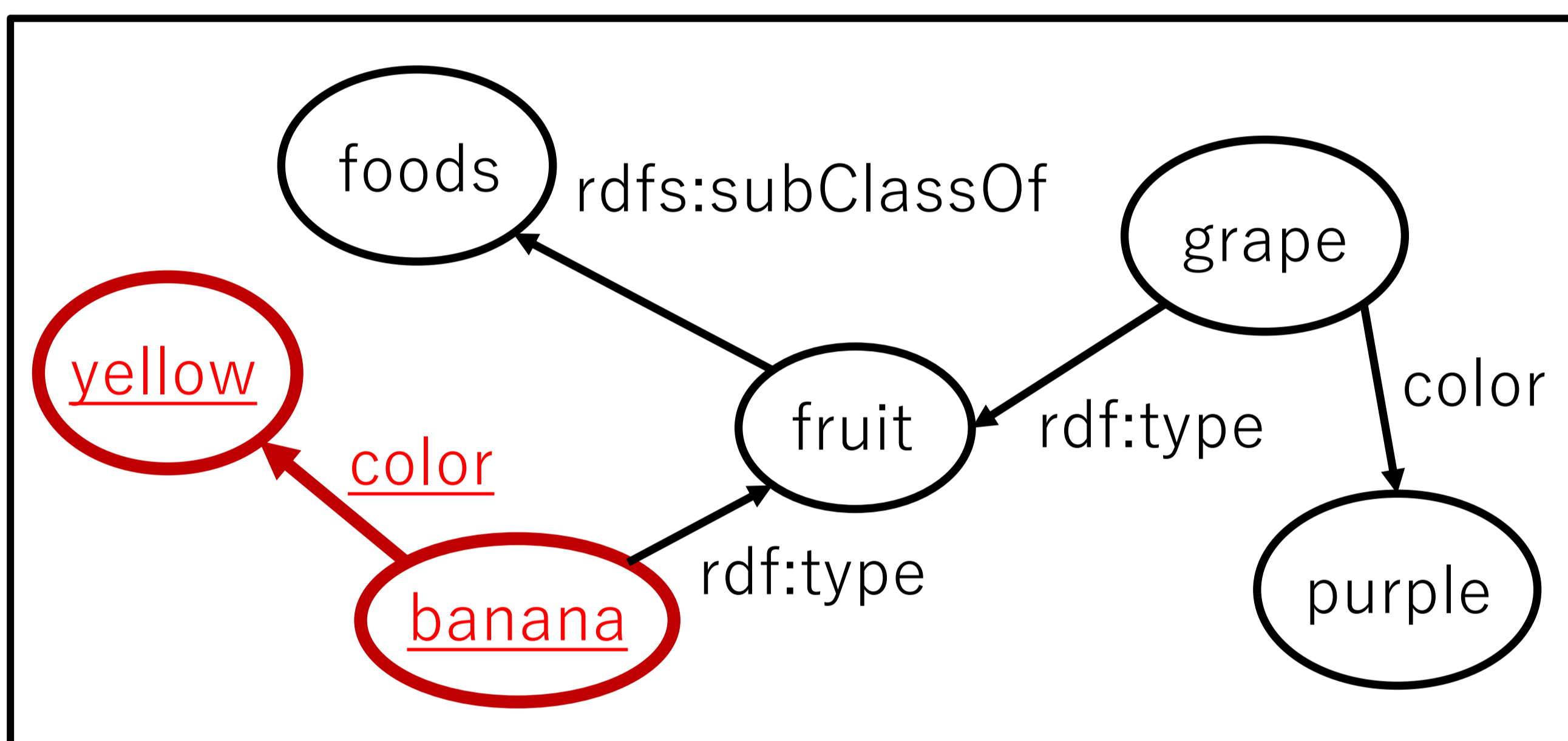
グラフニューラルネットワーク によるRDFデータの教師あり学習

情報・ネットワーク工学専攻 兼岩研究室 大石悠河

RDFデータ

- 主語 s, 述語 p, 目的語 o からなる三組みであるRDFトリプル(s, p, o)の集合として表現されるデータ形式
- RDFトリプルの例
(banana, color, yellow)
(banana, rdf:type, fruit)
(grape, color, purple)
- RDFデータはグラフ構造として表現することも可能(RDFグラフ)

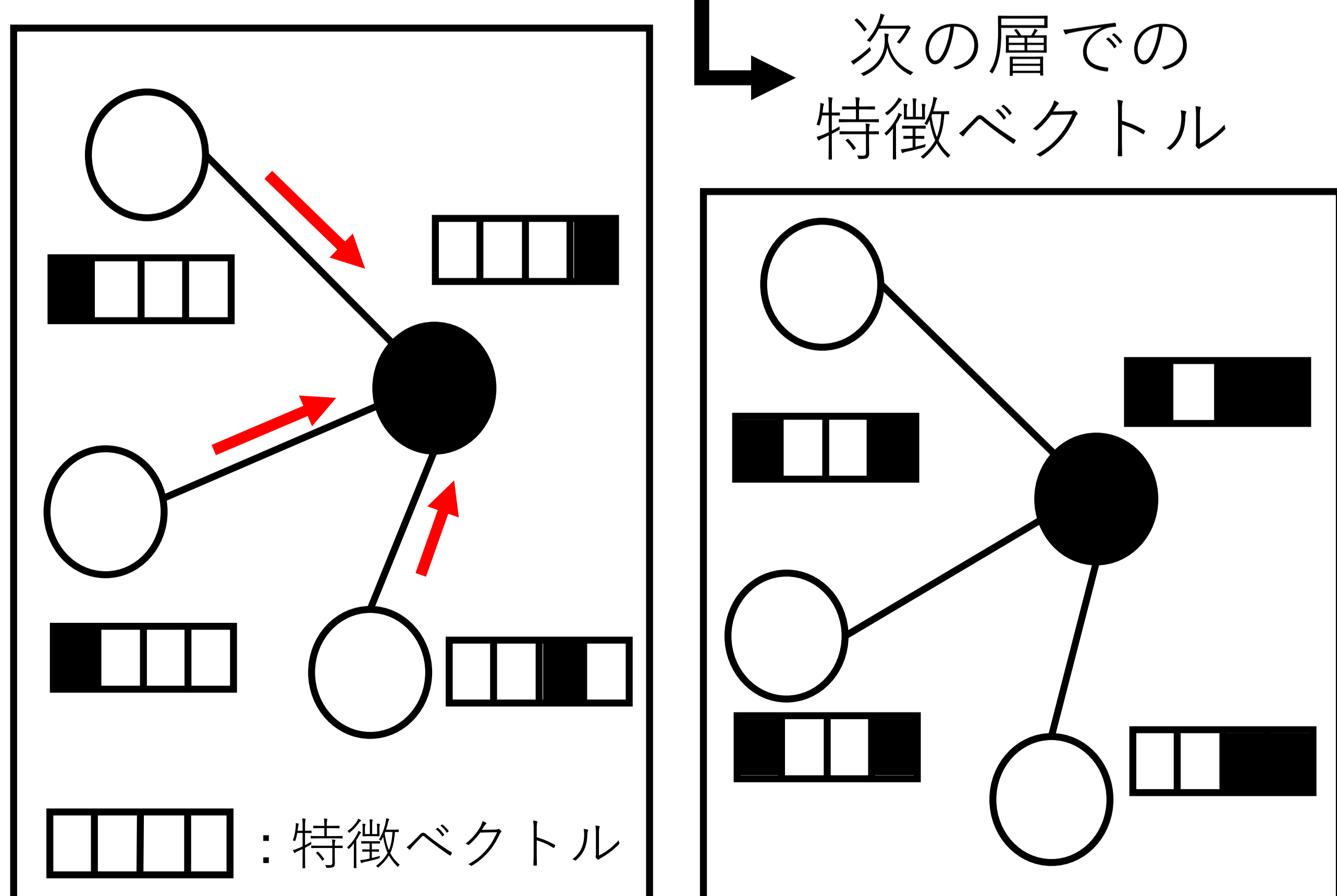
RDFグラフの例



グラフニューラルネットワーク (GNN)

- グラフデータの持つ構造**と, グラフ上のノードが持つ**特徴ベクトル**を用いることでグラフ情報を学習する機械学習手法の枠組みの1つ
- 各ノードが持つ特徴ベクトルはそれぞれのノード特有の性質を表現している
- GNNは各層ごとに近傍ノードの特徴ベクトルを集約していくことで学習を進める

集約のイメージ

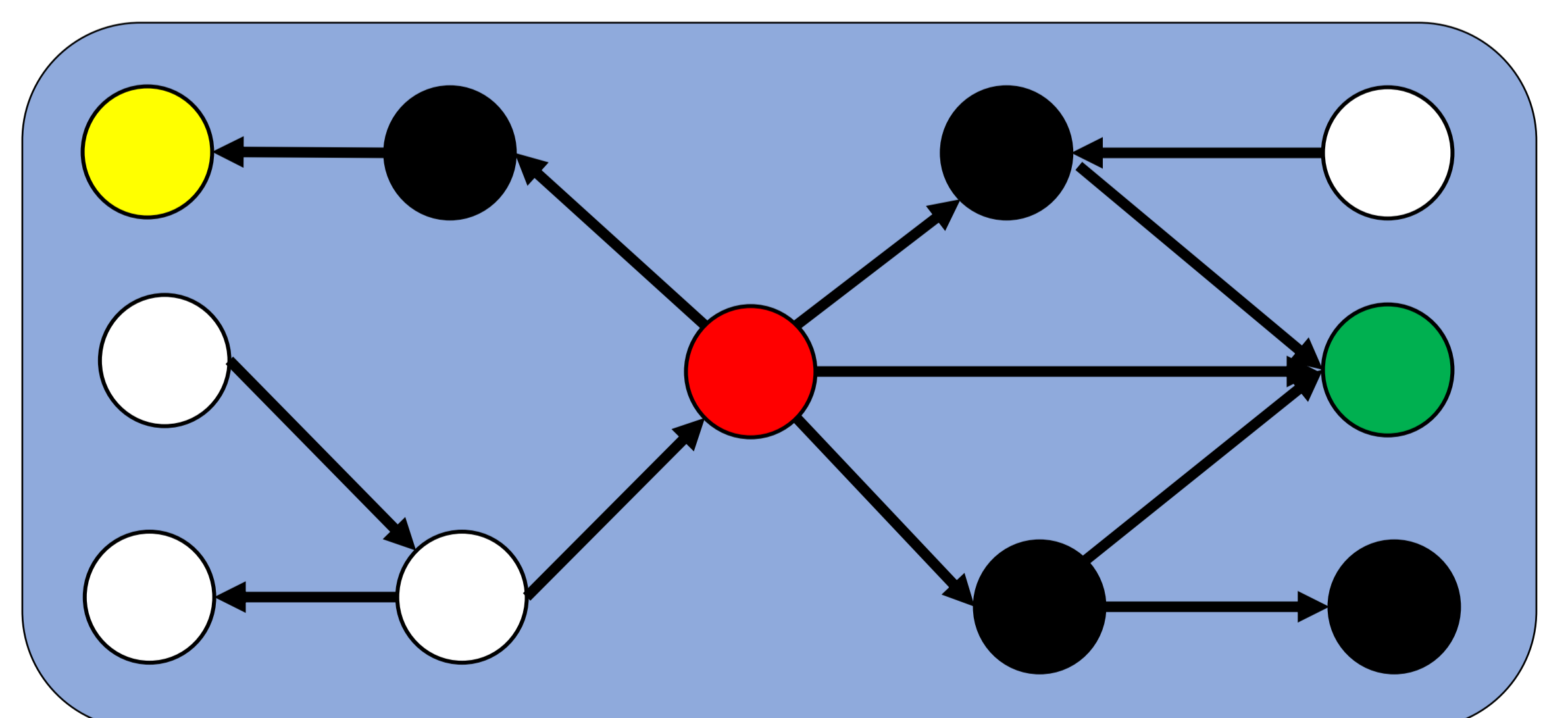


問題点

- RDFグラフ上のノードは特徴ベクトルを持たないため, GNNに対してそのままRDFデータを適用することができない

提案手法

- 特徴ベクトルの作成**
 - RDFデータの述語 p や目的語 o を特徴として扱う
 - ノードが p や o を持っていれば1, 持っていなければ0としてベクトルを作成する
 - 特徴ベクトルの例:
banana = [1(color), 1(fruit), 0(purple)]
- 特徴の制限**
 - RDFデータ上の全ての述語, 目的語を特徴として用いてしまうと計算コストが非常に大きくなってしまふ
 - 特徴の中には**学習に悪影響**を与えるものもある
 - 情報利得率**や**総経路数**を用いて特徴を制限する
- 情報利得率**
 - 特徴がクラス分類に有用であることを示す指標
- 総経路数**
 - 各訓練データから隣接ノードを辿り目的語まで到達できる経路の合計
 - 総経路数の小さい目的語は学習で重要ではないと考えられる



- 訓練データ ● 対象ノード: 総経路数 1 ● 対象ノード: 総経路数 3

今後の課題

- ノードや述語, 目的語に対するより有用な特徴の選別方法を見つける
- より性能の良いGNNを考案する