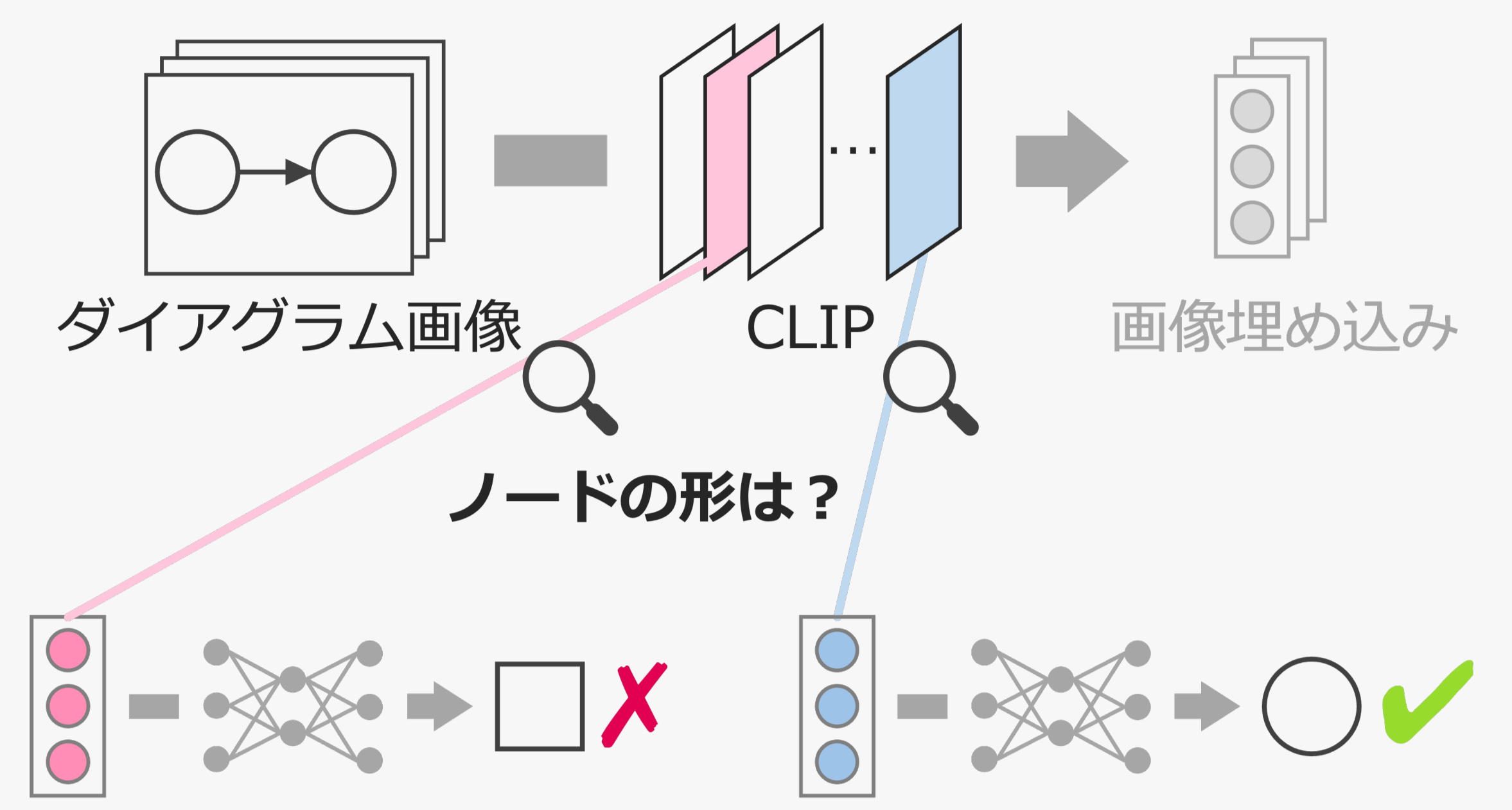


# 自然画像で学習された画像埋め込みに ダイアグラムを特徴づける情報は含まれているか？

吉田 遥音<sup>1,\*</sup>, 工藤 慧音<sup>1,2</sup>, 青木 洋一<sup>1,2</sup>, 田中 涼太<sup>1,3</sup>, 斉藤 いつみ<sup>1</sup>, 坂口 慶祐<sup>1,2</sup>, 乾 健太郎<sup>4,1,2</sup>  
 ( <sup>1</sup> 東北大学, <sup>2</sup> 理化学研究所, <sup>3</sup> NTT人間情報研究所, <sup>4</sup> MBZUAI ) \*yoshida.haruto.p1@dc.tohoku.ac.jp

## 概要

- CLIPの画像埋め込み（隠れ状態）にノードやエッジといったダイアグラムにおける基本的な要素の情報が含まれるか？
- CLIPの隠れ状態に対して分類によるprobingを行った
- CLIPがエッジの向きを認識できない可能性が明らかになった



## 背景

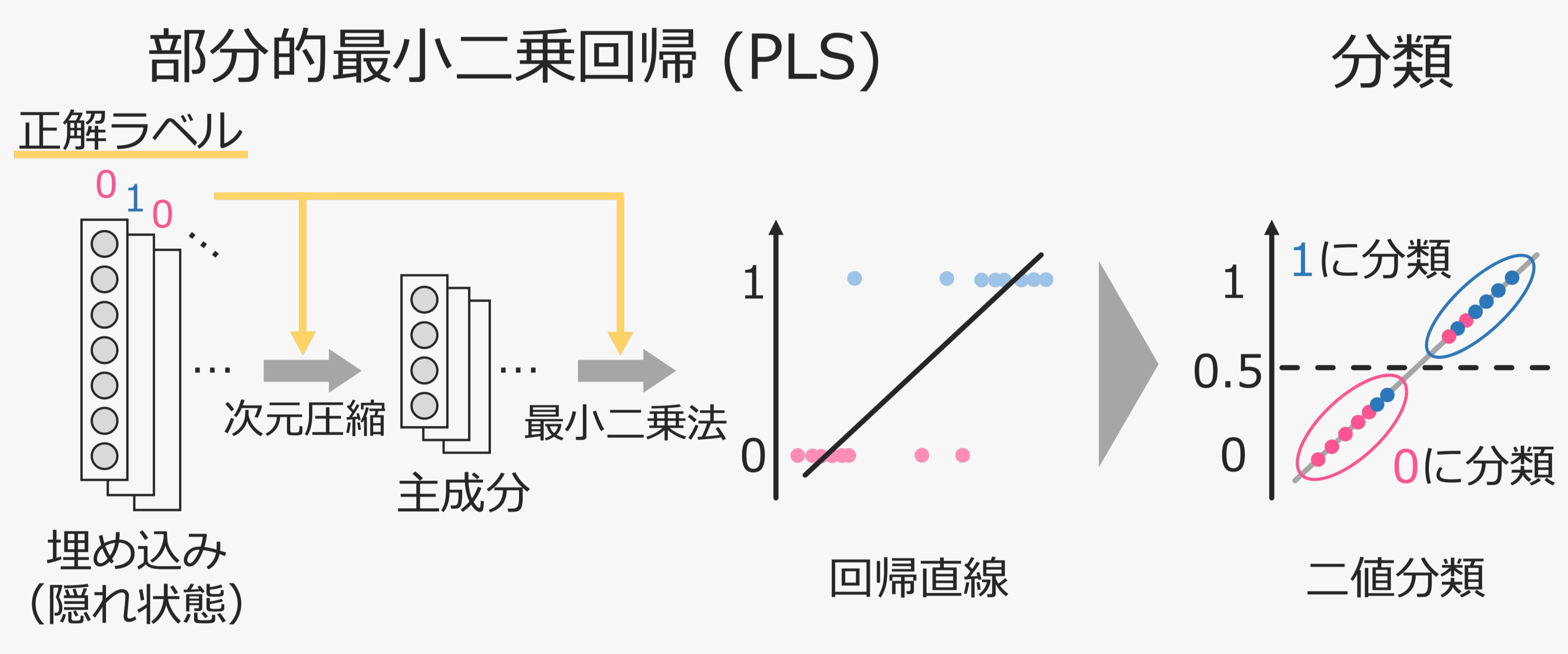
ニューラルモデルによるダイアグラムの理解を実現したい

- CLIPは言語と画像の両方を扱える
- CLIPの学習データの多くは自然画像

CLIPの隠れ状態は  
ノードやエッジの情報を含んでいるか？

- YES** → 検索や自動生成の評価に応用
- NO** → それらを認識できるモデルの実現

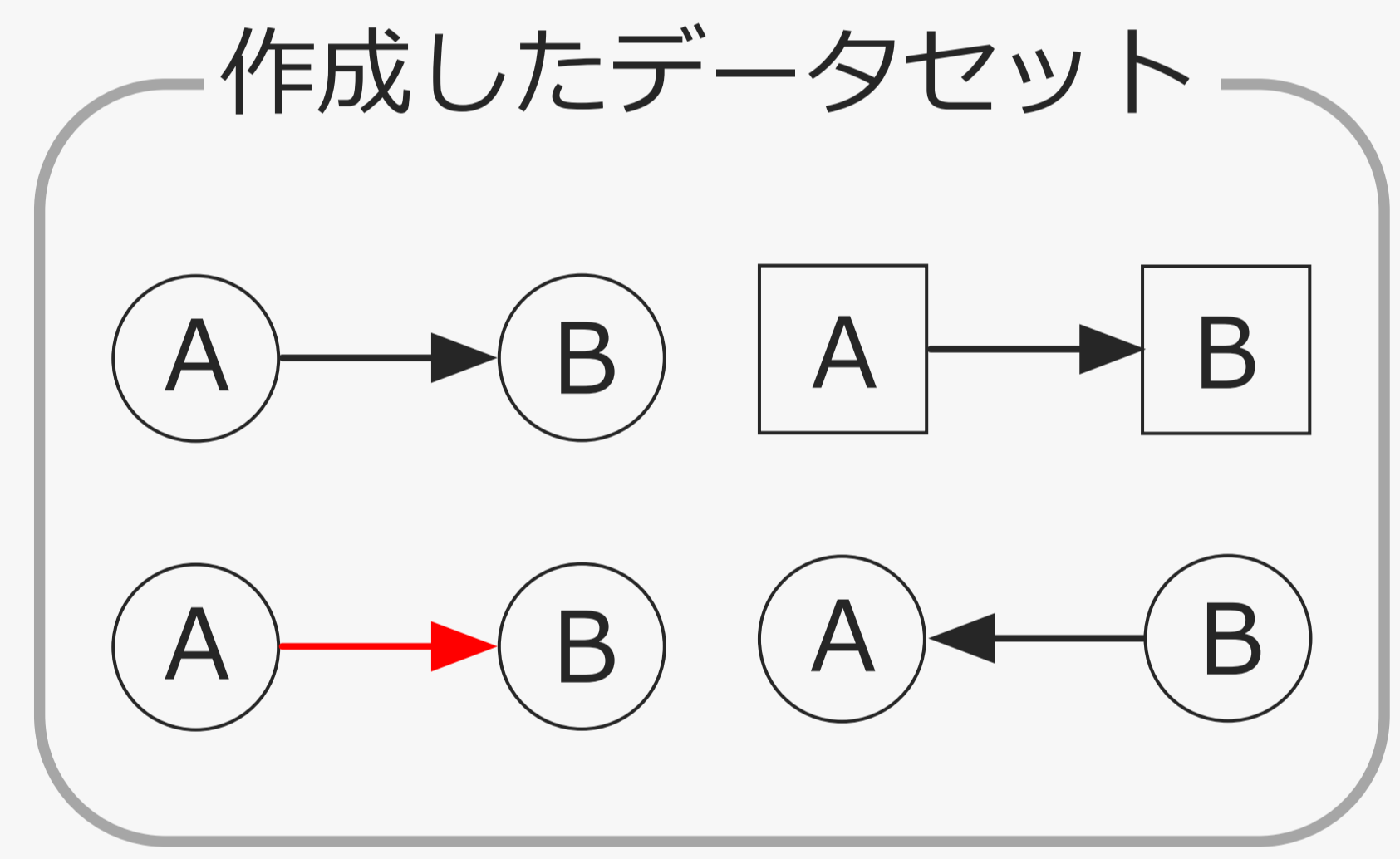
## 分析手法



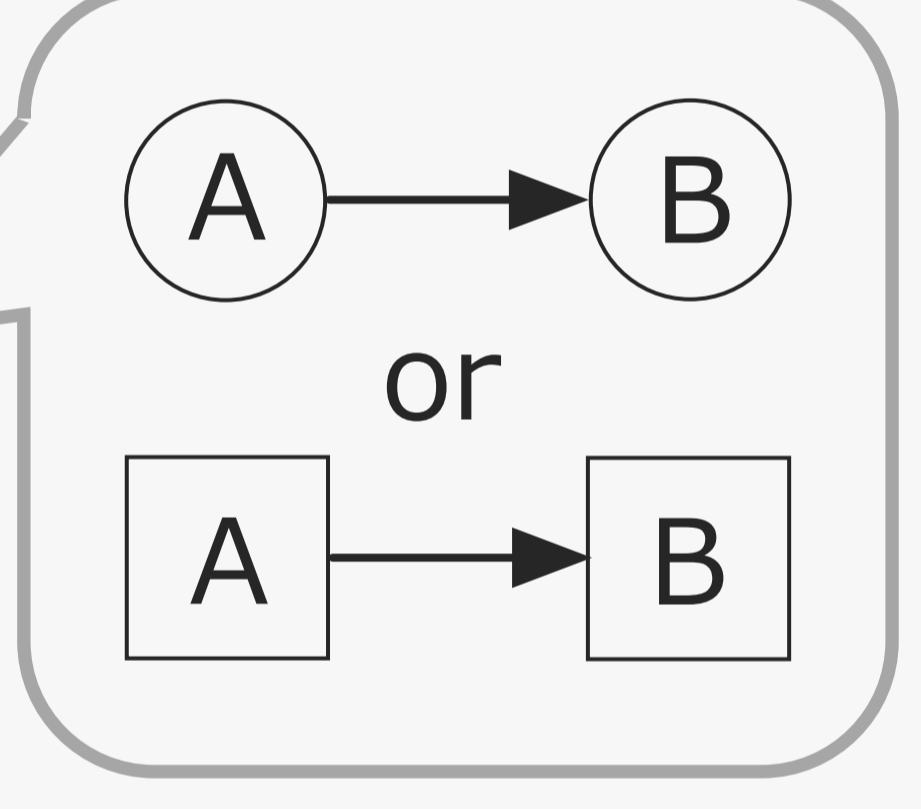
## 実験設定

データセット

- ノードが2つの有向グラフ
- 各要素を変化させて作成

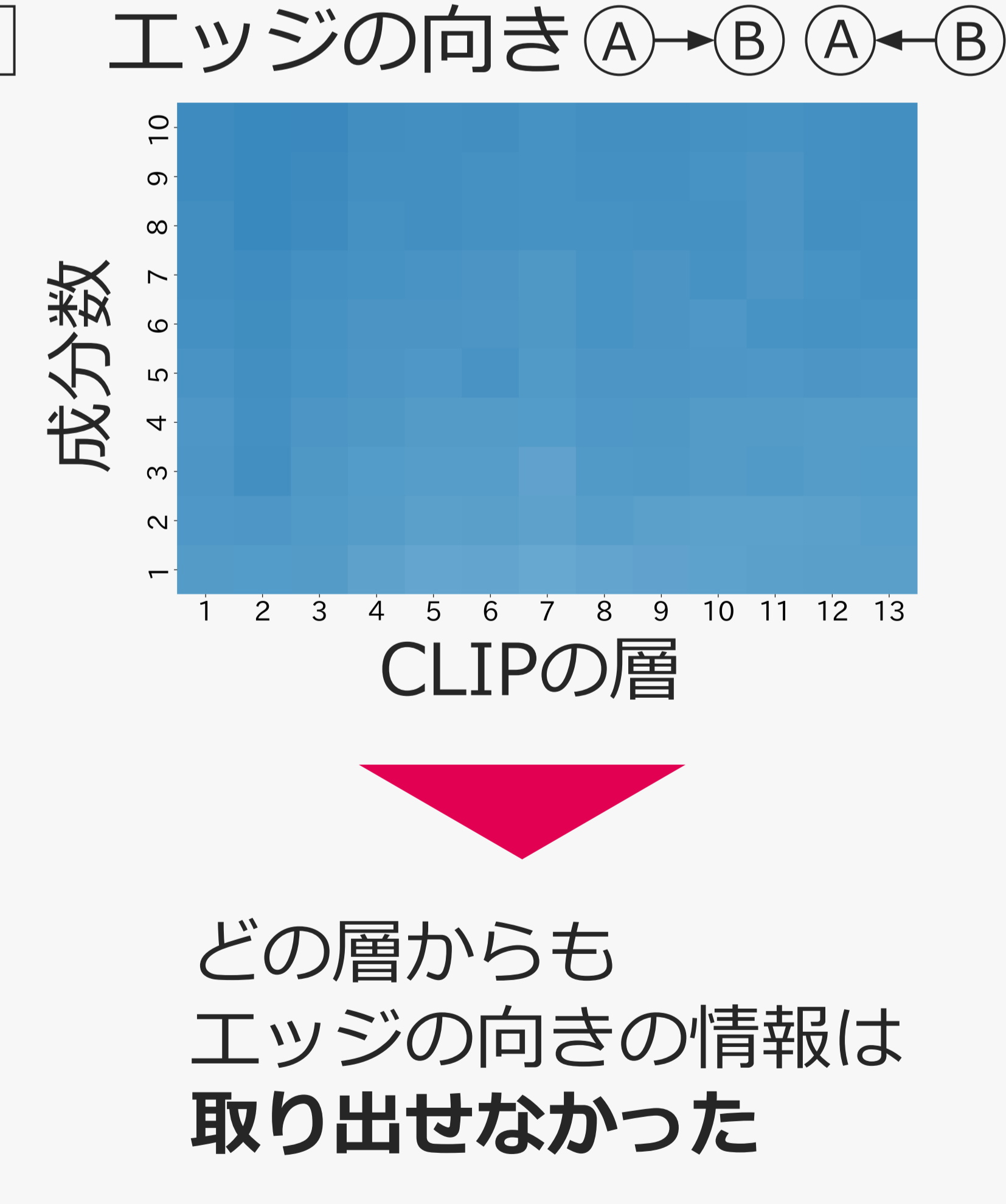
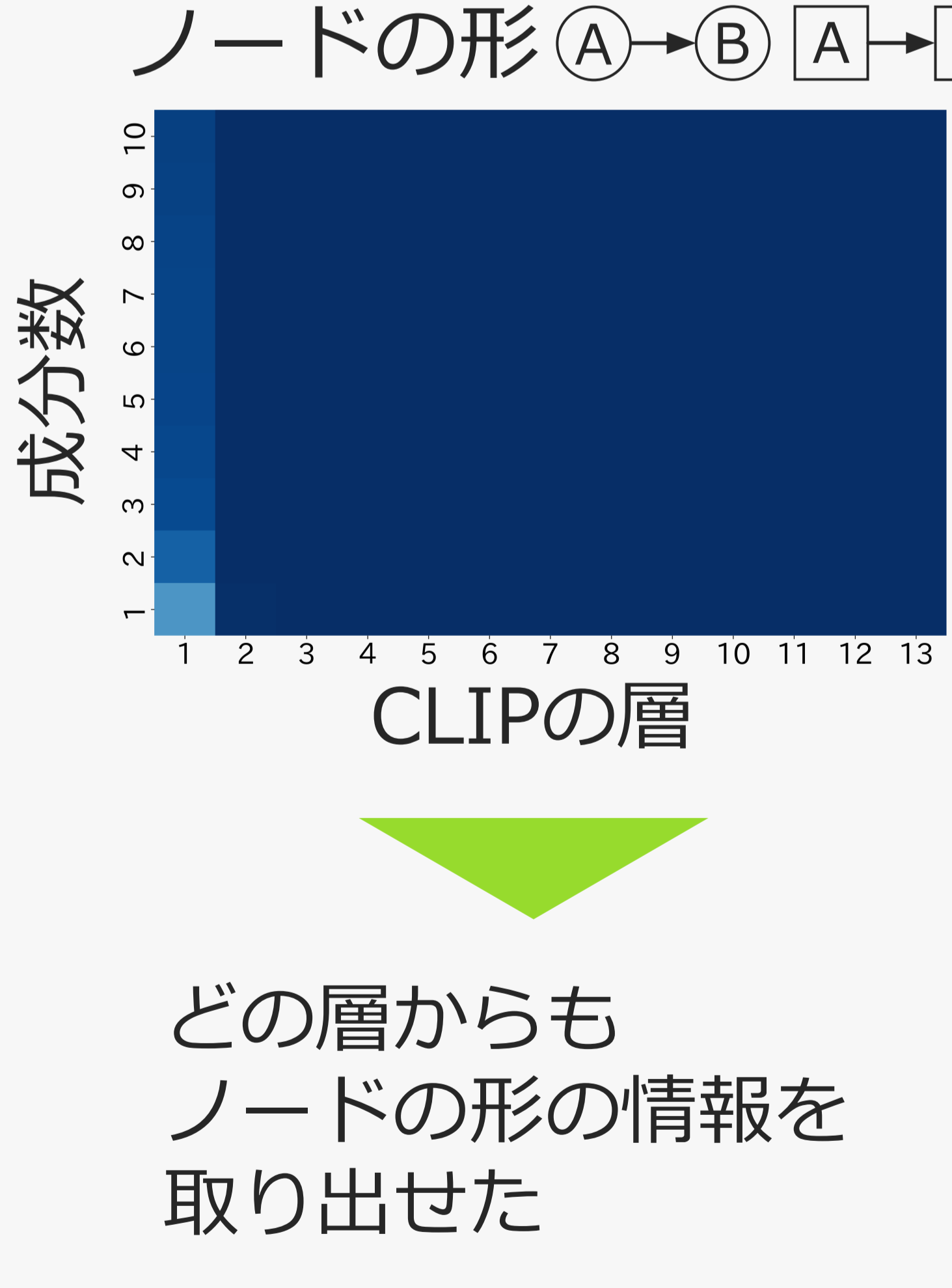
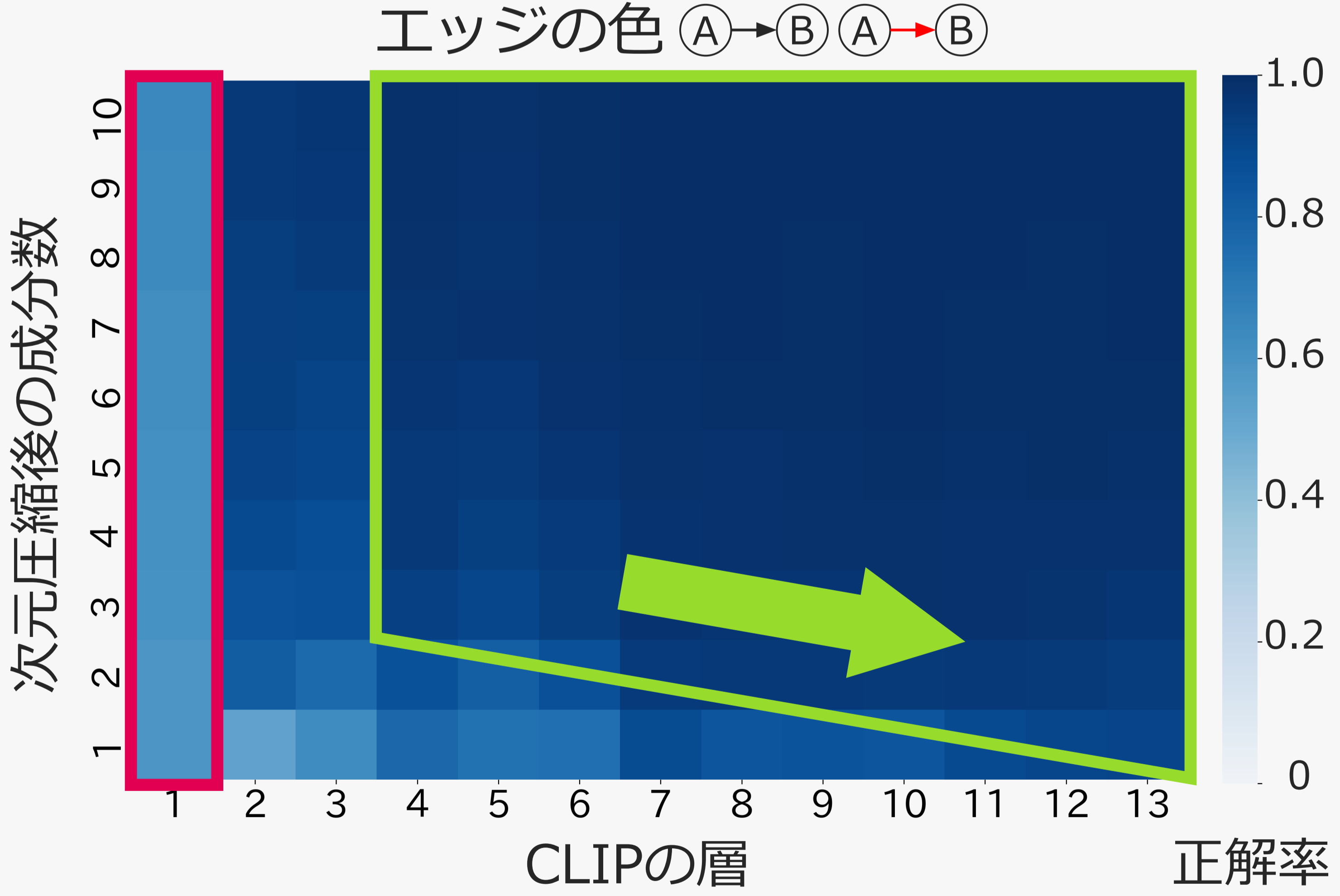


- ノードの色
- ノードの形
- エッジの色
- エッジの向き



で分類可能か？

## 結果・分析



- : 成分数によらず正解率が低い  
→ 情報は取り出せなかった
- : 成分数が少なくても正解率が高い  
→ 徐々に**低次元の部分空間**にエンコード

エッジの向きの情報は埋め込み（隠れ状態）に含まれていない可能性がある

- 今後は…
- エッジの向きを認識できるモデルの実現
  - テキストとの意味の対応を実現