

# 2022年度 修士論文

## 時系列・非時系列属性の混在データ に対する埋込モデルに関する研究

---

広島大学大学院 先進理工系科学研究科

情報科学プログラム

M216916 稲葉勇哉

---

**研究概要**

**先行研究**

**提案モデル（GDN応用モデル）**

GDN応用モデルを構築し  
株価の騰落を分類予測した

モデルについて説明する

**実験設定**

**実験結果**

**まとめ**

モデルの性能について説明する

# 研究概要

時系列属性と非時系列属性が混在したデータにおいて  
分類モデルを構築するとき・・・

## 【実務上の課題】

時系列属性の特徴量抽出（前処理）をして  
非時系列属性として扱って分類モデルで学習  
➡ 精度の低下が懸念される



## 【本研究の目的】

① 時系列属性の特徴量抽出    ② 分類器の学習  
を一連の学習として行う分類モデル  
（End-to-Endモデル）の構築  
➡ 「時系列属性の特徴量抽出」まで逆伝播できる

## 【End-to-Endモデルの処理イメージ】

企業	時系列属性（株価）				非時系列属性			
	●月●日	●月●日	●月●日	...	財務 指数	売上 規模	業種	...
A社	〇〇	〇〇	〇〇	...	〇〇	〇〇	〇〇	...
B社	△△	△△	△△	...	△△	△△	△△	...
C社	□□	□□	□□	...	□□	□□	□□	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

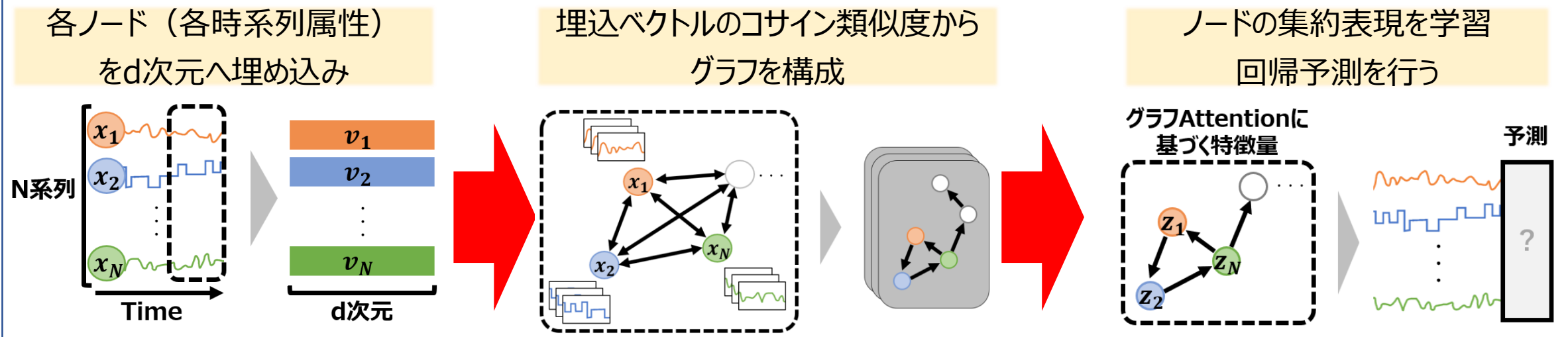
① 時系列属性の  
特徴量抽出

② 分類器の学習

2段階の処理を  
一連の学習として行う

# 先行研究

## 【GDN : Graph Deviation Network (多変量時系列モデル)】



## 【損失関数 : Dice Loss】

$N$  : サンプル数

$K$  : クラス数

$p_{ij}$  : 確率

$y_{ij}$  : 教師データ

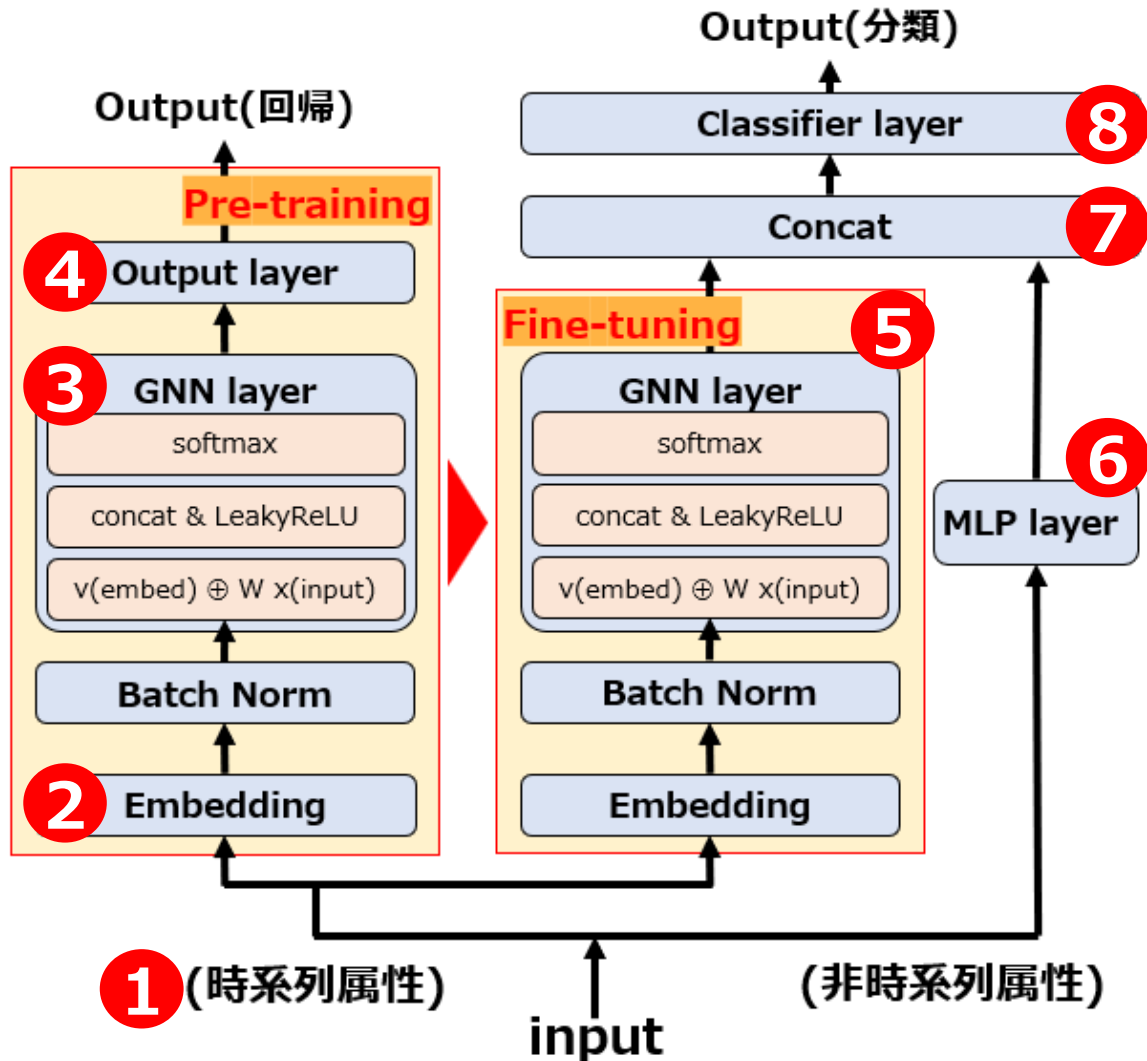
$\gamma$  : ハイパーパラメータ

$$\text{Dice Loss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K \left( \frac{2 p_{ij} y_{ij} + \gamma}{p_{ij} + y_{ij} + \gamma} \right)$$

- F値を応用した損失関数
- 特に不均衡データの分類予測においてバランスよく予測する

# 提案モデル（GDN応用モデル）

多変量時系列モデル：Graph Deviation Networkを用いた  
**GDN応用モデル（End-to-End 転移学習モデル）を構築**



**GDN応用モデルは  
次の2つ（転移学習）で構成される**

## 【Pre-training（回帰）】

- ① 時系列属性（部分時系列）を入力
- ② Embedding層により埋込ベクトルを出力
- ③ GNN層でノード集約表現を出力
- ④ ②埋込ベクトルと③ノード集約表現から回帰

## 【Fine-tuning（分類）】

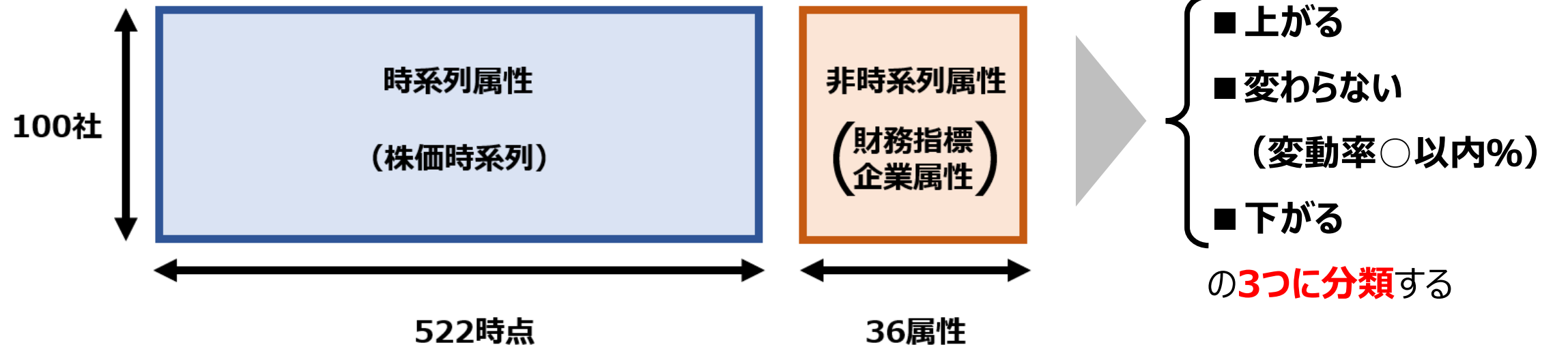
- ⑤ ②～④を再学習
- ⑥ 非時系列属性をMLP※層で集約
- ⑦ ⑤と⑥の結果を統合
- ⑧ 分類予測

※MLP（Multi-layer perceptron）：多層パーセプトロン 4

# 実験設定（使用データ）

## 【使用データ】

東証上場企業のうち一定条件を充足した  
企業からランダムに100社選択



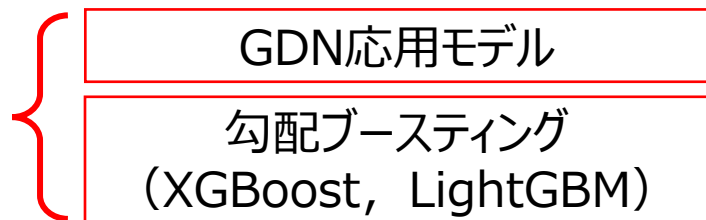
# 実験設定（実験① および 実験②）

## 【実験① 従来手法（勾配ブースティング）との比較】

クラス … 上がる／変わらない（変動率3%以内）／下がる  
約 1 : 1 : 1 に分割

損失関数 … Cross Entropy

### モデルを比較



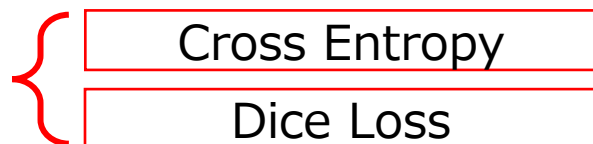
勾配ブースティングは  
前処理が必要  
➡ GDN回帰予測の  
埋込表現を使用

## 【実験② 損失関数 Dice Loss の適用】

クラス … 上がる／変わらない（変動率7%以内）／下がる  
約 1 : 3 : 1 に分割（不均衡データ）

モデル … GDN応用モデル

### 損失関数を比較



Dice Loss は不均衡データ  
において、各クラスをバランスよく  
予測することが期待される

## 【評価指標】

$$\text{Accuracy} = \frac{1}{N} \sum_{k=1}^K \text{TP}_k$$

➡ 正解率を測る

macro-F1

$$= \frac{1}{K} \sum_{k=1}^K \frac{2 \cdot \text{Precision}_k \cdot \text{Recall}_k}{\text{Precision}_k + \text{Recall}_k}$$

➡ 各クラスをバランスよく予測  
できているかを測る

# 実験結果①

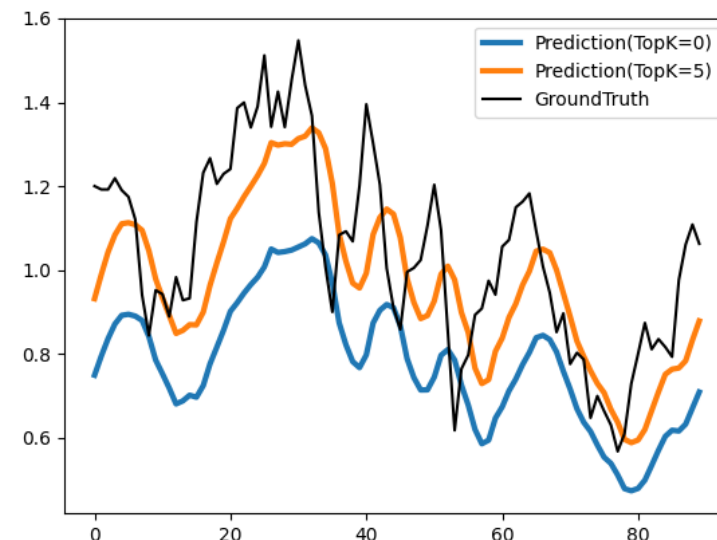
## 【実験① 勾配ブースティングとの比較の結果】

モデル	Accuracy		マクロ平均F値	
	平均	標準偏差	平均	標準偏差
GDN応用モデル	0.3674	0.0133	0.3466	0.0243
GDNプラグインモデル	0.3581	0.0090	0.3316	0.0162
GDN（※時系列のみ）	0.3619	0.0183	0.3345	0.0272
GDN（※非時系列のみ）	0.3504	0.0022	0.3157	0.0052
XGBoost	0.3523	0.0076	0.3078	0.0131
LightGBM	0.3526	0.0047	0.3117	0.0133

### • GDN応用モデル > 従来モデル

Pre-trainingにおいて  
Topk=5 で周辺ノードの  
取り込みを行っていることが  
回帰性能に寄与している

※Topk=5  
類似性の高い5つのノード  
特徴量を取り込んでいる



### • GDN応用モデル > GDNプラグインモデル※

※Pre-trainingの学習済みパラメータを凍結するモデル

GDN応用モデルの「時系列属性まで逆伝播する効果」あり

### • GDN（※時系列のみ） > GDN（※非時系列のみ）

時系列属性の特徴量抽出が精度向上に寄与している



## 実験結果②

### 【実験② Dice Loss の適用の結果】

GDN応用モデルの 損失関数	Accuracy		マクロ平均F値	
	平均	標準偏差	平均	標準偏差
Cross Entropy	0.4844	0.0853	0.3540	0.0388
Dice Loss	0.5350	0.0292	0.3743	0.0293

- Accuracy（単純な正解率）において  
**Dice Loss > Cross Entropy**
- マクロ平均F値において  
**Dice Loss > Cross Entropy**
- 不均衡データにおいて、損失関数にDice Lossを適用することによる精度向上の効果が示された

# まとめ

## 【総括】

本研究では、実務課題の解決のため**GDN応用モデル（GDNに基づくEnd-to-End 転移学習モデル）を構築**した  
従来手法と比較し、**一定水準の精度を示し、実務への応用の可能性も十分見いだせた**と考えている  
一方で、下記の課題も残っており、更なる性能改善の期待もできる

## 【今後の課題】

- ・ **時系列属性と非時系列属性の統合時の改善**

時系列属性と非時系列属性の相乗効果的な精度向上は十分に実現できなかったため NN層の工夫が必要

- ・ **不均衡データへの対処**

今回は程度の小さい不均衡データでの実験であったが程度の大きい（1:100やそれ以上）場合は学習自体がうまく進まなかった

- ・ **他手法との比較（ARモデルなど）**

時系列属性の特徴量抽出にはARモデルやその変形も適用できるため 比較実験も追加したい