

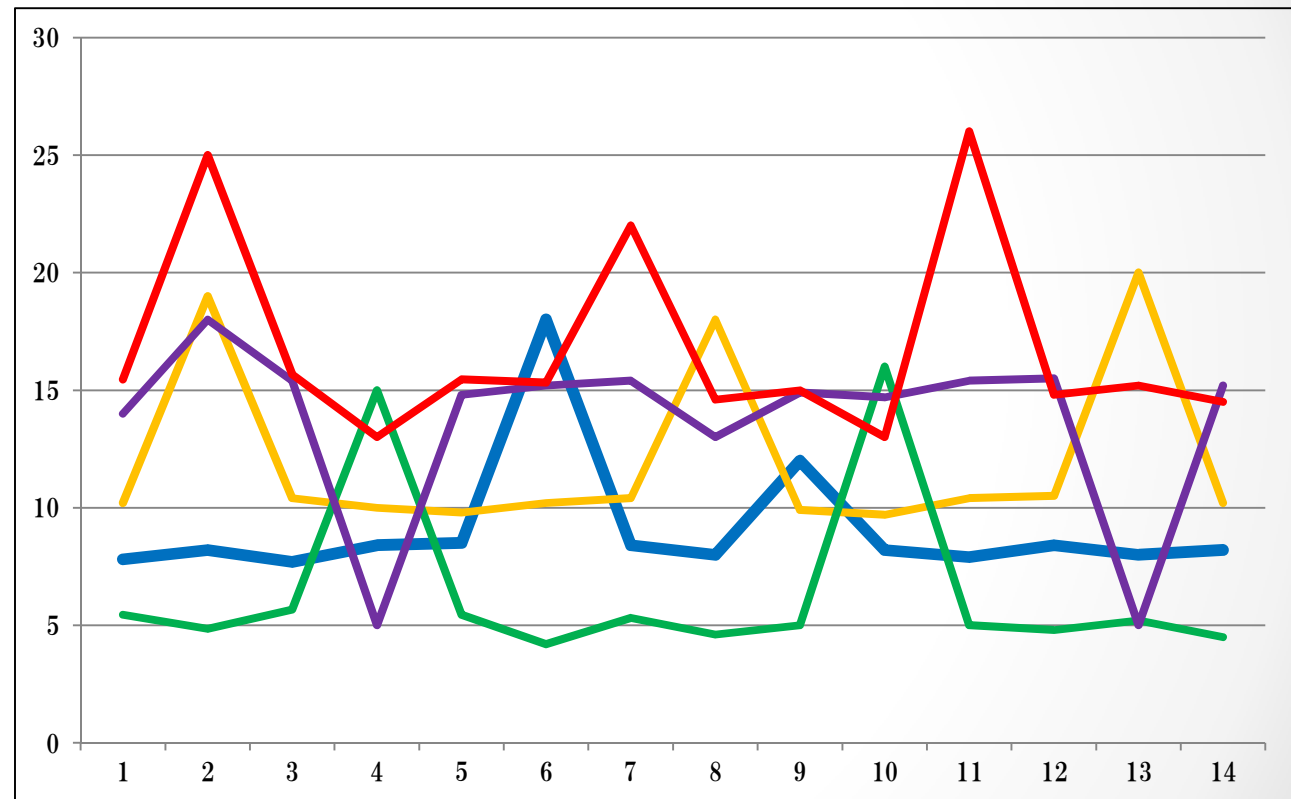
# 多次元時系列データからの 構造パターンマイニング

～区間系列イベント系列&部分グラフ～

鈴木 湧人

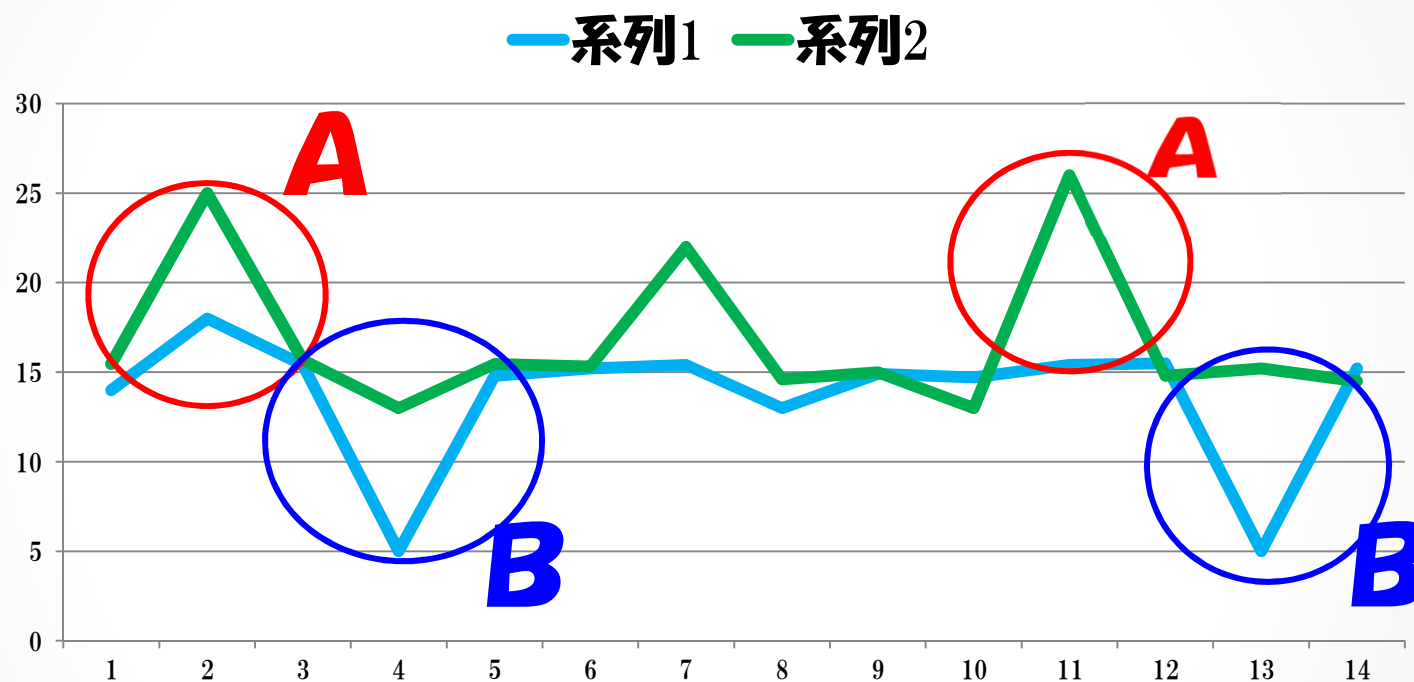
野口 哲哉

# 多次元時系列データ



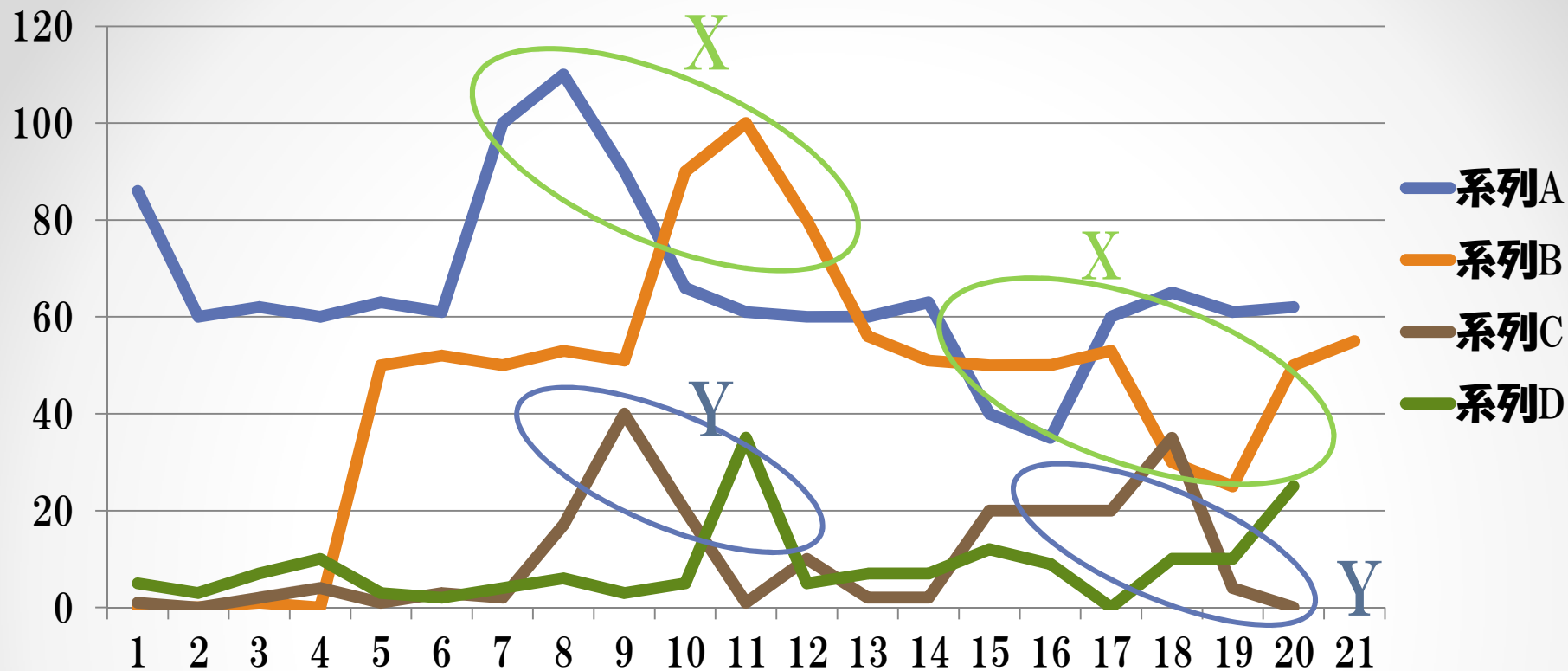
本研究では、**多次元時系列データ**を使用する。

# 一般的な多次元時系列データからの分析



時系列データの特徴的な部分を記号化する

**A**の後に**B**が出てくる



系列Aと系列Bの似ている箇所が

繰り返し出

系列Cと系

繰り返し出てくる

このパターンは  
既存の手法では出せない

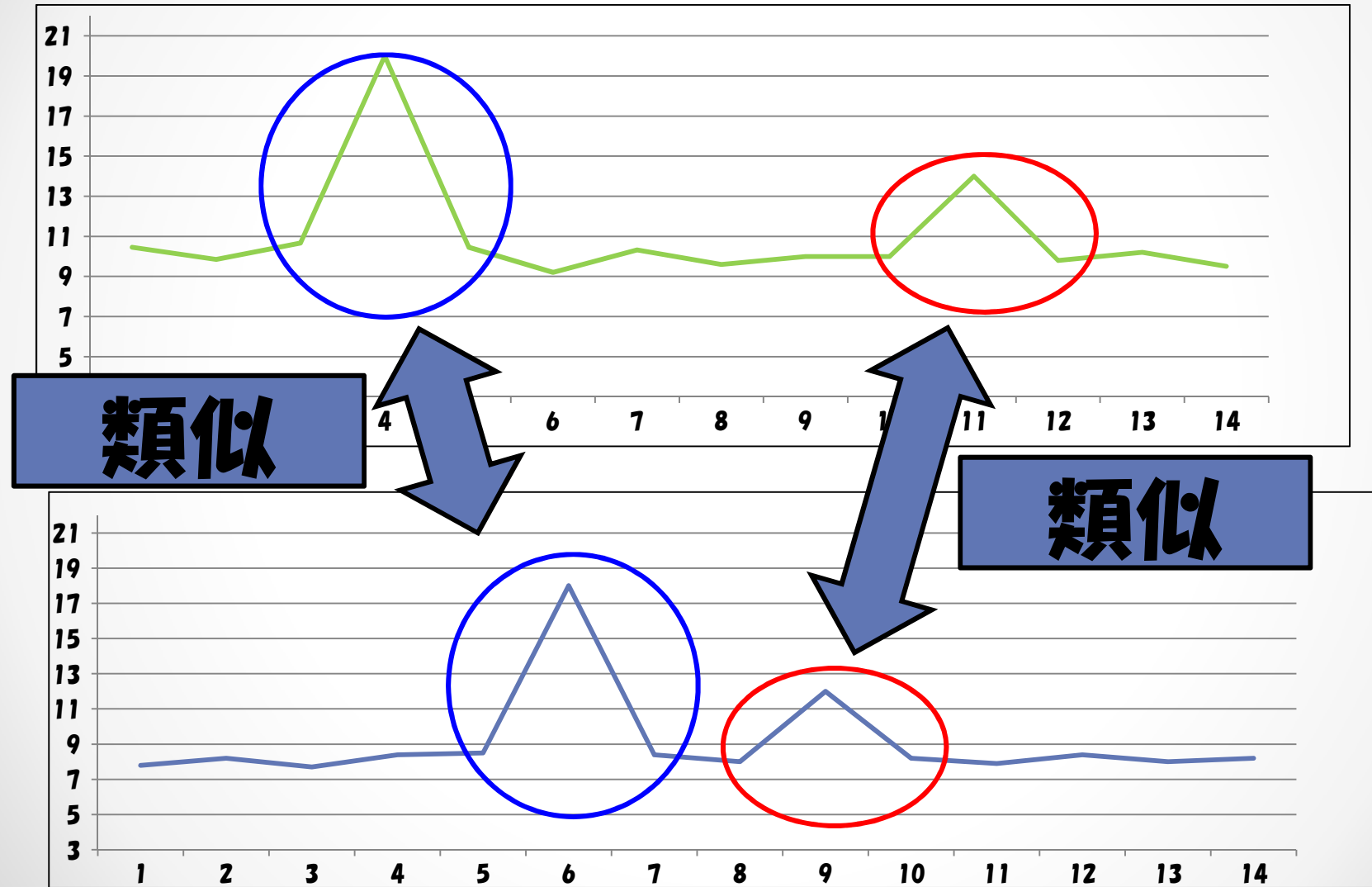
Y

出る

パターン

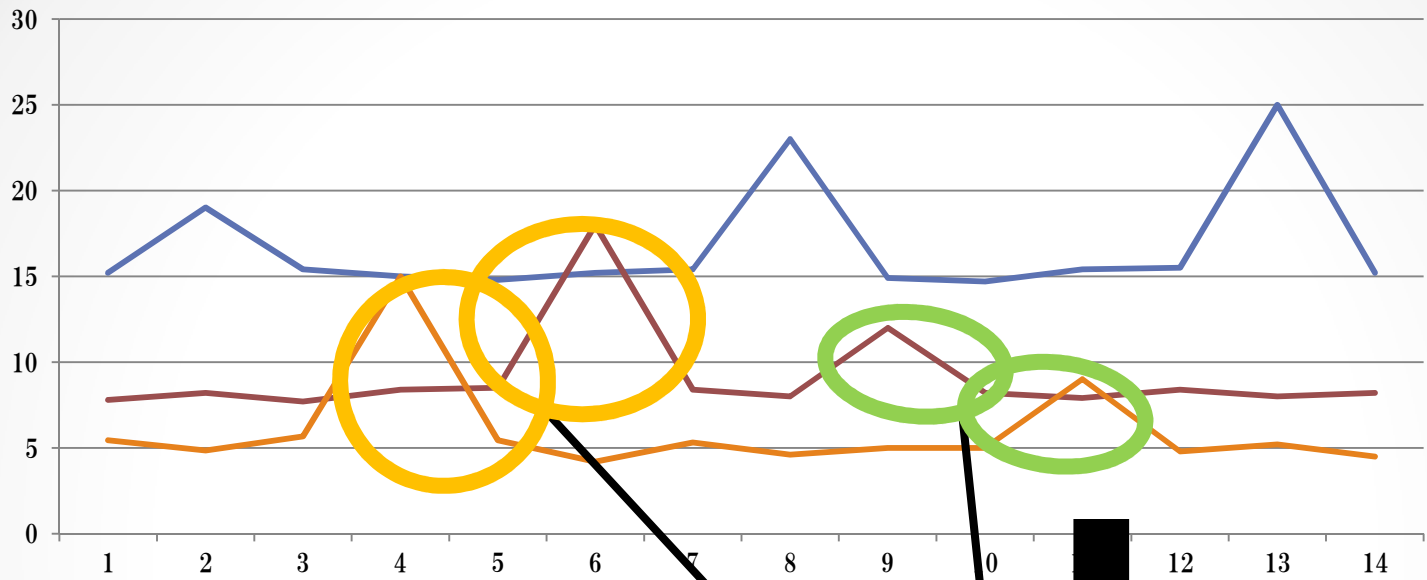
# CrossMatch

## 類似区間の抽出



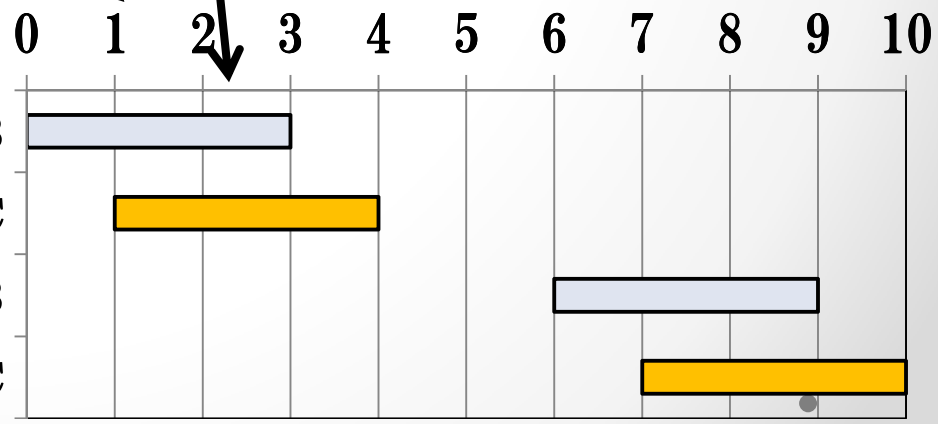
# 多次元時系列データ

# 類似区間抽出



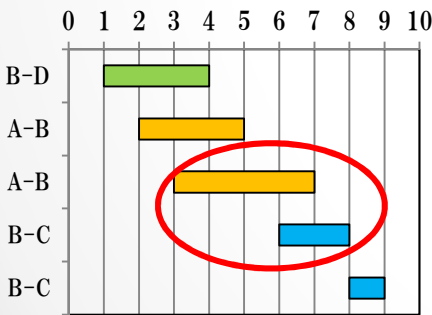
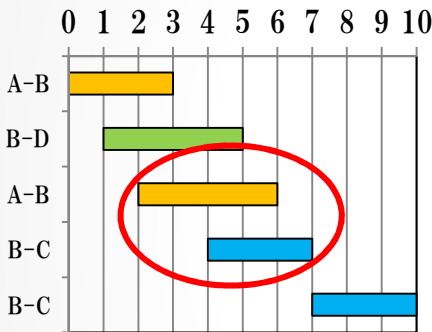
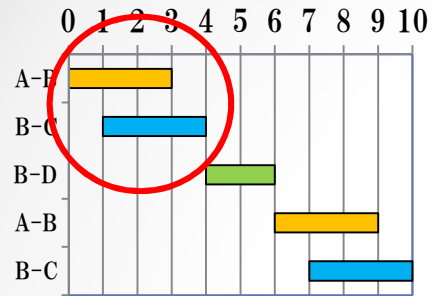
# 区間イベント系列

ある一定の時間幅を考慮した  
区間イベントパターン



グラフ構造と共に、  
順序関係を考慮した  
区間イベント系列パターン

# 関連研究

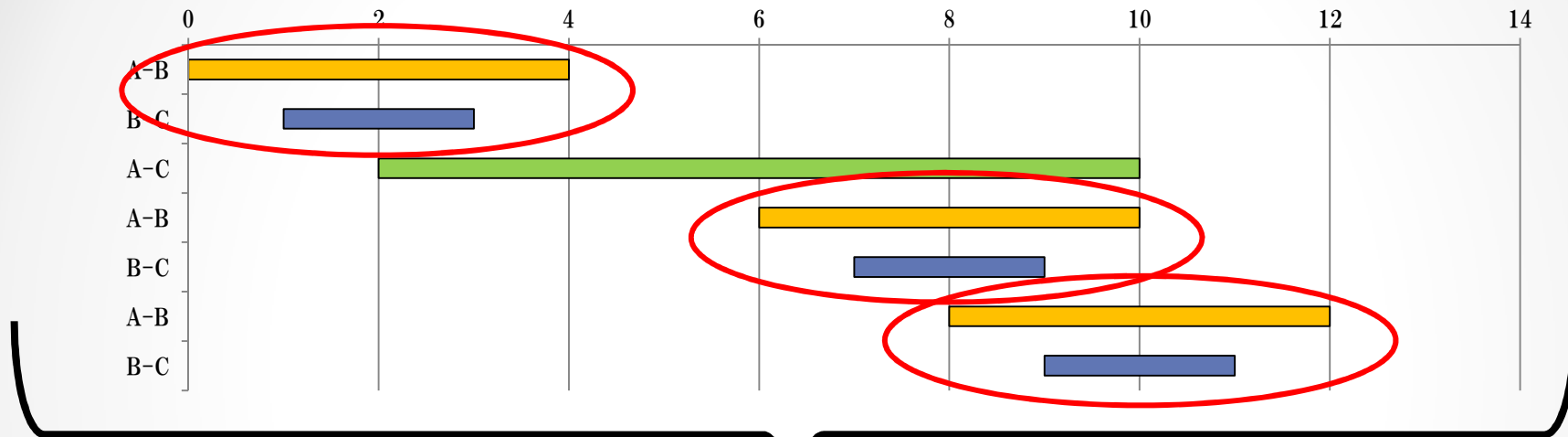


⋮

複数の区間イベント系列

どの区間イベント  
系列にも**共通**に  
含まれるパターンを  
見つける

# 一つの区間イベント系列

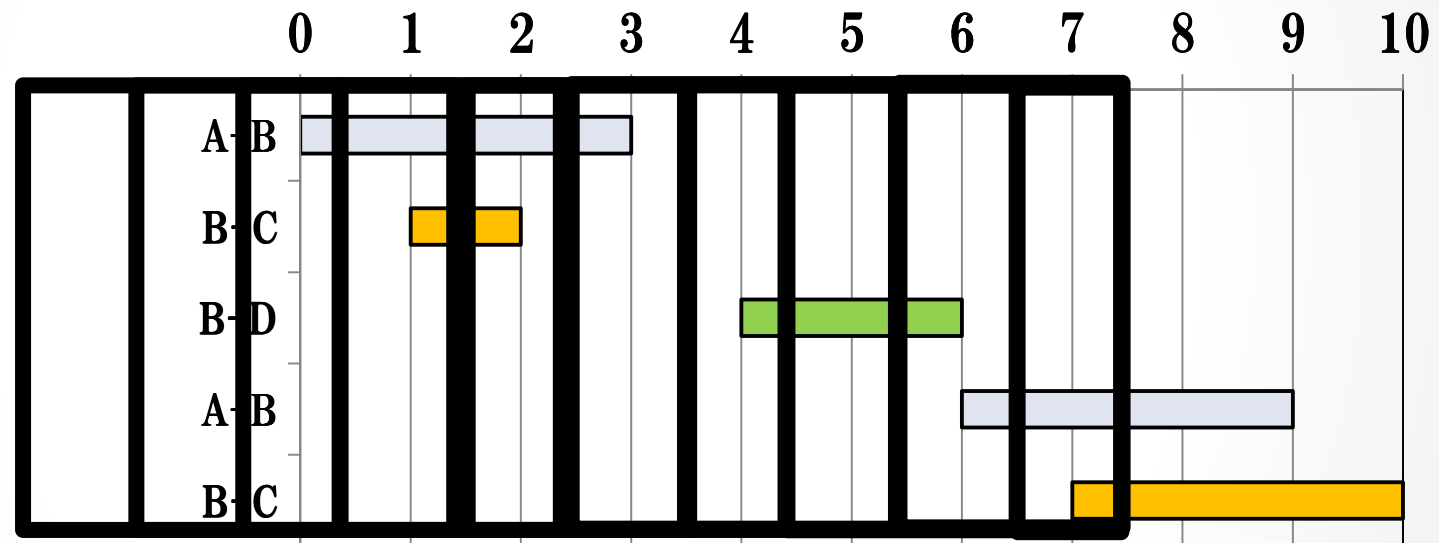


一つの区間イベント系列から  
**繰り返し**現れる  
パターンを発見する



# 時間幅を用いた区間系列パターン生成

系列データを**ウィンドウ**と呼ばれる  
時間幅をスライドさせる



ウィンドウ = 3

# 区間系列でのパターン

[TPrefixSpan] Shin-Yi Wu, Yen-Liang Chen: Mining Nonambiguous Temporal Patterns for Interval-Based Events, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING(2007)

区間イベントの**順序**と**重なり方**で異なるパターンになる

①



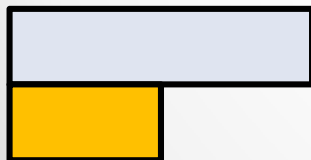
②



③



④



⑤



⑥



⑦



# 系列データの評価関数

問題 AとBが  
それぞれ

数の発音が異なる！  
いくつ含まれているでしょう？

[系列先頭頻度] 岩沼宏治: テキスト系列マイニングにおける有用性尺度について, 人工知能学会誌, 27 巻2 号, (2007)

**A A B B B**

ウィンドウ=3でABの頻度を求める

含む

**A A B**

含む

**A B B**

・ **先頭**を基準に数える

**B B**

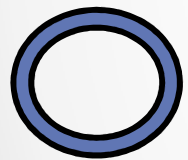
**ABの頻度は2回**

**B**

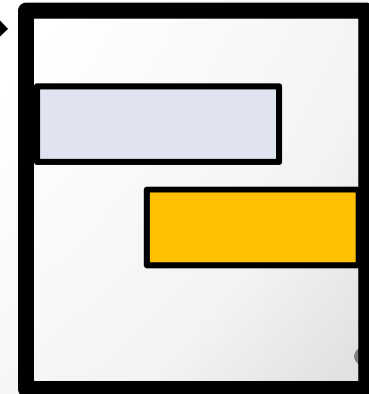
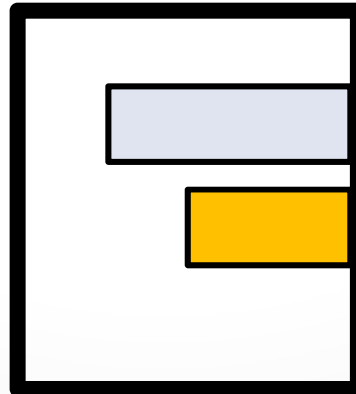
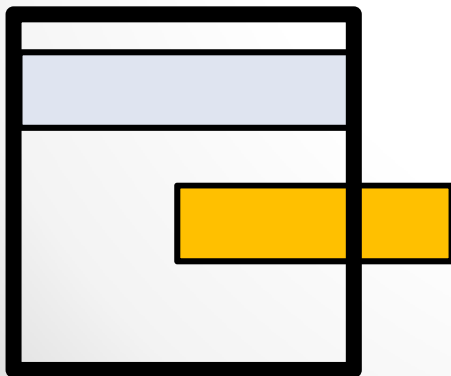
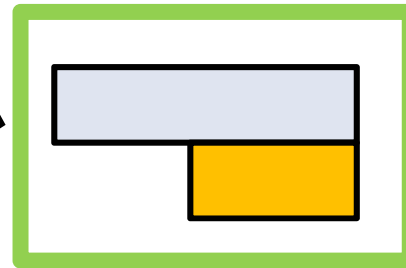


# 区間系列への適用

- 1、パターンの区間イベントの先頭とウィンドウの先頭が一致しているか
- 2、先頭の後ろのイベントの順序と重なり方が一致している



例 パターン



# 評価実験

以下の観点に着目し実験を行う  
パラメータの変化によるパターンの違いはどうか

実験データ … 株価データを使用 <http://k-db.com/>

2007年から2015年の取引のあった約2,200日分の100社分のデータ

最小支持度  $\sigma$  とウィンドウ  $\omega$  を閾値とし、  
30分で求められるパターン数を見る

# 実験結果

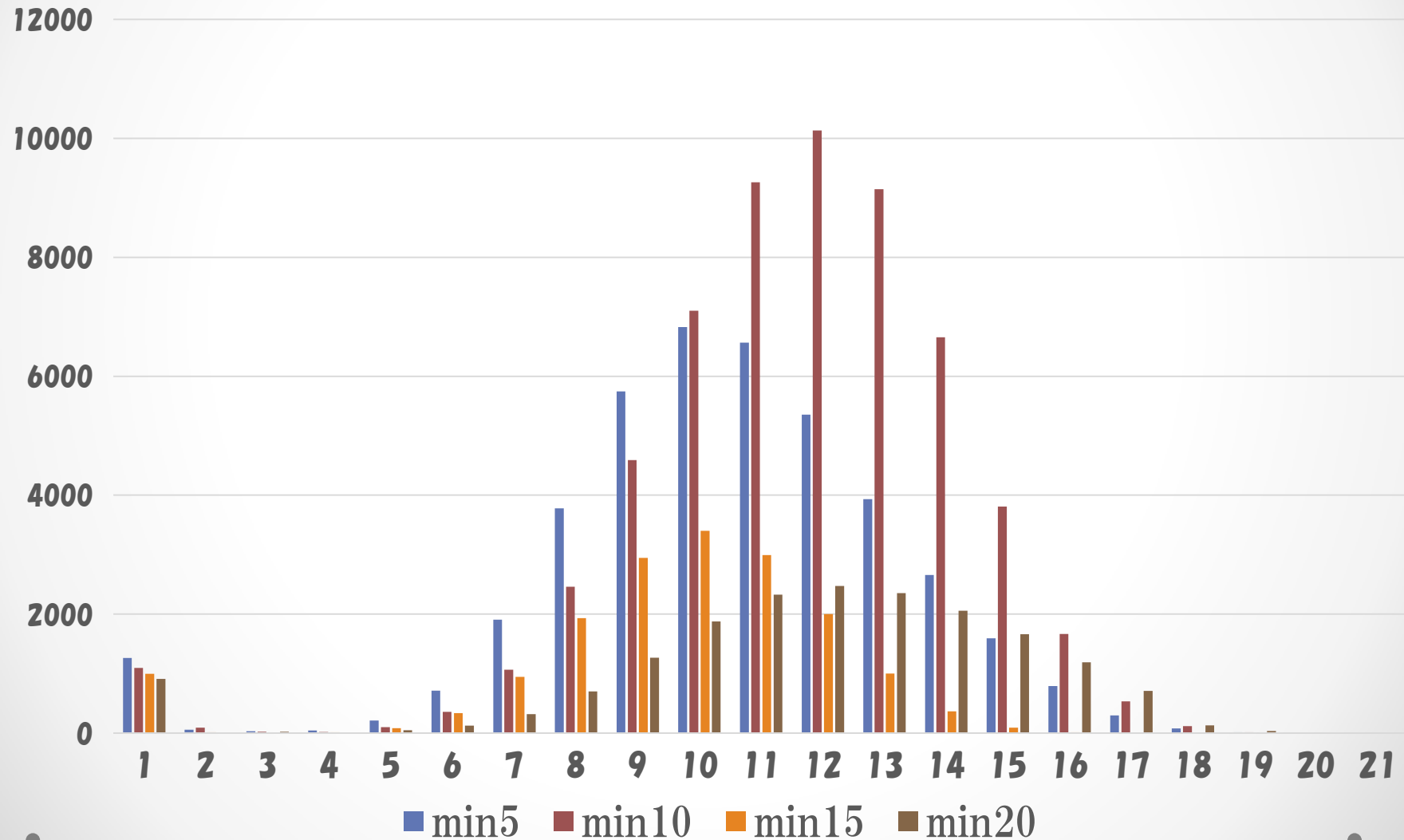
$\sigma$  最小支持度 5,10,15,20  
 $\omega$  ウィンドウ 5,10,15,20

パターン数

$\omega \setminus \sigma$	5	10	15	20
5	50573	42447	5139	3610
10	41859	58244	17140	18580
15	47848	57677	18824	29324
20	26198	8536	21563	21066



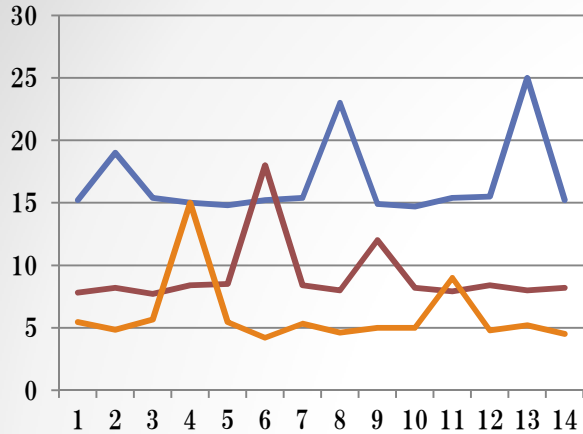
# ウィンドウ10の各最小出現回数での パターンのサイズ分布



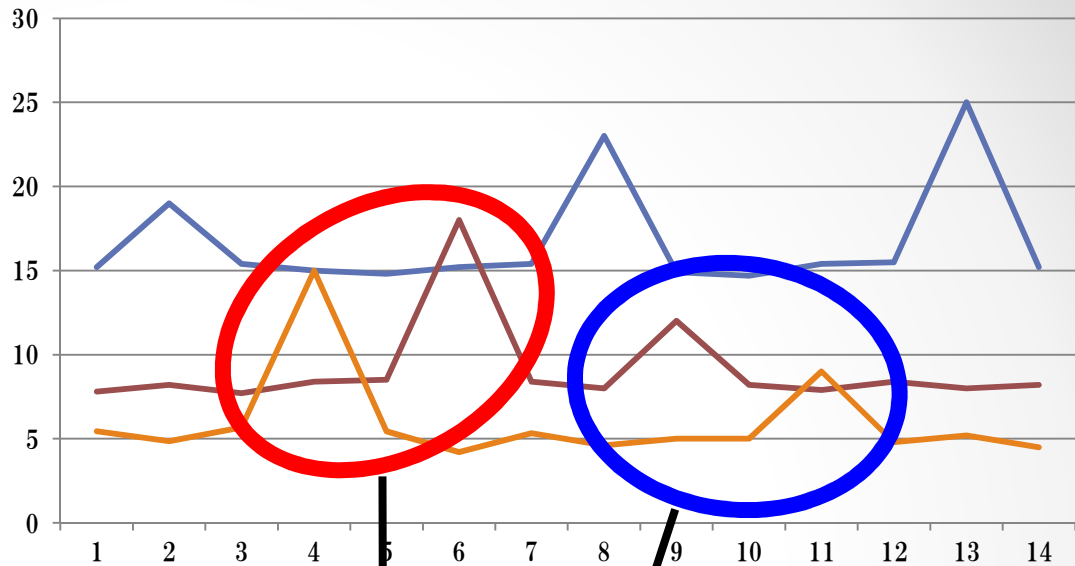
# まとめと今後の課題

- 多次元時系列データに対し、類似区間に着目した区間イベントへの変換をした。得られた区間イベント系列から、時間幅を考慮した区間系列パターン抽出手法を実装した。
- 今後の課題  
今回は類似区間に着目してイベント区間に変換  
⇒ **非類似区間**に着目したパターンを抽出する

# 多次元時系列データ



# 類似区間抽出



区間イベント系列を  
連続した辺の系列と捉える

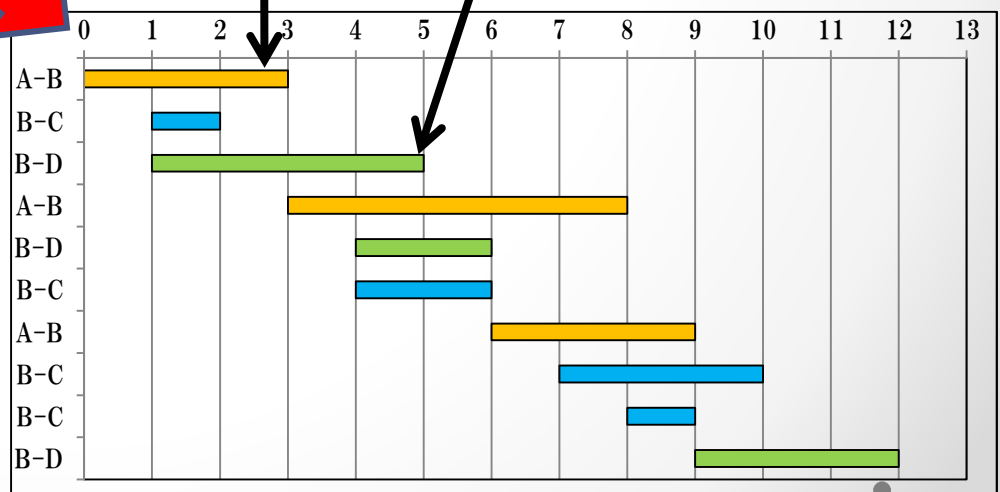


系列に繰り返し出てくる  
連結した部分グラフ



グラフ構造と共に、  
順序関係を考慮した  
区間イベント系列パターン  
連結性と辺の順序関係を  
考慮した区間系列パターン

# 区間イベント系列

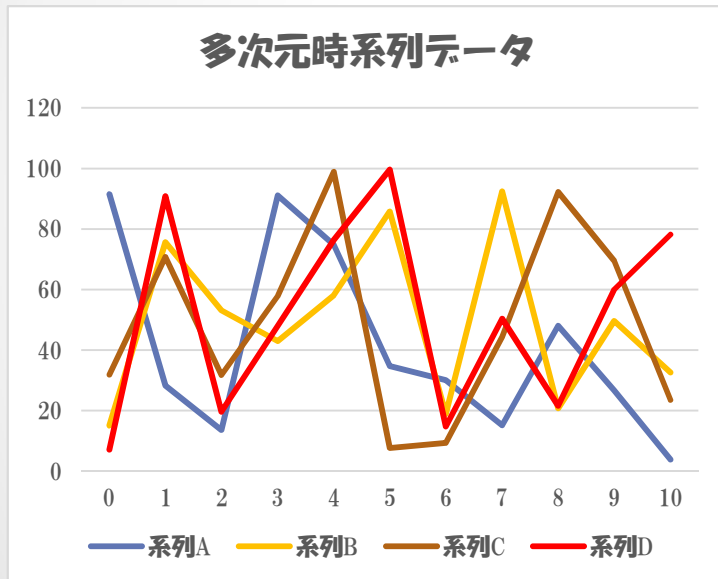


# 区系列からグラフへ

2つの系列間の関係に着目  
区間イベント系列



1系列を1頂点  
辺系列



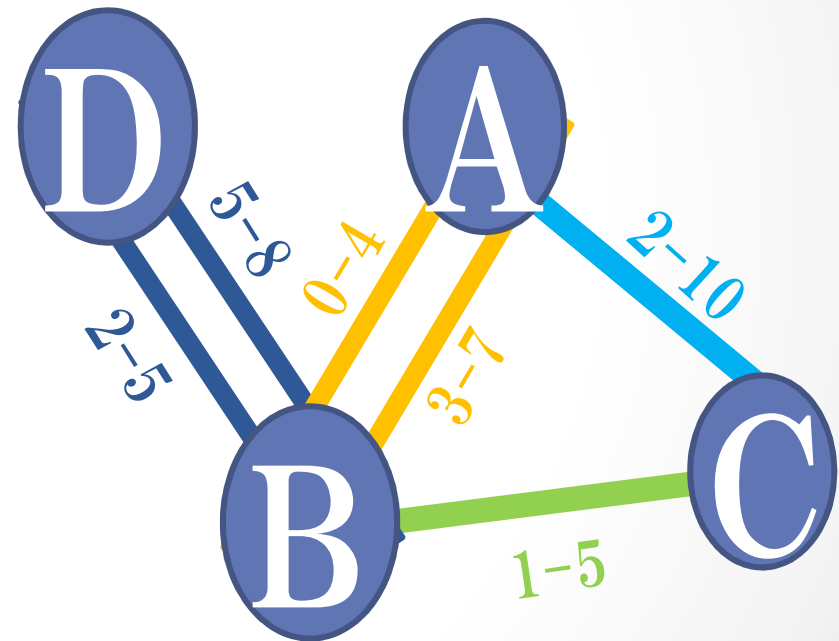
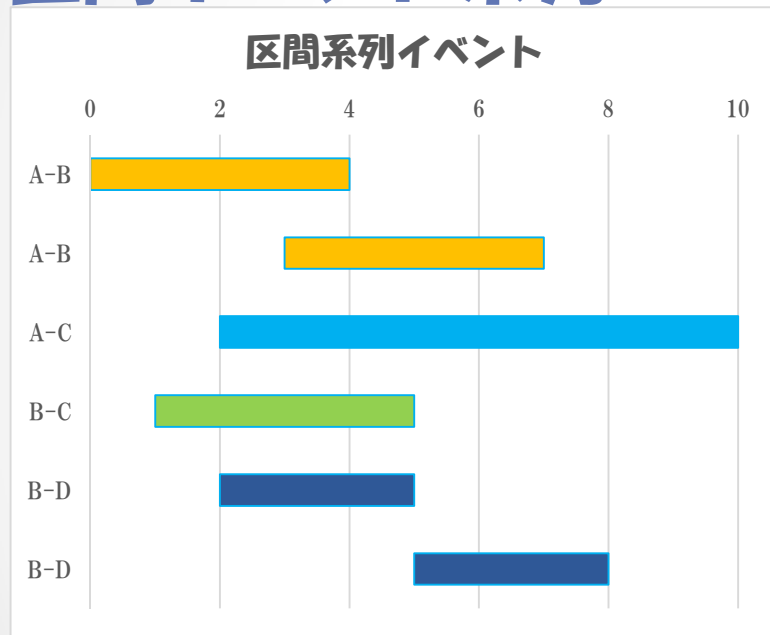
1つの系列を1つの頂点

# 区系列からグラフへ

2つの系列間の関係に着目  
区間イベント系列

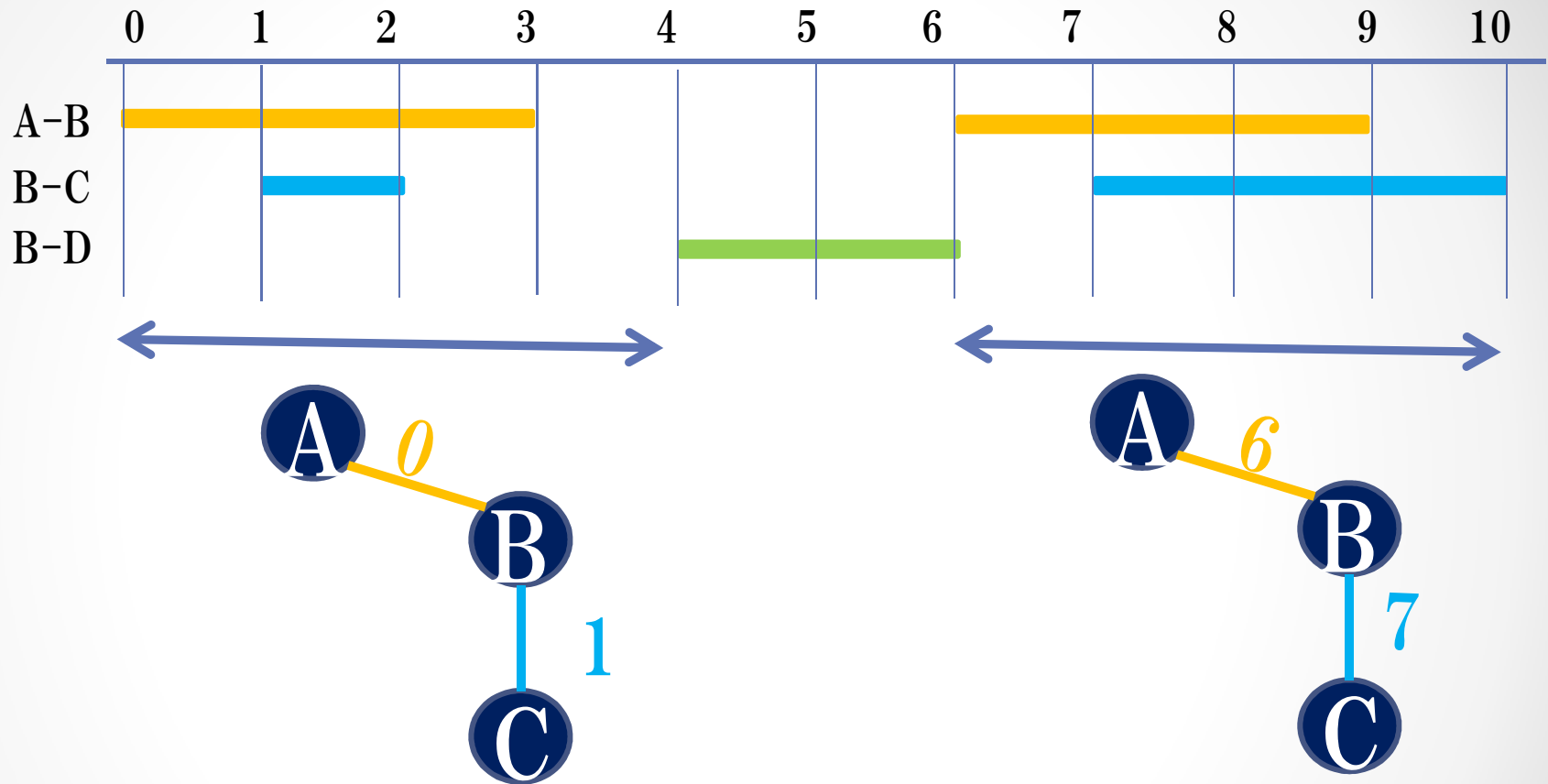


1系列を1頂点  
辺系列

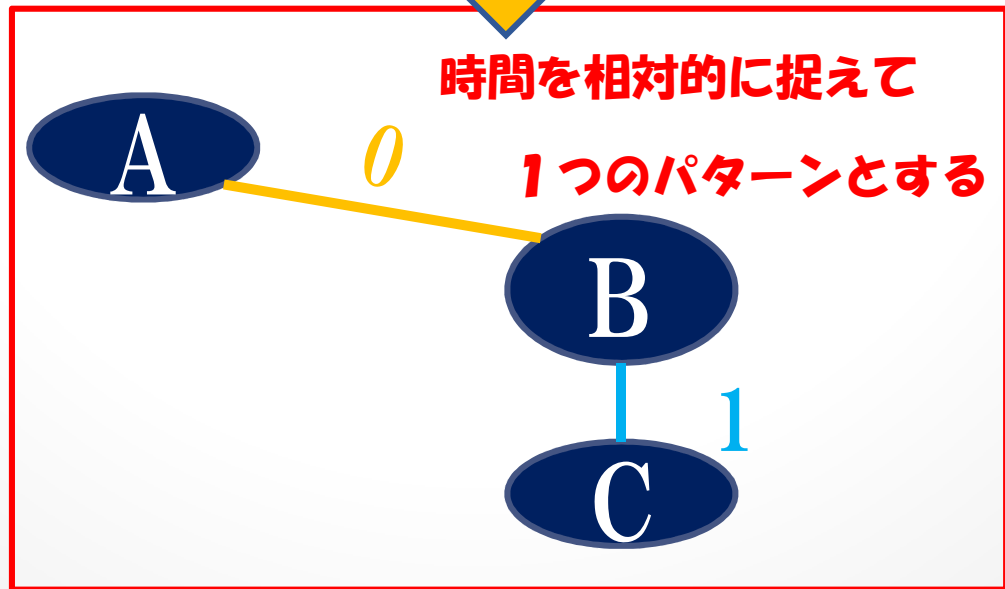
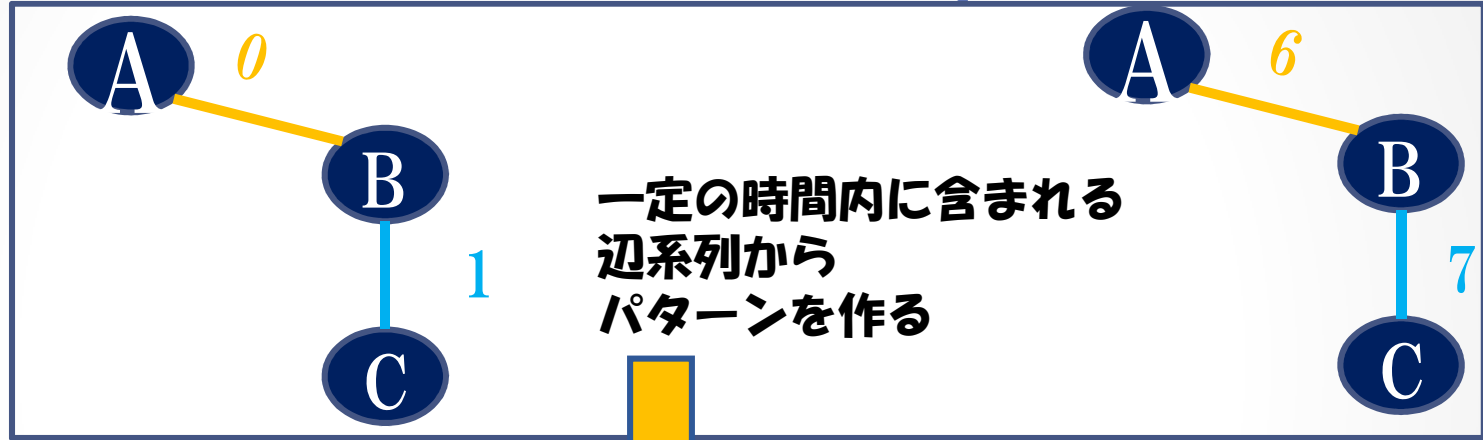
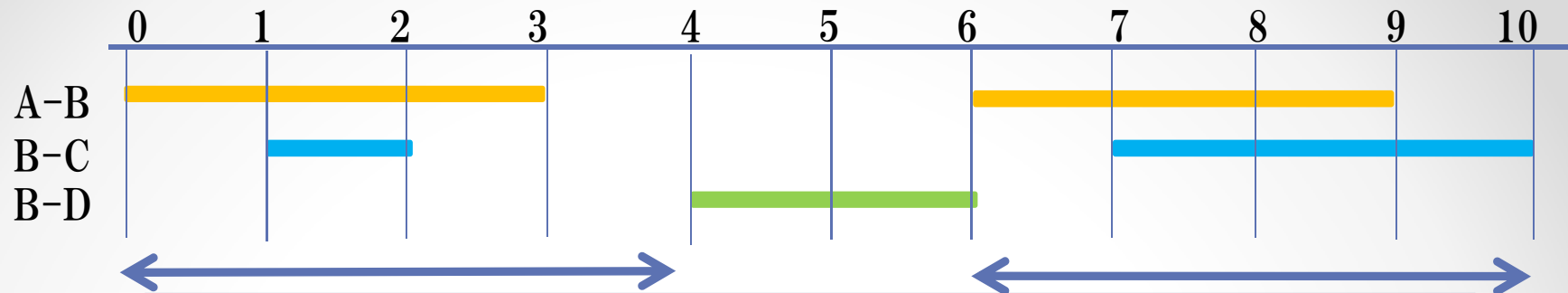


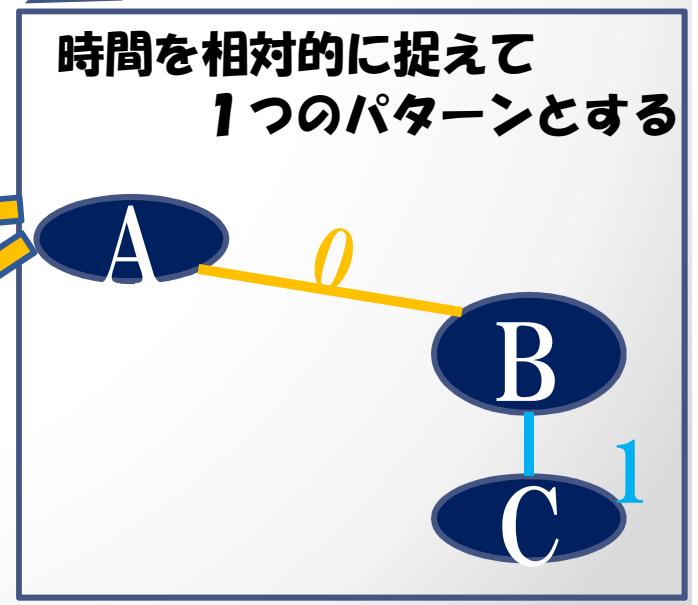
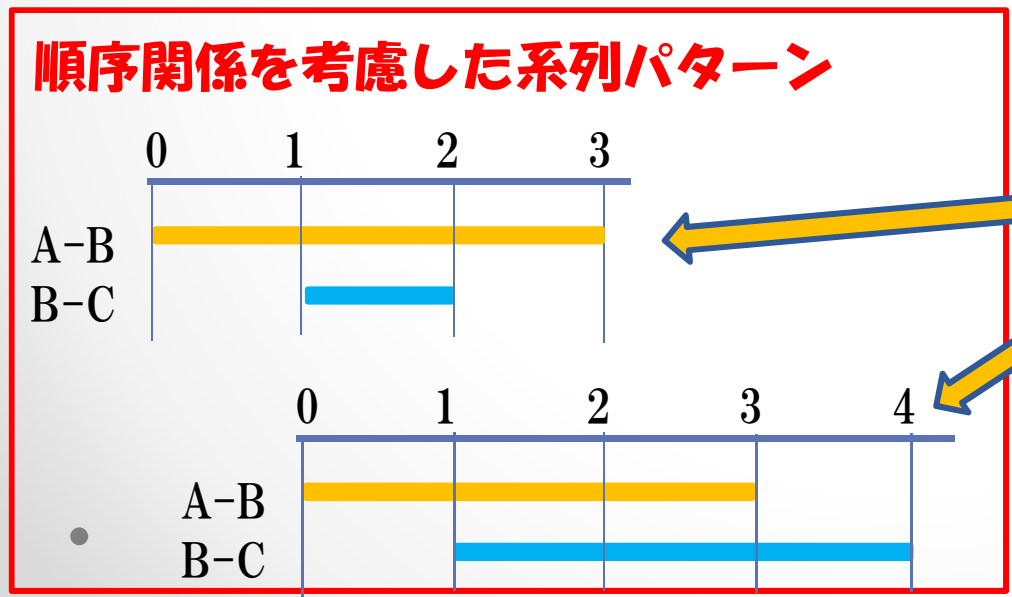
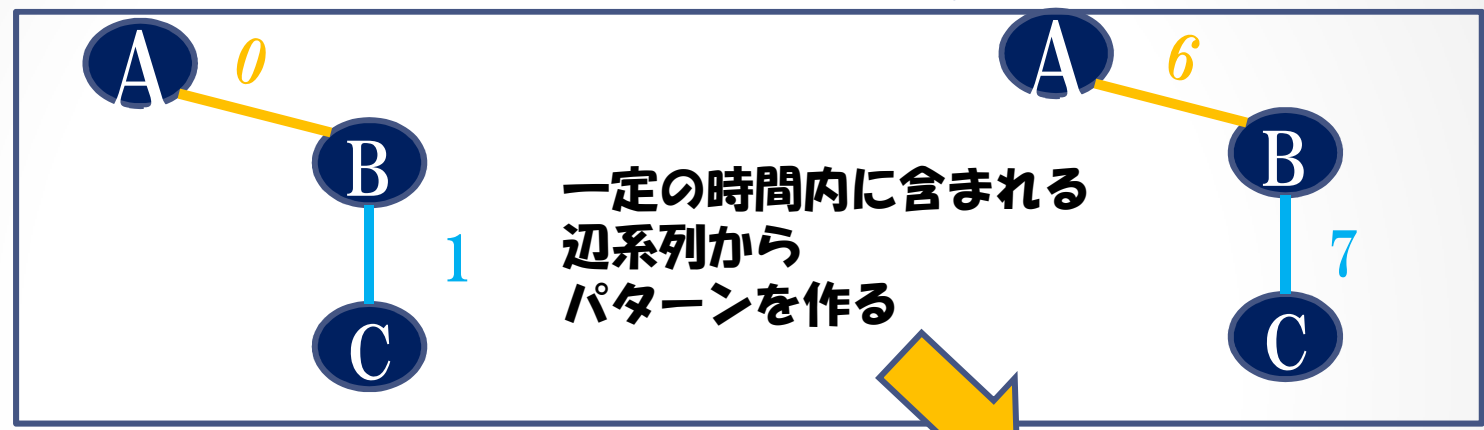
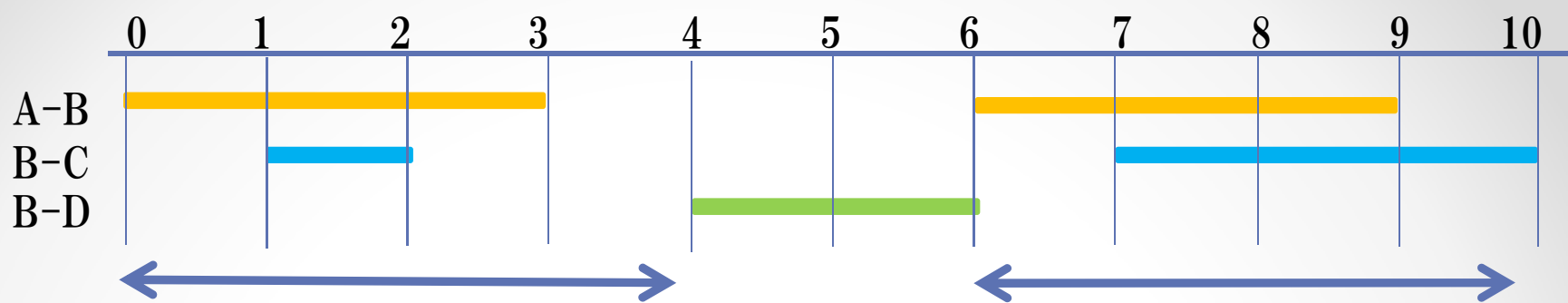
1つの類似区間を1つの辺  
開始と終了時刻を辺ラベルとした  
全連結で無向多重グラフ

# 全体の流れ



※ 辺ラベルは **一定の時間内に含まれる**  
開始時刻のみ **系列から繰り返し出る連結性を**  
**満たした多重辺のない部分グラフ**

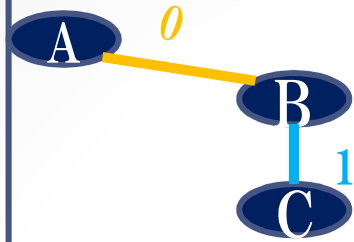




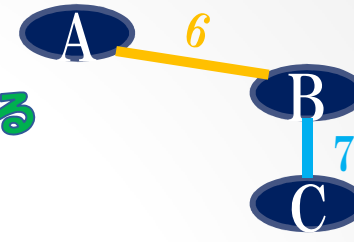


# アルゴリズムについて

単一グラフでの  
頻度計算



Window幅の中にある  
辺系列から  
パターンを作る



What Is Frequent in a Single Graph? [2]

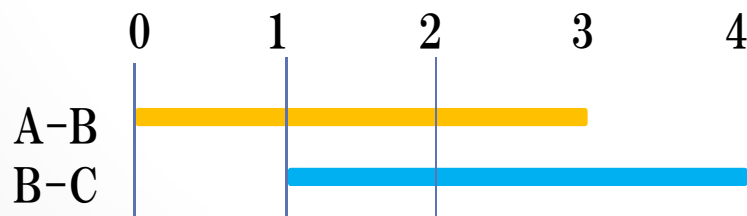
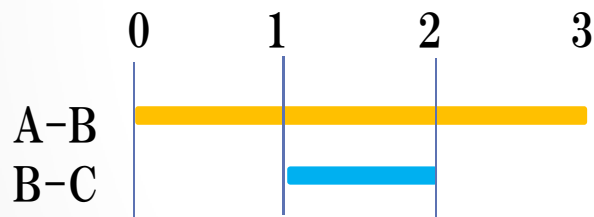
gSpan [3]

GERM [4]

TPrefixSpan [5]

時間を相対的に捉えて  
1つのパターンとする

順序関係を考慮した系列パターン



4つの技術を独自に組み合わせて作成

# 評価実験

以下の観点に着目し実験を行う

- 有効な時間内で列挙できるか
- パラメータの変化によるパターンの違いはどうか

実験データ … 株価データを使用 <http://k-db.com/>  
2007年から2015年の取引のあった約2,200日分のデータ

企業の数 50社

類似区間対 ( 区間イベント系列の数 ) 約 7,500

1つの企業 → 1つの頂点  
1つの類似区間 → 1つの辺 } 50の頂点と  
約7,500の辺からなるグラフ

# 実験結果1

$\sigma$  最小支持度 10,15,20,25

$\omega$  window幅 7,14,21,28

\* 1時間以内に結果が出なかったもの

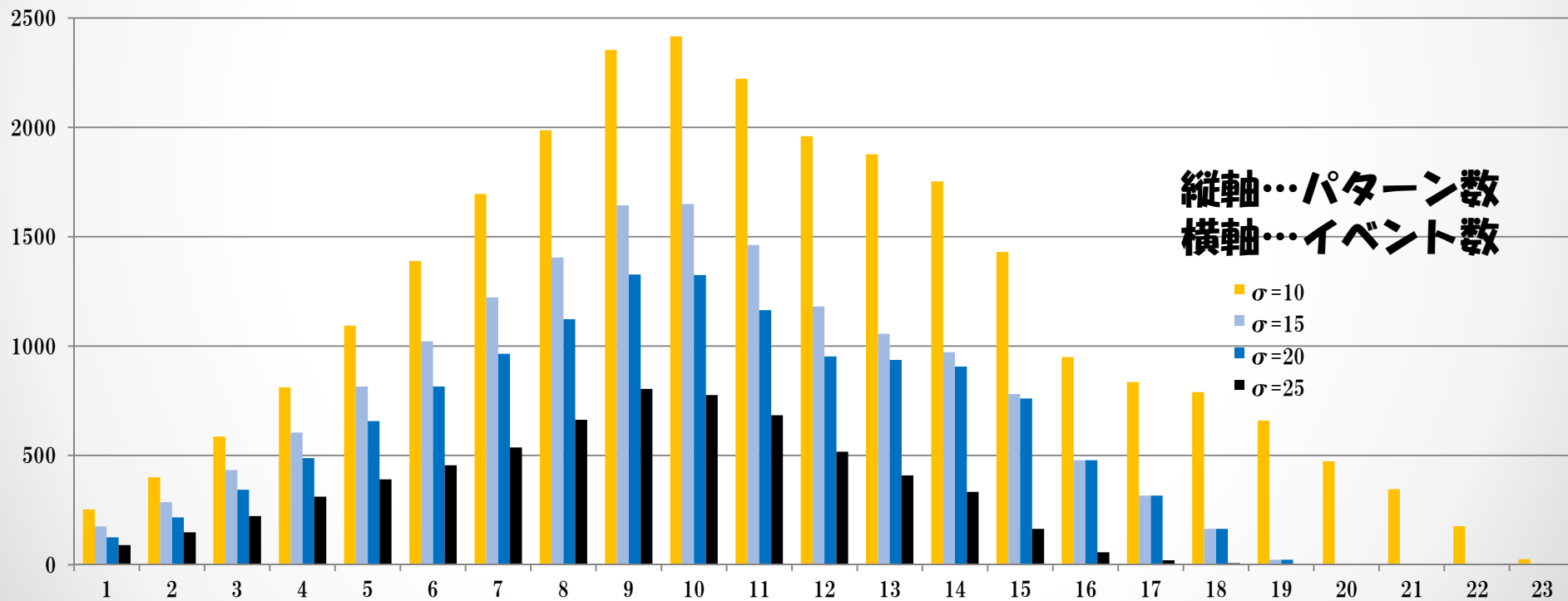
$\omega \setminus \sigma$	10	15	20	25
7	206	88	29	8
14	26,496	15,699	13,096	6,596
21	*	*	15,019	13,781
28	*	*	*	23,352

$\omega \setminus \sigma$	10	15	20	25
7	window幅が増えるほどパターン数も増える			
14	1,318	1,081	1,061	603
21	*	*	1,200	1,740
28	*	*	*	2,760

# 実験結果2

	パターン数 (個)			
$\omega \setminus \sigma$	10	15	20	25
14	26,496	15,699	13,096	6,596

Window幅14 の時の区間イベントパターンのイベント数の分布



各最小支持度もイベント数が10の時のパターン数が最大なことがわかる

# まとめと今後の課題

- **類似区間対に着目し多次元時系列データを区間系列化し辺系列と捉えることでグラフ構造と共に時間を考慮したパターンの抽出を提案した**

## 課題

- **類似区間ではなく非類似区間に適用**
- **時間による順序関係だけでなく時間差も考慮した区間系列パターンの列挙**

# 参考文献

- [TPrefixSpan] Shin-Yi Wu and Yen-Liang Chen,  
Mining Nonambiguous Temporal Patterns for Interval-Based  
Events”  
IEEE Transactions on Knowledge and Data Engineering,  
Vol.19,No.6 (2007).**
- [gSpan] Xifeng Yan and Jiawei Han,  
“gspan : Graph-based substructure pattern mining”  
Proceedings of the 2002 IEEE International Conference on  
Data Mining IEEE Computer Society , pp.721--724 (2002).**
- [GERM] Michele Berlingerio, Francesco Bonchi, Bjorn Bringmann, and  
Aristides Gionis. “Mining graph evolution rules” ,  
PKDD, pp.115--130 (2009).**
- [4] Bjorn Bringmann and Siegfried Nijssen,  
What Is Frequent in a SingleGraph?,  
PacificAsia Conference on Knowledge Discovery and Data  
Mining (PAKDD) (2008).**