

# カレンダーの予定を クラスタリングする手法の検討

石川大夢 乃村能成 谷口秀夫  
岡山大学大学院 自然科学研究科

2017年6月1日

DPS171/MBL83/ITS69

# カレンダー情報の活用

## <カレンダー情報>

(1) 日付, タイトルといった予定に関する情報

(2) 主に, 将来のスケジュールの確認という目的で利用

過去の情報を用いて予定を整理 ➡ 有効活用可能

類似している予定を集合にまとめる

(例) 「第1回会議」, 「第2回会議」, . . . を「会議」として整理

➡ 次回会議の計画立案に有用

∴ 会議の発生間隔, 去年の発生時期を把握可能

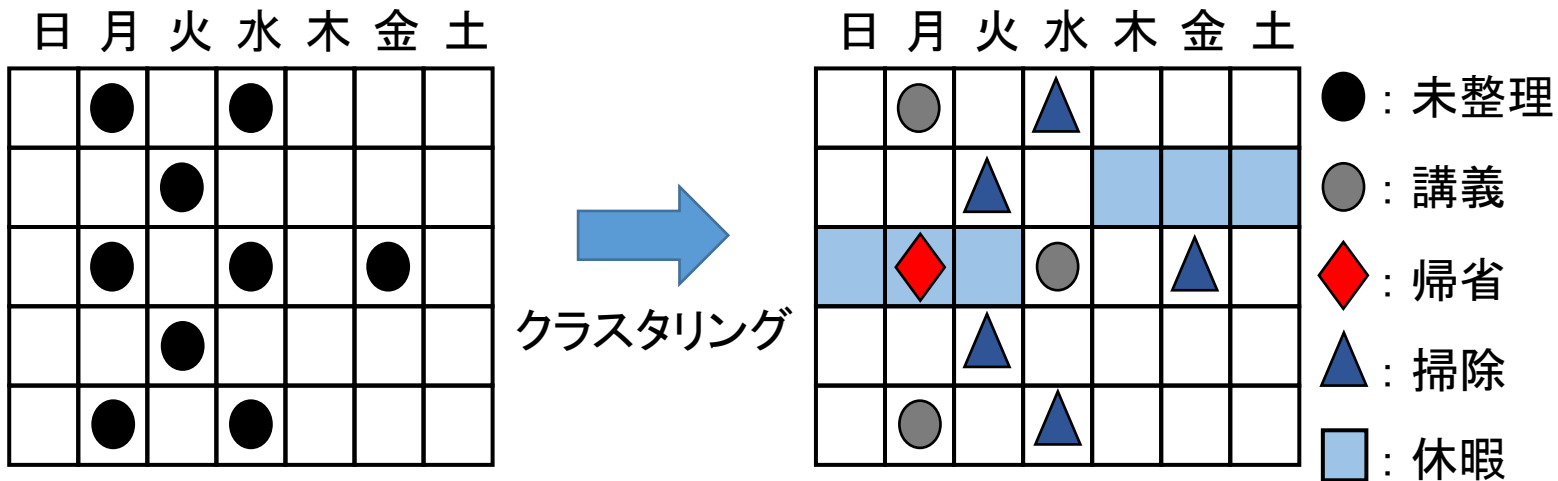
∴ 整理の支援方法として, 予定をクラスタリングする手法を提案

# クラスタリングによる整理支援

## <クラスタリング>

(1) 教師なし学習

(2) 着目する特徴に基づいたグループ(クラスタ)を作成



望ましいクラスタを得るためには、特徴の選択が重要

➡ 予定のクラスタリングの場合、**実施日が重要**

∴ カレンダー情報は時系列で管理

∴ クラスタリングに用いる特徴として実施日がもつ**周期性に着目**

# カレンダーの予定間に存在する周期

ほぼ全ての予定は何らかの曖昧な周期をもって発生

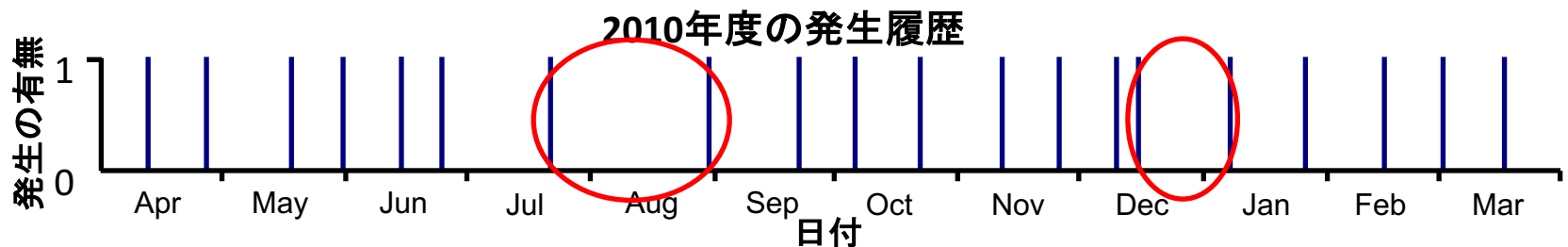
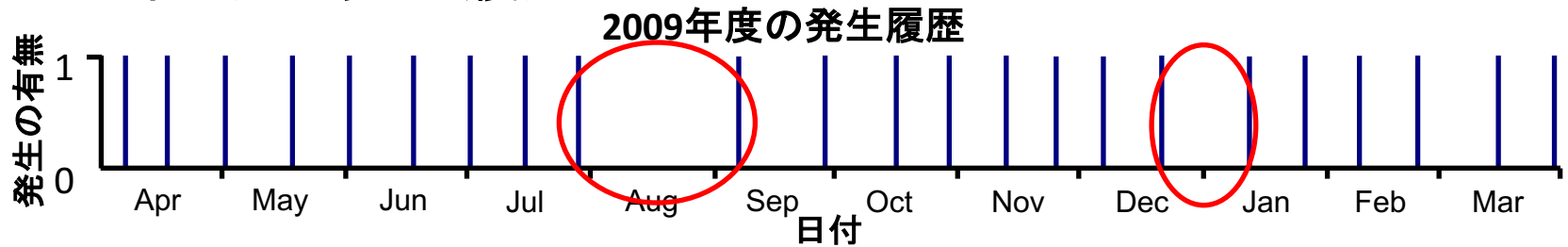
∴ 予定間の周期性はクラスタリングに有用

## <曖昧な周期>

(1) 固定的な周期(毎週月曜日)とは限らない

(2) 同様の作業が繰り返し発生し、**年を単位とした系列が相似形**

## <ミーティングの発生履歴>



➡ クラスタリングに用いる特徴として表現するには?

# 曖昧な周期の表現

周期的に発生する予定の集合をクラスタとして作成

➡ クラスタ内の予定が安定した周期をもつようにクラスタリング

∴ 安定度合いを定量化する指標があれば、クラスタリングに利用可

<安定度合いの定量化に用いる概念>

(1) **タスク**: カレンダーに登録されている個々の予定

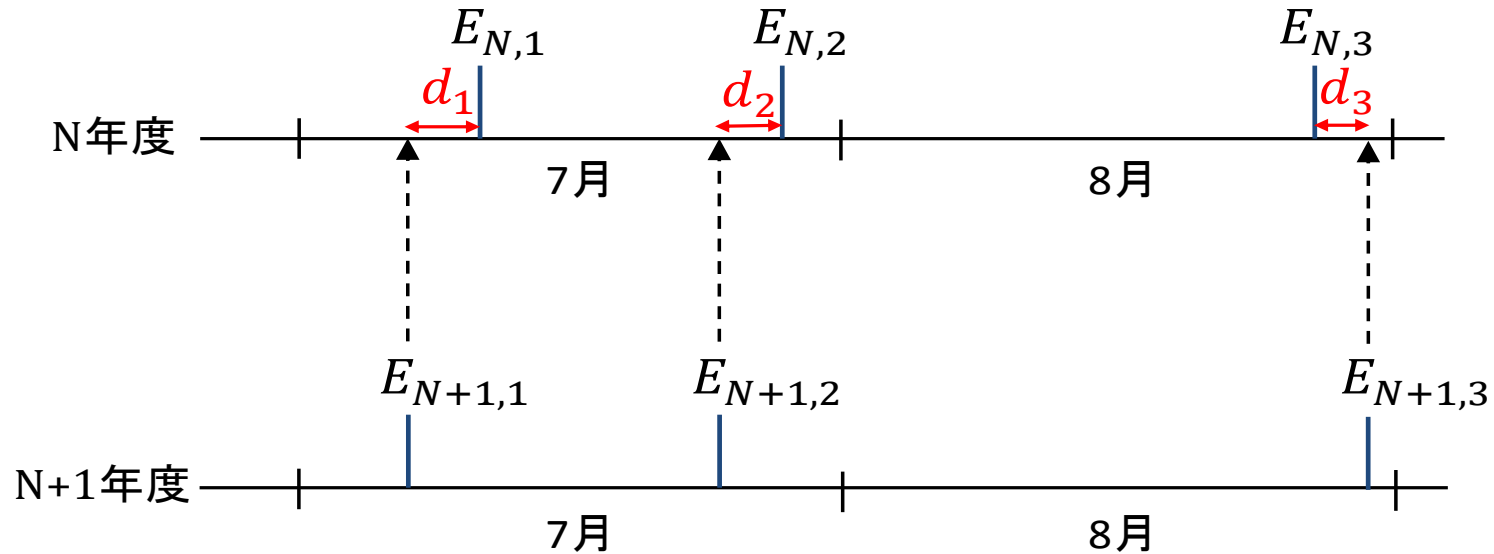
(2) **リカーレンス**: 繰り返し発生する同様のタスクの集合  
=クラスタリングによって作成するタスクの集合

リカーレンス内に含まれるタスクがもつ周期の安定度合いを評価

➡ リカーレンスの信頼性を評価

# リカーレンスの信頼性

作業発生系列の自己相関を用いて評価



対応する組の差分の総和小

➡ 作業発生系列の自己相関強

= 作業発生が年を単位とした相似形

∴ 自己相関が高くなるようなクラスタを作成

➡ リカーレンスに相当するクラスタを作成可能

# タスクのクラスタリング手法

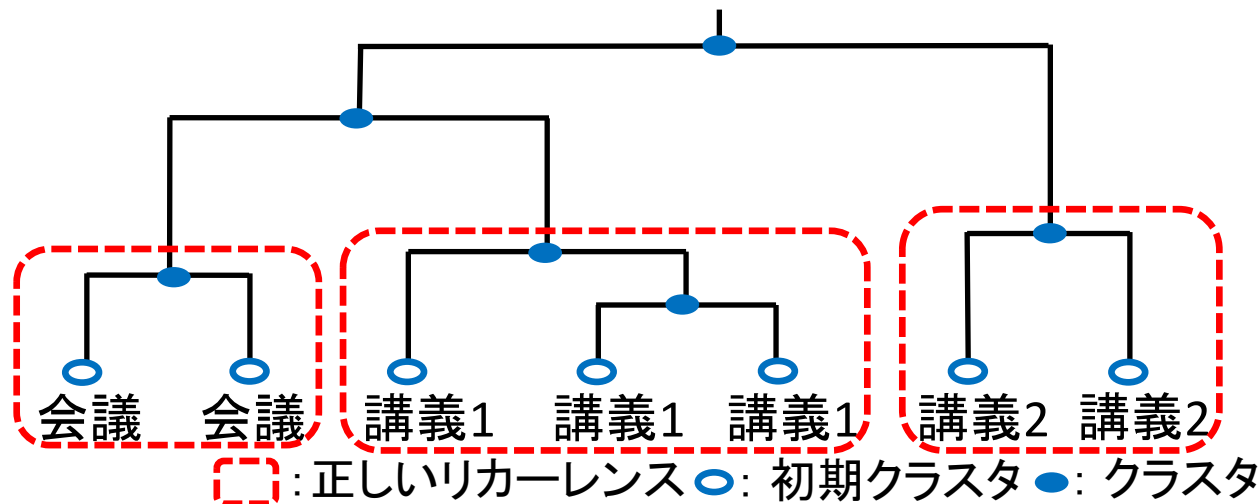
タスクの整理時において、分類するグループ数は未定

∴階層型クラスタリング手法を用いた手法を提案

∴クラスタ数をあらかじめ与える必要無

## ＜階層型クラスタリング手法＞

- (1) 1つのタスクのみを要素としてもつ初期クラスタを複数作成
- (2) クラスタ間距離が最小となるクラスタの組を順に結合
- (3) 結合するクラスタがなくなった時点で終了
- (4) 階層構造から望ましいクラスタを選択



# 信頼性によるクラスタ間距離

なるべくリカーレンスの信頼性が高いクラスタを作成する必要有

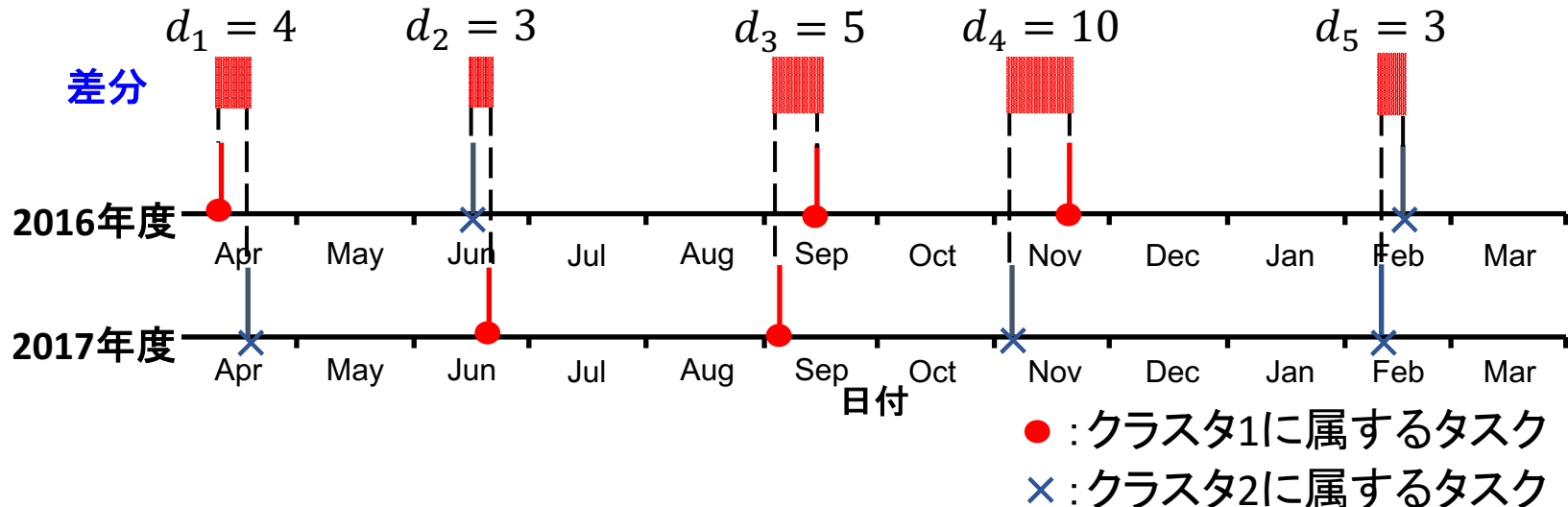
➡ クラスタを結合した際の自己相関をクラスタ間距離として使用

実施日の集合

クラスタ1 2016/4/13, 2016/9/12, 2016/11/15, 2017/6/18, 2017/9/7

クラスタ2 2016/6/15, 2017/2/16, 2017/4/17, 2017/11/5, 2018/2/13

結合後の自己相関を算出





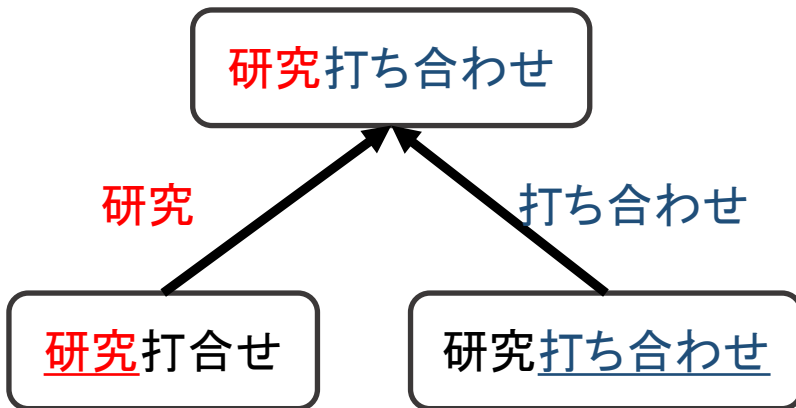
# クラスタ名の生成

- クラスタリングの結果得られる階層中には多くのクラスタが存在  
∴ 一目で望ましいクラスタか否かを判断可能なクラスタ名を生成  
➡ クラスタの選択作業を支援可能

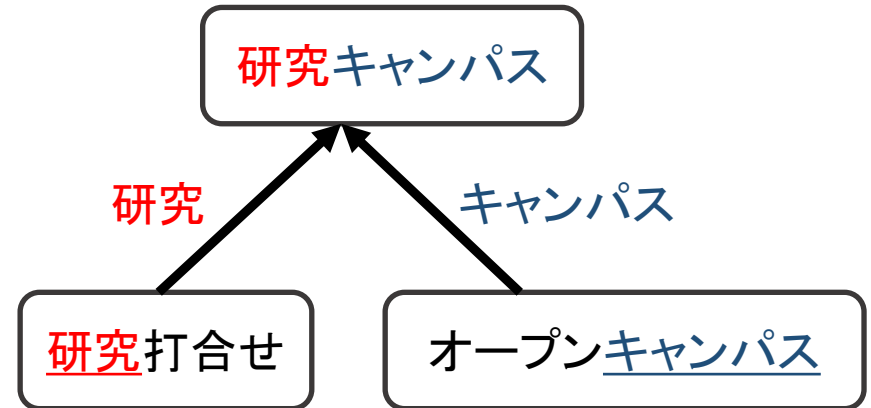
## <クラスタ名の生成手法>

- (1) 結合元の2つのクラスタ名を形態素に分割
- (2) 一方の上位半分と、もう一方の下位半分の形態素を結合

### <正しく分類されている場合>



### <複数のリカーレンスを含む場合>



# 評価

予定の整理作業にかかる手間を評価

## <想定するユーザインタフェース>

- (1) プルダウン形式
- (2) チェックボックスによる選択

## <理想的な階層構造>



チェックボックスの選択数が最小

## <評価尺度>

選択するチェックボックスの数

= 階層構造中の 部分集合クラスタ 数

整理の結果得られるリカーレンスの  
部分集合となるクラスタ

Tasks  

- ▼  研究打ち合わせ
  - ▼  研究打ち合わせ
    - ▼  研究打合せ
      - 研究の打ち合わせ
      - 研究打合せ
    - ▼  研究打ち合わせ
      - 研究の打合せ
      - 研究打ち合わせ
  - ▼  幹事打ち合わせ
    - 幹事の打ち合わせ
    - 幹事打ち合わせ
- ▼  研究室忘年会
  - ▼  研究室ミーティング
    - 研究室ミーティング(2)
    - ミーティング
  - ▼  研究室忘年会
    - 研究室忘年会
    - 研究室の忘年会

# 評価環境の準備と評価方法

## <評価環境の準備>

- (1) 実データと似た性質を持つタスク(ダミーデータ)を使用
  - (A) 研究室の共有カレンダーから抽出した実データから作成
  - (B) 30種類のリカーレンスを表現
- (2) 同名タスクをあらかじめ初期クラスタとして結合
  - (A) 同名タスクの割合を変えて評価

## <評価方法>

- (1) 同名タスクの割合が 0~100%の場合においてそれぞれ評価
- (2) 提案手法とタスク名のみを用いた手法の比較
- (3) 各リカーレンスの部分集合クラスタ数の平均値を評価

# 評価結果

手法	同名タスクの割合(%)										
	0	10	20	30	40	50	60	70	80	90	100
(A)タスク名のみ (個)	20.0	19.0	17.0	15.0	13.0	11.0	9.0	7.0	5.0	3.0	1.0
(B)提案手法 (個)	16.3	15.7	14.0	11.8	10.4	9.0	7.4	6.4	4.7	2.9	1.0
(A)-(B)	3.7	3.3	3.0	3.2	2.6	2.0	1.6	0.6	0.3	0.1	0

(a) タスク名のみに比べ提案手法で部分集合クラスタ数少

 選択するチェックボックス数少

# 評価結果

手法	同名タスクの割合(%)										
	0	10	20	30	40	50	60	70	80	90	100
(A)タスク名のみ (個)	20.0	19.0	17.0	15.0	13.0	11.0	9.0	7.0	5.0	3.0	1.0
(B)提案手法 (個)	16.3	15.7	14.0	11.8	10.4	9.0	7.4	6.4	4.7	2.9	1.0
(A)-(B)	3.7	3.3	3.0	3.2	2.6	2.0	1.6	0.6	0.3	0.1	0

大 ←

(a) タスク名のみに比べ提案手法で部分集合クラスタ数少

➡ 選択するチェックボックス数少

(b) 同名タスクの割合が小さい場合に評価値の差大

# 評価結果

手法	同名タスクの割合(%)										
	0	10	20	30	40	50	60	70	80	90	100
(A)タスク名のみ (個)	20.0	19.0	17.0	15.0	13.0	11.0	9.0	7.0	5.0	3.0	1.0
(B)提案手法 (個)	16.3	15.7	14.0	11.8	10.4	9.0	7.4	6.4	4.7	2.9	1.0
(A)-(B)	3.7	3.3	3.0	3.2	2.6	2.0	1.6	0.6	0.3	0.1	0

(a) タスク名のみに比べ提案手法で部分集合クラスタ数少

➡ 選択するチェックボックス数少

(b) 同名タスクの割合が小さい場合に評価値の差大

∴ 予定間の周期性が予定のクラスタリングに有用

# まとめ

- (1) カレンダの予定をクラスタリングする手法の検討
  - (A) 予定間の周期性を用いたクラスタリング
  - (B) 階層型クラスタリング手法を応用
- (2) 提案手法の評価
  - (A) 整理にかかる手間を評価
  - (B) 整理ユーザインタフェースを想定
  - (C) 整理に要するチェックボックスの選択数を評価

予定間の周期性はカレンダの予定のクラスタリングに有用

<残された課題>

- (1) クラスタリングに用いる特徴の検討

# 予備スライド



# 同名タスク

リカーレンス内のタスク名には以下の3種類の分類が存在

(分類1) リカーレンス内の他のタスクと全く同じタスク名

(分類2) 年度, 回数を示す部分のみ異なるタスク名

(分類3) その他(送り仮名, 用いる単語が相違)

(分類1)と(分類2)を同名タスクとして定義

# 同名タスクの割合の調査

## <調査目的>

リカーレンス中の同名タスクの割合を調査

## <調査対象>

研究室の共有カレンダーから抽出した280件のタスク

➡ 30種類のリカーレンスを表現

## <調査結果>

84.3%のタスクがリカーレンス中において同名タスク

# タスク名を用いた初期クラスタの作成

初期クラスタのようなタスクの数が少ないクラスタ

➡ 誤ったクラスタの組が結合される可能性高

∴ **タスク数が大きい初期クラスタを作成**

＜タスク名を用いた初期クラスタの作成＞

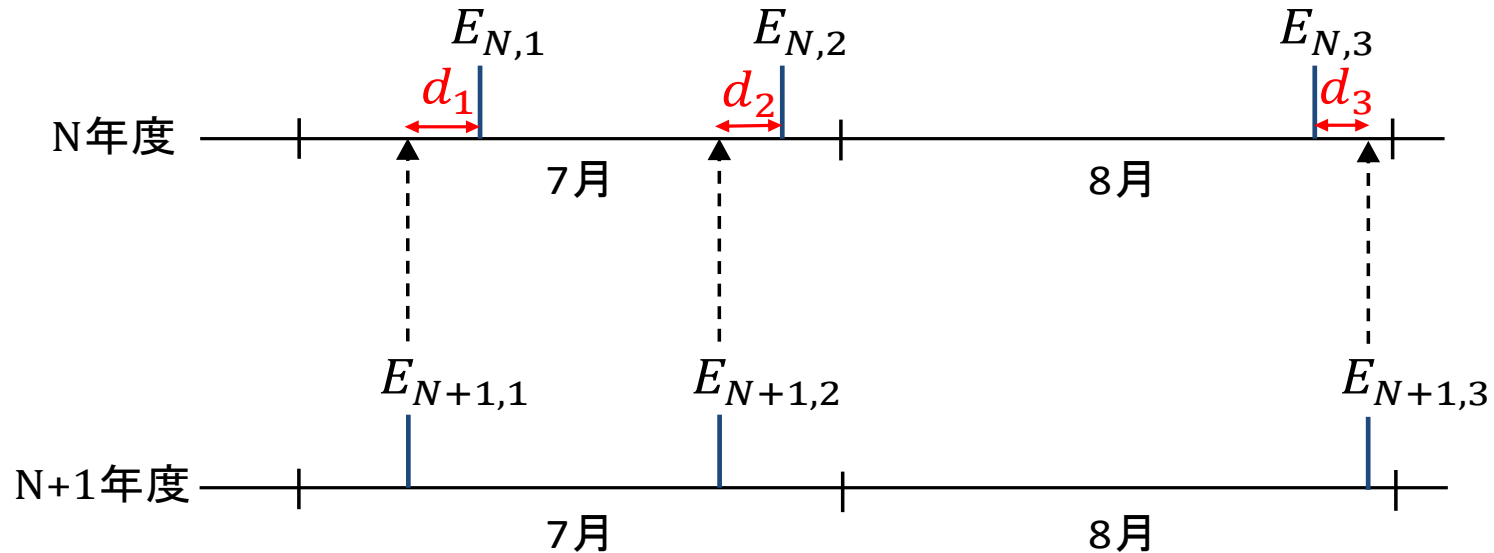
繰り返される同様の予定に対して規則性のあるタスク名が命名

➡ リカーレンス内のタスク名がある程度統一

∴ **同名のタスクをまとめたものを初期クラスタとして作成**

# リカーレンスの信頼性

リカーレンスの信頼性を作業発生系列の自己相関を用いて評価



$$R(\tau) = \sum_{t \in D} (f(t) - f(t - \tau))^2$$

$\tau$  = ラグ

$D$  = 作業発生日の集合

$f(t)$  = 日付  $t$  に最も近い  $D$  の要素を返却

$R(365)$  が小さいほど、自己相関強  
= 作業発生が年を単位とした相似形

∴ 自己相関が高くなるようなクラスタを作成

# 曖昧な周期の表現

作業発生が年を単位とした相似形

➡ 予定間に存在する周期が安定するかを表現する特徴

∴ 安定度合いを定量化する指標があれば、クラスタリングに利用可

< 安定度合いの定量化に用いる概念 >

(1) **タスク**: カレンダーに登録されている個々の予定

(2) **リカーレンス**: 繰り返し発生する同様のタスクの集合  
= クラスタリングによって作成するタスクの集合

リカーレンス内に含まれるタスクがもつ周期の安定度合いを評価

➡ リカーレンスの信頼性を評価