

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¹ 乃村 能成¹ 谷口 秀夫¹

概要: 日々発生する予定は、カレンダーを用いて管理することが多い。これらの予定について、その特徴に基づいて、類似している予定を集合として整理しておくことは計画立案において有用である。そこで、本稿では、予定の整理を支援する方法として、類似している予定をグループにまとめる手法を述べる。これは、着目した特徴に基づいてデータをグルーピングする手法であるクラスタリングをおこなうことに相当する。具体的には、予定間に存在する周期性に着目したクラスタリング手法を提案する。そして、ユーザが予定を整理する際にかかる手間を観点として提案手法によって得られるクラスタリング結果を評価し、予定間の周期性が予定のクラスタリングに有用であることを述べる。

キーワード: カレンダー, クラスタリング, 計画立案

1. はじめに

日々発生する予定は、カレンダーを用いて管理することが多い。カレンダーの予定は、日付、タイトルといった予定に関する情報（以下、カレンダー情報と呼ぶ）をもっている。カレンダー情報は、将来のスケジュールの確認という目的で利用することが多い。文献 [1] では、国内企業の 8 割がスケジュール管理を支援するアプリケーションを利用していることが確認されており、将来発生する予定に関するカレンダー情報を活用する場面は多い。さらに、カレンダー情報を有効活用する事例として、過去のカレンダー情報を用いて、類似している予定を整理しておけば、将来の計画立案に有効活用できることがある。類似している予定の整理の方法として、例えば、「第 1 回会議」、「第 2 回会議」、…、「第 n 回会議」のように定期的に発生する定例会議を表す予定を「会議」という集合にまとめる手法がある。定期的に発生する「会議」のまとまりを時系列で一覧することができれば、どの程度の間隔で会議が行われているのか、去年の今頃はいつ会議が行われたのかといった情報を容易に参照でき、次回会議の日程を決める際に有効活用できる。したがって、カレンダー情報が類似している予定をまとめておくことは有用である。

カレンダー情報を用いた予定の整理において、カレンダー情報を用いて予定を整理するといった考え自体はあるものの、カレンダーやスケジュール管理を支援するアプリケーション

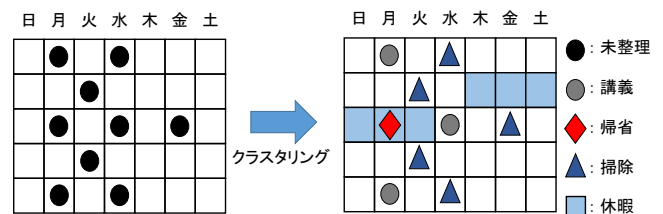


図 1 カレンダーの予定をクラスタリングする例

は整理に関するユーザ支援が十分ではない。このため、予定を適切に整理することは難しい。

そこで、本稿では、カレンダー情報を用いた予定の整理を支援する方法として、類似している予定をグループにまとめる手法を述べる。これは、着目する特徴に基づいてデータをグルーピングする手法であるクラスタリングをおこなうことに相当する。

クラスタリングとは、教師なし学習の 1 つであり、対象となるデータを着目する特徴に基づいたグループであるクラスタにまとめる手法である。図 1 は、未整理状態の予定をクラスタリングする様子である。図 1 の例では、クラスタリングの結果、「講義」、「帰省」、「掃除」といった 3 つのクラスタに分類されている。クラスタリングの分野において、望ましいクラスタを得るためには、特徴の選択が重要である。また、選択すべき特徴は分類対象によって異なる。例えば、テキスト分類の分野では、単語間の類似性を単語の属性空間から算出する [2][3]。ブログのエントリをストーリーごとにクラスタリングする研究では、ブログ間に貼られたリンクを用いた分類手法 [4] が提案されている。また、人の興味を特徴として抽出し情報をクラスタリング

¹ 岡山大学大学院自然科学研究科
Graduate School of Natural Science and Technology,
Okayama University

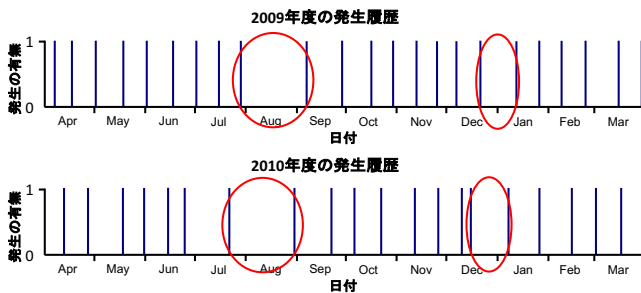


図 2 年を単位とした相似形で発生する作業の例

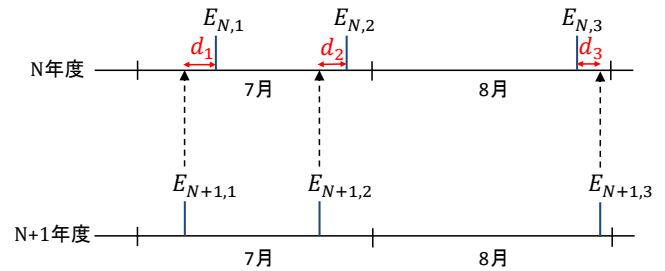


図 3 作業発生履歴の系列における自己相関

する手法 [5] や、写真撮影時の時刻とその時間的距離から写真を分類する手法 [6] もある。

予定は、一般的に時系列で管理されており、計画立案においても予定の実施日を参考にすることが多い。このため、カレンダー情報においては、実施日という特徴が最も重要な特徴の 1 つである。そこで、本稿では、カレンダー情報をもつ実施日という特徴を扱う方法として、予定間に存在する周期性に着目してクラスタリングをおこなう手法を述べる。

2. カレンダーの予定間に存在する周期

2.1 予定間の曖昧な周期性

カレンダーに含まれる予定の分類を行った調査 [7] によると、ある組織におけるカレンダー全体の予定のうち、約 30% が固定的な周期を持つ予定で、約 60% が曖昧な周期を持つ予定であり、ほぼすべての予定は何らかの周期性をもつことが確認されている。このため、予定のクラスタリングにおいて、予定間の周期性は有用な特徴であると考えられる。ここで、曖昧な周期とは、毎週月曜日といった固定的な周期で予定が発生するとは限らないが、同様の作業が繰り返し発生し、年を単位とした系列がおおよそ相似形になっているという性質を示す。

この性質について、図 2 を用いて概略を説明する。図 2 は、あるミーティングにおける 2009 年度と 2010 年度の 2 年間の発生履歴を示している。横軸は日付で、縦軸は発生の有無を 1/0 で表現している。作業発生が年を単位とした相似形をもつという性質が明確にあらわれている部分として、7 月下旬から 9 月上旬にかけての作業発生がある。この部分は他の作業発生と比べ、発生間隔の幅が広がっている。2009 年度においては、7 月下旬から 9 月上旬まで作業が発生していない。また、2010 年度においては、7 月下旬から 8 月下旬まで作業が発生していない。各年度において 7 月下旬から 8 月下旬が共通して作業が発生していない期間である。2009 年度と 2010 年度の作業の発生履歴から、2011 年度も 7 月下旬に作業が発生し、その次は 8 月下旬に発生すると推測できる。12 月から 1 月までの作業発生についても同様のことが言える。このように多くの予定は作業発生が年を単位とした相似形をもつという性質を持つ

いる。

次節にて、予定がもつ曖昧な周期をクラスタリングに用いる特徴として表現する方法について説明する。

2.2 曖昧な周期の表現

前節にて、多くの予定が年を単位とした相似形で表現される曖昧な周期をもつと述べた。この性質は、言い換えれば、予定間に存在する周期が安定しているかを表現する特徴である。このため、予定間に存在する曖昧な周期の安定度合いを定量化する指標があれば、クラスタリングの特徴として用いることができる。ここで、曖昧な周期の安定度合いを表現する指標としてリカーレンスの信頼性が定義されている [8]。これは、カレンダーに登録されている個々の予定を **タスク**、周期的に繰り返し発生する同様のタスクの集合を **リカーレンス** と定義し、リカーレンス内のタスクが表現する周期の安定度合いを表現する指標である。また、リカーレンスは、本稿において、クラスタリングによって作成する予定の集合に等しい。

リカーレンスの信頼性は、作業発生履歴を 1/0 の系列と考えたときの系列の自己相関を用いて評価する。図 3 を用いて、まず、作業発生履歴の系列における自己相関について説明する。図 3 は N 年度と $N+1$ 年度の 7、8 月付近の作業の発生履歴である。 $E_{N,i}$ は N 年度の i 回目の発生日を表す。同様に、 $E_{N+1,i}$ は $N+1$ 年度の i 回目の発生日を表す。 $E_{N,1}$ に注目すると、 $E_{N,1}$ は $E_{N+1,1}$ の日付の 1 年前付近に発生していることが分かる。このとき、 $E_{N+1,1}$ には対応する発生日が 1 年前に存在すると言える。同様に、 $E_{N+1,2}$ の日付の 1 年前付近に $E_{N,2}$ が対応する。このように、1 年前の系列と対応する組が作れるとき、系列は年を単位とした相似形であると言える。次に、系列の自己相関の強さについて説明する。 $E_{N+1,1}$ の 1 年前の日付と $E_{N,1}$ の日付の距離を d_1 とする。同様に、 $E_{N+1,2}$ の 1 年前の日付と $E_{N,2}$ の日付の距離を d_2 とする。このとき、 i は組の数だけ存在し、 d_i の総和が小さいほど系列の自己相関が高くなる。系列の自己相関が高いほど相関が強くなり、系列が相似形に近いことを表す。

作業の発生履歴の系列における自己相関関数の定義を式 1 に示す。なお、 τ はラグであり、 D は作業発生日の日付の集合、 $f(t)$ は日付 t に最も近い D の要素を返す関数を

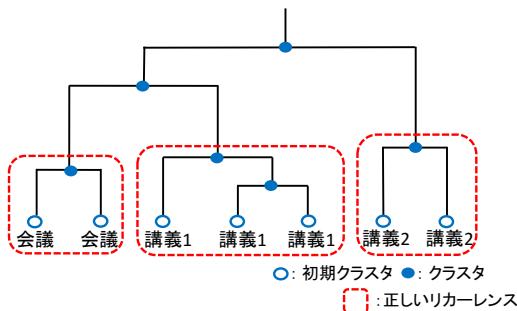


図 4 階層構造に整理したクラスターの例

表す。 $R(\tau)$ の値が小さい、つまり、自己相関が高いことは、系列が τ 日を単位とした相似形に近いことを意味する。

$$R(\tau) = \sum_{t \in D} (f(t) - f(t - \tau))^2 \quad (1)$$

この自己相関関数において、 $\tau = 365$ とすることで、リカーレンス内の作業発生が年を単位とした相似形であるかどうかを評価することができる。

3. タスクのクラスタリング手法

3.1 方針

本稿では、タスクの特徴として、2章にて述べた予定間の周期性に着目したクラスタリング手法を提案する。クラスター内の作業発生系列に対する自己相関がなるべく高くなるようにクラスタリングをおこなうことで、周期が安定した、つまりリカーレンスに相当するクラスターを作成可能である。また、クラスタリングに用いるタスクの特徴として、タスク名も用いる。タスク名は、カレンダーに登録される予定のタイトルである。提案手法では、タスク名をクラスタリングの初期状態を作成する際に用いる。

また、カレンダーの予定を整理する状況において、あらかじめいくつのグループに分類するかを決定しておくことは少なく、グループの数は、整理の結果得られるものである。このため、クラスタリング手法の1つであるk平均法など、あらかじめクラスター数を与える必要がある非階層型クラスタリング手法はタスクのクラスタリングにおいては適用できない。そこで、本稿では、クラスター数を決定しておく必要がない、階層型クラスタリング手法を応用した手法を提案する。

3.2 階層構造を用いたクラスタリング

階層型クラスタリングは、全ての対象ノードをノード1つからなるクラスターとみなし、全てのノードが1つのクラスターに結合されるまで、最も類似している（クラスター間の距離が最も小さい）クラスターの組を結合することで階層構造を作成するクラスタリング手法である。提案手法において、対象ノードはタスクを意味する。また、クラスタリング開始時のクラスターを初期クラスターと呼ぶ。階層型クラ

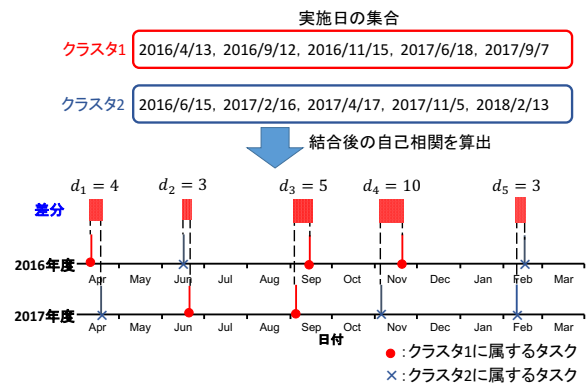


図 5 クラスタ間距離を算出する例

スタリングでは、結合の様子を階層構造で表現する。図 4 に、クラスタリングの結果得られる階層構造を樹形図として表した例を示す。図 4 は、「会議」、「講義 1」、および「講義 2」という 3 種類のリカーレンスに分類されるべき 7 つのタスクをクラスタリングした結果の階層構造を示している。クラスタリングの結果、初期クラスターを除いて 6 つのクラスターが得られている。ユーザはこれらのクラスターの中から、正しいリカーレンスを表現している 3 つのクラスターを選択することでタスクの整理を行う。

3.3 リカーレンスの信頼性によるクラスター間距離

2章にて、リカーレンスを表現する性質としてタスクがもつ周期性について述べた。多くのリカーレンスは作業発生が年を単位とした相似形をもっており、2.2 節で述べたリカーレンスの信頼性が高い。このため、クラスタリングによって、できるだけリカーレンスの信頼性が高いクラスターを作成することが望まれる。つまり、結合した際に得られるクラスターの信頼性がより高くなるようにクラスター間の距離を与える必要がある。そこで、提案手法では、式 1 で定義されるリカーレンスに含まれるタスクの実施日に対する自己相関をクラスター間距離として用いる。ここで、式 1 で算出される自己相関は、対象となるタスク数が増えると値が増加する傾向にある。このため提案手法では、自己相関の値を、算出する際に得られる対応するタスクの組の数で正規化したものを用いる。図 5 は自己相関によってクラスター間距離を算出する例を示したものである。図 5 では、それぞれ 5 つのタスクをもつクラスター 1 とクラスター 2 の距離を算出している。2 つのクラスターを結合して得られるクラスターに対して、式 1 によって算出される自己相関は $d_1 + d_2 + d_3 + d_4 + d_5 = 25$ である。また、実施日の系列中に 5 つの対応するタスクが存在するため、クラスター間距離は $\frac{25}{5} = 5$ となる。ここで、自己相関を用いて距離を測定する方法では、クラスタリングの開始時に問題が発生する。自己相関は 1 年前の対応するタスクとの日数差によって算出するため、タスクを 1 つしかもたない初期クラスターのようなタスク数の少ないクラスターにおいて、偶然誤った組合

わせが結合される可能性が高くなる。これに対し、初期クラスタのタスク数を大きくすることで対処する。3.4節にて、タスク名を用いて初期クラスタを作成する方法について述べる。

3.4 タスク名を用いた初期クラスタの作成

人がカレンダーに予定を登録する際、繰り返される同様の予定に対してある程度規則性のあるタスク名をつけることが多いと考えられる。そこで、実際のカレンダーのデータを用いて、リカーレンス内におけるタスクの命名規則の種類を調査した。調査には著者らの所属する研究グループの共有カレンダーから抽出した280件のタスクを用いた。また、このデータは30種類のリカーレンスで表現される。調査の結果、リカーレンス内のタスク名には、大きく3種類の分類が存在することがわかった。以下に3種類の分類を示す。

- (分類1) 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(送り仮名が異なる、用いる単語が異なる等)のタスク名

(分類1)、(分類2)に関しては、タスク名をつける際、リカーレンスにおけるタスクの命名規則を意識して作成されたタスク名であるといえる。そこで、提案手法では、このようなタスク名をもつタスクを同名のタスクとして扱う。一方、(分類3)に関しては、別名のタスクとして扱う。

次に、同じリカーレンスに含まれるタスクに関して、タスク名がどの程度統一されているかを知るためにリカーレンス内の同名タスクの割合を調査した。調査には、タスク名の分類を調査したのと同じデータを用いた。調査結果の表を表1に示す。

表1より、リカーレンス内において、同名タスクは約84%の割合で存在することがわかった。このため、実際のカレンダーにおいて、繰り返し発生するタスク、すなわち1つのリカーレンスに含まれるタスクでは、そのタスク名はある程度統一されていると考えられる。そこで提案手法では、クラスタリングを始める際、同名タスクをあらかじめまとめたものを初期クラスタとして作成する。

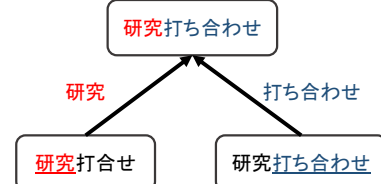
3.5 クラスタ名の生成

クラスタリングによって階層構造を作成すると、階層中に多くのクラスタが作成される。これらのクラスタに対し、ユーザが一目で望ましいクラスタかそうでないかを判断可能なクラスタ名を与えることで、クラスタを選択する作業を支援することができる。そこで、正しくクラスタリングできているクラスタ、すなわち1種類のリカーレンスに分類されるべきタスクのみを含んだクラスタに対しては、そのリカーレンスを代表する名前を与え、クラスタ内に複数のリカーレンスのタスクが混在している場合は、一見して

表1 リカーレンスに含まれるタスクにおける同名タスクの割合

通番	作業名	タスク数(個)			同名タスクの割合(%)
		同名タスク	その他	計	
1	研究打合せ	68	1	69	98.6
2	進捗報告会	36	1	37	97.3
3	ミーティング	34	0	34	100.0
4	幹事打合せ	17	16	33	51.5
5	誕生日会	3	2	5	60.0
6	忘年会	3	1	4	75.0
7	M2 中間発表	2	1	3	66.7
8	B4 中間発表	2	1	3	66.7
9	M1 論文紹介	3	0	3	100.0
10	M2 論文紹介	6	1	7	85.7
11	オープンキャンパス	6	1	7	85.7
12	特別研究発表会	3	0	3	100.0
13	修士論文発表会	3	0	3	100.0
14	暑気払い	3	0	3	100.0
15	研修会	0	3	3	0.0
16	新人歓迎会	2	1	3	66.7
17	研究室送別会	4	1	5	80.0
18	送別会	0	3	3	0.0
19	大掃除	3	4	7	42.9
20	推薦入試	2	0	2	100.0
21	編入試験	3	0	3	100.0
22	卒業式	3	0	3	100.0
23	研究室忘年会	3	0	3	100.0
24	防災訓練	3	0	3	100.0
25	レクリエーション	11	0	11	100.0
26	ボウリング大会	4	5	9	44.4
27	前期入試	3	0	3	100.0
28	後期入試	2	1	3	66.7
29	入学式	3	0	3	100.0
30	新人幹事歓迎会	2	1	3	66.7
計	-	236	44	280	84.3

<正しく分類されたクラスタの名前生成>



<複数のリカーレンスからなるクラスタの名前生成>

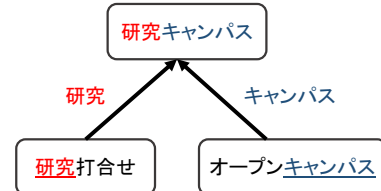


図6 クラスタ名生成の例

おかしいと判断可能な名前を与える方法を検討した。

以下に、クラスタリングの結果が階層構造となることに着目して検討したクラスタ名生成手法について述べる。階層構造において、初期クラスタを除く全てのクラスタは、結合元の2つのクラスタが存在する。そこで、結合元のクラスタをそれぞれ形態素に分割し、一方のクラスタの上位半分の形態素と、もう一方のクラスタの下位半分の形態素を結合した名前をクラスタ名として与える。図6に、正しく分類できているクラスタと、複数のリカーレンスのタ

スクが含まれるクラスタそれぞれの場合においてクラスタ名を生成する例を示す。図6の正しく分類できている例では、「研究打合せ」というクラスタと「研究打ち合わせ」というクラスタを結合している。結合されたクラスタには、「研究打合せ」の「研究」と、「研究打ち合わせ」の「打ち合わせ」を結合することで、「研究打ち合わせ」という名前を生成する。一方で、複数のリカーレンスのタスクを含むクラスタでは、「研究打合せ」というクラスタと「オープンキャンパス」というクラスタを結合している。この場合、「研究打合せ」の「研究」、「オープンキャンパス」の「キャンパス」を結合した「研究キャンパス」というクラスタ名を生成する。また、初期クラスタに対しては、クラスタに含まれるタスクの名前が一致しているため、そのタスク名をそのままクラスタ名として与える。この方法でクラスタ名を生成することで、1つのリカーレンスに含まれるべきタスクのみを要素としてもつクラスタに対して、一見してユーザが正しいと判断可能な名前を与えられる。

3.6 手順

提案手法におけるクラスタリングの手順を以下に示す。

- (1) 未整理のタスクのうち、同名のタスクをまとめて初期クラスタを作成する。
- (2) 全てのクラスタの組み合わせに対し、クラスタ間距離を算出する。
- (3) クラスタ間距離が最も小さくなる2つのクラスタを結合し、新たなクラスタを作成する。
- (4) 新たに作成したクラスタに対し、クラスタ名を生成する。
- (5) 全てのタスクを要素としてもつクラスタが作成されるまで、(2)の処理に戻る。

次章にて、本手法によるクラスタリングの評価について述べる。

4. 評価

4.1 評価観点

4.1.1 ユーザにかかる整理の手間

本評価では、クラスタリング結果を用いてタスクを整理する際に、ユーザにどの程度の手間がかかるかを評価観点とする。ここで、タスクの整理作業とは、階層構造の中から正しく分類されているクラスタを選択し、リカーレンスとして整理することである。この整理作業に対する手間は、クラスタリングの精度だけではなく、クラスタリング結果を表示するユーザインタフェース形態にも影響される。このため、まず、ここで想定するユーザインタフェースについて述べる。

4.1.2 想定するユーザインタフェース

図7に想定するユーザインタフェースを示す。想定するユーザインタフェースはプルダウン形式で階層構造をたど

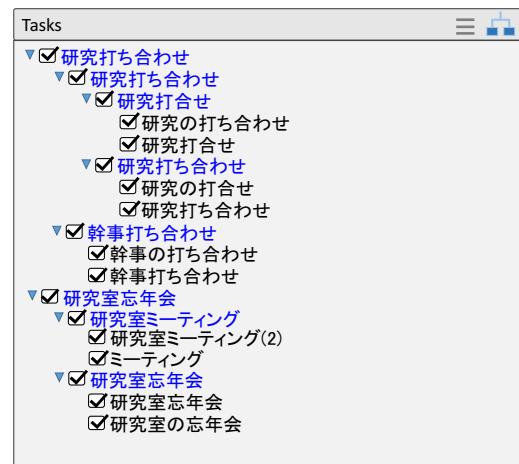


図7 階層構造をもつクラスタリング結果を扱うユーザインタフェース

ることが可能なものである。ユーザは、階層構造を上位のクラスタから確認していき、正しいと判断したクラスタをチェックボックスで選択することで、整理を行う。

4.1.3 理想的な階層構造と評価尺度

想定するユーザインタフェースを用いた、タスクの整理作業にかかる手間を評価する。そこで、まず、ユーザにかかる手間が小さくなるような理想的な階層構造を検討した。なぜならば、クラスタリングの結果得られた階層構造が理想的な階層構造にどの程度近いかによってクラスタリング結果によってどの程度手間がかかるかを評価可能なためである。

理想的な階層構造は、できる限り少ない操作回数で全てのタスクが整理可能な階層構造である。想定するユーザインタフェースによる整理作業において、ユーザは、正しく分類されているクラスタをチェックボックスを用いて選択するという操作を行う。理想的な階層構造では、選択する必要があるチェックボックスの数が最小、すなわち、最終的に分類されるべきリカーレンスの数に等しくなっている。理想的な階層構造ではない場合、ユーザは、最終的に分類されるリカーレンスの部分集合であるクラスタ（以下、部分集合クラスタと呼ぶ）を複数選択しなければならない。このため、選択する必要があるチェックボックスの数は、階層構造中に存在する各リカーレンスの部分集合クラスタの数に等しいと考えられる。そこで、以下に示す尺度を用いて階層構造をもつクラスタリング結果に対する整理にかかる手間の評価を行った。

(尺度) 各リカーレンス毎の、階層構造中に存在する部分集合クラスタの数

4.2 評価環境の準備

まず、評価に用いたデータについて述べる。評価用データには、実際のカレンダーに登録されたタスク（以下、実データと呼ぶ）ではなく、機械的に作成したタスク（以下、ダミーデータと呼ぶ）を用いた。例として、研究打合せとい

表 2 各手法におけるリカーレンス毎の部分集合クラスタ数の平均の比較

手法	同名タスクの割合 (%)										
	0	10	20	30	40	50	60	70	80	90	100
(A) タスク名のみを用いた手法 (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)(個)	3.7	3.3	3.0	3.2	2.6	2.0	1.6	0.6	0.3	0.1	0

うリカーレンスに似た性質をもつダミーデータを作成する場合について述べる。研究打合せは「2週間間隔で発生」、「土日祝日には行われない」、および、「長期休暇には行われない」といった性質をもつ。ダミーデータはこれらの性質のみを似せ、「n日間隔で発生」、「特定の曜日に影響される」、および「非発生期間がある」といった性質をもつデータとなる。このダミーデータを用いて評価を行った。具体的には、3.4節における調査に用いた実データで表現されていた30種類のリカーレンスを対象に、それらのリカーレンスと似た性質を持つ30種類のダミーデータを作成し、評価に用いた。

また、提案手法では、リカーレンス中の同名タスクの割合によって、作成される初期クラスタの大きさが変わる。このため、同名タスクの割合はクラスタリングに大きく影響すると考えられる。そこで、同名タスクの割合と評価結果との関係を確認するために、ダミーデータにおける同名タスクの割合を変えて評価を行った。ここで、ダミーデータは機械的に作成するデータであるため、同名タスクの割合を任意の値に設定することは難しい。このため、本評価では、初期クラスタの大きさを変えることで、同名タスクの割合を擬似的に表現した。

4.3 実測評価

4.3.1 評価方法

本評価では、正しいリカーレンスにあらかじめ整理されているタスクを未整理のタスクであるとみなしてクラスタリングを行った。また、本評価では、正しいリカーレンスをクラスタリングの正解データとした。4.2節で述べたように、本評価では初期クラスタの大きさを変えてリカーレンス中の同名タスクの割合を擬似的に表現した。具体的には、各正解データに含まれるタスクのうち、任意の割合のタスクをランダムに選択し、同名タスクであるとみなし、初期クラスタを作成した。本評価では、0%から100%の範囲において、10%ずつ同名タスクの割合を変え、それぞれの場合について評価した。また、本評価では、各リカーレンスにおける部分集合クラスタの数を平均した値を評価値として、提案手法とタスク名のみを用いた場合の比較を行った。以下に評価手順を述べる。

- (1) 各リカーレンスについて、設定した割合のタスクをランダムに同名タスクとする。
- (2) 各手法によりクラスタリングをおこなう。

- (3) 各手法のクラスタリングの結果に対して、評価尺度での評価をおこなう。

4.3.2 評価結果

4.1.3項で述べた評価尺度により評価を行った。提案手法を用いた場合とタスク名のみを用いた場合について、それぞれのクラスタリング結果に対する、同名タスクの割合毎の評価結果を表2に示す。直感的にいうと、タスク名のみを用いた手法の評価において、同名タスクの割合が100%の場合は、1回のチェック操作で同様の予定の集合を整理できることを示している。同様に50%の場合では、11回のチェック操作で整理できることを示している。表2から、全ての同名タスクの割合において、提案手法を用いたクラスタリング結果における各リカーレンスの部分集合クラスタの数の平均、つまり整理の際に選択する必要があるチェックボックスの数の平均が、タスク名のみを用いた場合と比べ小さくなっていることがわかる。このことより、予定間の周期性を表現する指標であるリカーレンスの信頼性をクラスタリングの要素として用いることで正しく結合されるクラスタが存在しているといえる。

また、同名タスクの割合が小さい場合において、提案手法を用いた場合の評価結果とタスク名のみを用いた場合との差が大きくなっていることがわかる。例えば、同名タスクの割合が90%、80%、70%の場合では、提案手法を用いた場合とタスク名のみを用いた場合の差は、それぞれ、0.1、0.3、0.6であることに對し、同名タスクの割合が30%、20%、10%の場合では、3.2、3.0、3.3となっている。つまり、リカーレンス中の同名タスクの割合が30%以下となるような場合では、提案手法を用いることによって、チェックボックスを選択する作業が各リカーレンス毎に約3回減少することを示している。

5. おわりに

カレンダーの予定をクラスタリングする手法について述べた。まず、カレンダーに登録されている予定を類似している予定のグループにまとめておく、つまり、カレンダーの予定を着目した特徴に基づいてクラスタリングすることで、計画立案に有用であることを述べた。

そして、クラスタリングに用いる予定の特徴として、予定間に存在する曖昧な周期について述べた。具体的には、周期的に発生する予定が年を単位とした相似形をもって発生することに着目し、その自己相関を算出することで、予

定間の周期を表現する指標であるカーレンスの信頼性を挙げた。

次に、カレンダーに登録されている予定をクラスタリングする手法として、予定間の周期に着目した、階層型クラスタリング手法を提案した。

最後に、チェックボックスの選択によって整理を行うユーザインタフェースを想定し、提案手法の評価について述べた。評価では、提案手法を用いた整理作業にかかる手間を観点とし、整理の手間を表現する尺度として、選択する必要があるチェックボックスの数に着目した。また、予定間の周期性を特徴として用いたことによる影響を調べるために、提案手法を用いた場合と、予定の名前のみを用いた場合において評価を行った結果について述べた。評価の結果、提案手法を用いることで選択する必要があるチェックボックスの数が小さくなることを示し、予定間の周期性が予定のクラスタリングに有用であることを示した。

残された課題として、予定間の周期性だけでなく、日付や、曜日の情報を特徴としてクラスタリングに用いる方法の検討がある。

参考文献

- [1] 株式会社アイ・ティ・アール：クラウド時代のコラボレーション・ツールの方向性, ITR White Paper (オンライン), 入手先 (https://www.itr.co.jp/library/whitepaper/ITR_WP_C10090023-pdf.html) (参照 2017-04-14).
- [2] 富浦洋一, 田中省作, 日高 達：共起データに基づく名詞の多次元空間への配置, 人工知能学会論文誌, Vol. 19, pp. 1-9 (2004).
- [3] 笠原 要, 稲子希望, 加藤恒昭：単語の属性空間の表現方法, 人工知能学会論文誌, Vol. 17, pp. 539-547 (2002).
- [4] Qamra, A., Tseng, B. and Chang, E. Y.: Mining Blog Stories Using Community-based and Temporal Clustering, *Proceedings of the 15th ACM International Conference on Information and Knowledge Management, CIKM '06*, New York, NY, USA, ACM, pp. 58-67 (2006).
- [5] 樋口賢治, 原田史子, 島川博光：J-003 情報整理のための手動クラスタリングによる人の興味の抽出, 情報科学技術フォーラム講演論文集, Vol. 7, No. 3, pp. 405-408 (2008).
- [6] Ryu, D.-S., Park, S.-Y., Kim, K. and Cho, H.-G.: A priority queue-based hierarchical photo clustering method using photo timestamps, *Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on*, Vol. 3, IEEE, pp. 152-156 (2011).
- [7] 吉井英人, 北垣千拡, 乃村能成, 谷口秀夫：作業発生の規則性に基づく作業予測手法と評価, 情報処理学会論文誌, Vol. 56, No. 2, pp. 543-553 (2015).
- [8] 北垣千拡, 乃村能成：作業発生の規則性に基づく作業予測の信頼性評価手法, 情報処理学会研究報告, Vol. 2015-DPS-162, No. 9, 情報処理学会第 162 回マルチメディア通信と分散処理・第 68 回コンピュータセキュリティ合同研究発表会, pp. 1-8 (2015).