

リカーレンスの信頼性を用いたタスク分類手法

北垣 千拡¹ 乃村 能成¹

概要: 過去のカレンダー情報に基づいて周期的に発生する作業から将来の作業の発生を予測する手法がある。この手法では予測に先立ち、カレンダーの予定(タスク)を同様のタスクの集合(リカーレンス)として分類しておく必要がある。この作業が完全でないと、不完全なリカーレンスを用いた予測となり、予測精度が低下してしまう。そこで本研究では、カレンダー情報からタスク間の関係を学習し、未分類のタスクに対して分類先のリカーレンスを提示することで、ユーザの分類の手間を軽減する。

キーワード: カレンダー, 計画立案, 情報整理

1. はじめに

オフィスにおける様々な作業は、何らかの周期性や関連性を持つことが多い。たとえば、会議という作業は「約2週間に1回」といった周期を持ち、会議に関連して「会議通知」、「会場準備」、「議事録送付」といった作業を伴うことが挙げられる。このように、作業は周期性や関連性を持つため、将来の予定について計画するとき、過去の作業履歴を参考にすることが多い。たとえば、昨年の行事に関する予定を確認し、今年のおおよその実施時期を検討したり、昨年の当該行事の周辺に関連する行事があったか確認したりすることがある。このように、カレンダーの情報は将来の計画を立てる作業(計画立案)や仕事の引継ぎの際に有用であるといえる。IT分野専門のリサーチ企業の調査においても、労働者の8割がスケジューラを利用していることが確認されており[1]、日常において、カレンダーの情報を活用する場面は多い。

情報は整理することにより有効活用される。そこで、日々発生する作業や会議の予定に関する情報を「カレンダー情報」とし、それらの整理について考えた場合、カレンダー情報は収集後の整理が適切にされているとは言い難い。これは、これまでカレンダー情報を整理するといった考え方がなく、カレンダーアプリケーションやスケジューラも整理に関するユーザ支援の方法は提供していなかったためである。そこで、我々はこれまでに、整理されたカレンダー情報とは何かについて論じ、それを表現する手段として作業の曖昧な周期性と作業間の関連性を表現するモデルを提案

し、このモデルを利用した具体的なユーザ支援の方法として作業発生の規則性を扱うカレンダーシステムを提案、実装した[2]。このシステムは、整理された予定の周期性や関連性を基に、近い将来発生しうる予定を予報する機能を持つ。予定の予報によって、将来の予定の数や時期がわかるため、計画立案の際に有用である。

この予報機能を実現するためには、予定に対して規則性の情報を事前に付与しておく必要がある。この作業を本稿では、カレンダー情報の整理と呼ぶ。この整理の実現方法として、作業発生の規則性を扱うカレンダーシステムでは、予定の新規登録時に規則性の情報を半自動で付与する方式を採用しており、これによって、予定を整理する手間を削減している。しかし、日々の生活の中では、システムの機能を用いずに予定を登録せざるを得ない場面もある。たとえば、外出中のためシステムを使える環境になく、スマートフォンのカレンダーアプリケーションにタスクを登録する場合がある。この場合、登録したタスクを作業発生の規則性を扱うカレンダーシステムに同期し、登録された予定を後から整理する必要がある。予定の整理をしなかった場合、システムの提供する機能が本来の性能を発揮できない。たとえば、予定の予報機能は予定間の周期性に基づいて将来の予定を予測するため、未整理の予定があると、システムの認識する予定の周期性が崩れ、予報の精度が低下してしまう。このため、ユーザは予定の整理漏れがないよう整理を定期的に行う必要がある。

しかし、予定の整理はユーザにとって手間のかかる作業である。このため、整理にかかる手間をできるだけ軽減することが望まれる。そこで本研究では、予定の整理先をリコメンドすることで整理の手間を軽減する。提案手法で

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

は、予定が既存のグループのどこに属するかを分類し、その分類先を整理先としてリコメンドする。

クラス分類やクラスタリングの分野において、高い分類精度を実現するために、特徴選択が重要である。また、このパラメータは分類対象の特徴によって異なる。たとえば、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求めたり [3][4]、写真の分類においては、写真撮影時の時刻とその時間的距離を用いたり [5] する。また、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 [6] もある。本研究では、周期的に発生する予定の周期性に着目し、発生周期に関する自己相関を示す指標であるリカーレンスの信頼性 [7] を用いて分類を行う。

以降では、まずはじめに、作業発生の規則性を扱うカレンダーシステムの機能と課題について述べる。次に、予定の分類手法について述べ、最後に、提案手法の評価について述べる。

2. 作業発生の規則性とカレンダーシステム

2.1 作業発生の規則性

我々は、作業発生の周期性と関連性を表現する**作業発生の規則性**というモデルを提案した [8]。作業発生の規則性は、作業間の関係を集合の包含関係で定義している。以下に概要を述べる。

タスク (Task): タスクは、作業を扱う最小の単位である。タスクは開始時刻と終了時刻を持ち、この間で連続的に行われる作業を表現する。

リカーレンス (Recurrence): リカーレンスは、タスクを要素とする集合である。リカーレンスは繰り返し発生している同様のタスクを1つの集合とする。例として、「会議」というリカーレンスについて考える。ここで、会議は1カ月周期で行われるものだという共通認識があり、実際に「第1回会議」、「第2回会議」、...の実施日の間隔が1カ月という周期を持っていたとすると、繰り返し発生するタスクである「第1回会議」、「第2回会議」、...はリカーレンス「会議」に属する。リカーレンスはうまく利用することで、将来のタスク発生を予測できる。リカーレンスを用いた作業予測の例を図1に示す。図1のように、リカーレンス内に複数のタスクが存在する場合、リカーレンス内のタスクの周期性を解析することにより、次回タスクの発生を予測できる。

ミッション (Mission): ミッションはタスクまたはミッションを要素とする集合である。ミッションは関連する複数のタスクまたはミッションを1つの集合とする。例として、「第1回会議関連の仕事」というミッションについて考える。ここで、「第1回会議」というタスクに関連して、「会議通知」、「会場準備」、「議



図1 リカーレンスを利用した作業予測

事録送付」という3つのタスクがあるとすると、これらの4つのタスクは関連する1つのミッションとして「第1回会議関連の仕事」に属する。

2.2 カレンダー情報の整理

作業発生の規則性において、カレンダー情報とは主にタスクを指し、通常のカレンダーに記入する予定に相当する。カレンダー情報の整理とは、予定に周期性と関連性を付与することであり、タスクの集合からリカーレンスとミッションを作成することである。つまり、カレンダー情報の整理とは、2.1節のモデル関係を作成することである。このように、カレンダー情報の整理では、関連性からタスクを分類する方法もあるが、本稿では、リカーレンスに関する整理を対象とし、ミッションに関する整理については取り扱わない。

カレンダーに含まれるタスクの分類を行った調査 [9] によると、カレンダー全体のタスクのうち、29.2%が固定的な周期を持つタスクで、62.6%が曖昧な周期を持つタスクであり、ほぼすべてのタスクは何らかの周期性を持つことが確認されている。このことから、カレンダーにおけるリカーレンスの割合は高く、整理によって得られる恩恵は大きいと考えられる。

2.3 作業発生の規則性を扱うカレンダーシステム

作業の規則性を扱うカレンダーシステム [2] は、将来の計画立案や仕事の引継ぎで有用である。このシステムは、以下の3つの特徴を持つ。

(特徴1) 過去のタスクを参照しながら複製するタスク登録方式

具体的には、1年前の同月付近のタスクを参照しながら簡便な複製操作で当月のタスクを作成する登録方式を指す。この操作により登録されたタスクは、複製元のタスクと同じリカーレンスとして登録されるため、リカーレンス整理の手間が軽減される。

(特徴2) 周期性の継承によるタスクの予報機能

本機能は、リカーレンス内のタスクの周期性を解析することで、近い将来発生しうるタスクを予測して提示する機能である。

(特徴3) Inboxによるカレンダー情報の整理機能

Inboxとは、予定の情報を時系列とは異なる軸で整理可能な、メールアプリケーションの受信箱に似たユーザインタフェースである。本機能を用いることで、カレンダー情報を効率的に付与できる。

2.3.1 課題

作業発生の規則性を扱うカレンダーシステムでは、カレンダー情報の整理における課題がある。課題について説明する前に、タスクを整理する状況を考える。

(ケース1) システムの機能を用いて予定を新規作成

作業発生の規則性を扱うカレンダーシステムでは、(特徴1)のタスク複製機能によってユーザのタスク登録を支援する。この場合、タスクの登録とともにカレンダー情報にリカーレンスの分類が付与される。これは既存のリカーレンスからのコピーによって入力するインタフェースになる。

(ケース2) システムの機能を用いずに予定を新規作成

(手順1)

過去に同様のタスクを登録したことを忘れており、システムの機能を使わず登録することがある。この場合、後からこのタスクに対してカレンダー情報を付与する必要がある。あるいは、タスク登録時に、このカレンダーシステムを使える環境にない場合、別のカレンダーアプリケーション(たとえば、スマートフォンのカレンダーアプリケーション)にタスクを登録する場合がある。この場合、ここで登録したタスクを作業発生の規則性を扱うカレンダーシステムに同期し、後からカレンダー情報を付与する必要がある。

(手順2)

システムの機能であるInboxを用いて整理を行う。作業発生の規則性を扱うカレンダーシステムでは、(特徴3)の機能によって未整理のタスクの整理を支援する。カレンダー情報が付与されていないタスクはすべてInboxに入る。対して、ユーザはInboxに入ったタスクにカレンダー情報を付与する。この操作は定期的に行う必要があり、Inboxが常に空になるよう整理することが望まれる。

作業発生の規則性を扱うカレンダーシステムでは、ユーザのタスク整理を支援する機能を持っており、これは(ケース1)と(ケース2)(手順2)に相当する。

しかし、これらの機能を用いてカレンダー情報の整理ができない状況も考えられる。(ケース2)の場合、カレンダー情報を付与されずにタスクが登録されるため、未整理のタスクとしてInboxに入ることになる。この場合、Inbox内にある未整理のタスクが増え、整理の手間が増大する可能性がある。通常、Inbox内にある未整理のタスクが増大することはない。なぜならば、1度整理したタスクはカレンダー情報が付与され、2度目以降に同様のタスクを登録する際は(特徴1)の機能を用いて登録されるため(ケース1に相当)、これらのタスクは2度目以降にはInboxに入らない。ゆえに、Inboxで整理するタスクの量が線形的に増大することはない。しかし、(ケース2)が多発した場合、

この限りではない。また、ユーザが扱うリカーレンスは数十を越える[2]ため、Inbox内のタスクはより効率的に整理できることが望まれる。

2.3.2 対処

Inbox内のタスクをより効率的に整理する方法として、タスクの整理先のリコメンド機能を考える。リコメンドを行う状況として、未整理のタスクを既存のリカーレンスに分類する状況を扱う。つまり、分類対象のない既存のリカーレンスが存在しないタスクについては扱わない。3章にて、未整理のタスクの分類手法について詳細に述べる。

3. 未整理のタスクの分類手法

3.1 分類手法

本章では、タスクの分類手法について述べる。今回扱う問題は、未整理のタスクを既存のリカーレンスのどこに分類されるかを解くクラス分類問題である。本研究では、一般的なクラス分類手法の1つであるk近傍法を用いる。k近傍法では、特徴となるパラメータの選択と距離の与え方が重要である。以降の節でこれらについて検討する。

3.2 タスクの特徴

本項では、タスクの分類のために必要な要素について述べる。ここで、分類に必要な要素とは、タスクの特徴そのものである。

文献[2]の調査によると、作業予測に必要な情報を**繰返情報**として以下の8つを挙げている。

作業発生間隔の平均: 作業発生間隔の平均日数と標準偏差

昨年の作業発生間隔: 昨年の同時期に発生した作業の作業発生間隔

曜日情報: 日月火水木金土の7つに祝日を加えた8つの曜日の発生割合

月情報: 1月から12月までの12の月の発生割合

週情報: 月の第1週から第5週までの5つの週の発生割合

日付情報: 1月1日から12月31日のまでの1年の各日付の発生割合

時間帯情報: 開始時刻の上位3つと発生回数の最も多い作業時間を組み合わせた時間帯の3つのパターン

最終作業情報: リカーレンス内の直近の作業に関する情報

この8つの特徴に加え、予定名も分類においては重要である。なぜならば、タスクの分類を人間が手動で行う際、多くの場合、まず初めに確認する項目が予定名であるためである。このため、以下の要素を新たに加える。

予定名: タスクのタイトル

一方、昨年の発生間隔は作業発生間隔の平均を包含しているため[2]、分類において作業発生間隔の平均は必要ない。また、最終作業情報は作業予測のためだけに必要な要素であるため、分類においては必要ない。以上より、昨年の作業発生間隔、曜日情報、月情報、週情報、日付情報、

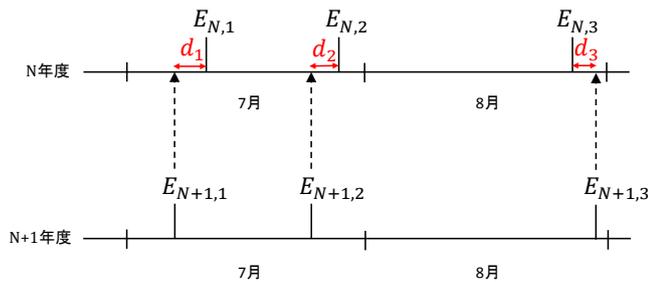


図 2 作業の発生履歴の系列における自己相関

時間帯情報、および予定名の7つの情報をタスクの特徴とし、これらを用いてタスクを分類できると考えられる。

3.3 タスク間の距離

k近傍法はオブジェクト間の距離を基にクラスを分類する手法である。そこで、3.2節にて述べたタスクの各特徴について、次のようにタスク間の距離を与える。

- (距離 1) 昨年の作業発生間隔の距離: 未知のタスクを既知のタスクのリカーレンスに分類した際、リカーレンスの信頼性(3.4節にて後述)が向上した場合0、低下した場合1を与える。
- (距離 2) 曜日情報の距離: 各タスクの曜日が同じ場合0、異なる場合1を与える。
- (距離 3) 月情報の距離: 各タスクの月の差分を与える。
- (距離 4) 週情報の距離: 各タスクの週の差分を与える。
- (距離 5) 日付情報の距離: 各タスクの日付の差分を与える。
- (距離 6) 時間帯情報の距離: 各タスクの時間帯(終日か否か)が同じ場合0、異なる場合1を与える。
- (距離 7) 予定名の距離: 各タスクのタイトルのレーベンシュタイン距離 [10] を与える。

3.4 リカーレンスの信頼性

リカーレンスの信頼性 [7] は、リカーレンスの定義に基づき、その信頼性を測るものである。2.1節で述べたように、リカーレンスとは、繰り返し発生する同様のタスクの集合である。つまり、あるリカーレンスに属するタスクの実施日はなんらかの周期をもつものとなる。このため、この周期の乱れ具合を算出する、つまり、リカーレンスの自己相関を算出することでリカーレンスの信頼性の指標となると考えられる。

リカーレンスの信頼性について、図 2 を用いて詳細に説明する。あるリカーレンスに属するタスクの集合は、タスクの実施日を1、それ以外を0とすると1/0のインパルス列とみなせる。このインパルス列は、系列の間隔のみでその性質が決まるため、リカーレンスの自己相関は、時間軸方向のずれのみから評価されるべきである。図 2 を用いて、時間軸方向のずれのみの評価に着目した自己相関の考

え方について説明する。図 2 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。まず、 $E_{N,1}$ 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。ここで、時間軸方向のずれに着目した評価を行う。 $E_{N+1,1}$ 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、系列が相似形に近いことを表す。

4. 評価

4.1 評価観点

未整理のタスクをリカーレンスに分類した際の分類精度を評価観点とする。ここで分類精度とは、未整理のタスクを正しい整理先に分類できた割合のことを指す。分類の際、提案手法で扱うパラメータのうち、リカーレンスの信頼性を加えたことによる分類精度の向上を確認する。そこで、リカーレンスの信頼性を加えた場合と加えない場合についてそれぞれ評価する。また、リカーレンスの信頼性がタスクの周期性を測る手法であることから、実験データには、何らかの周期性を持つタスクを用いて実験を行う。周期性のない不定期に発生するタスクは実験に用いない。以上を考慮して、未整理のタスクの分類精度を評価する。

4.2 評価環境の準備

まず、評価に用いる実験データについて述べる。実験データには、実際のカレンダーに登録されたタスクではなく、機械的に作成したタスクを用いる。なぜならば、実際のカレンダーに登録されたタスク(実データ)は複雑な要因から実施日が決まるため、評価に適していないためである。そこで、実際のカレンダーに登録されたタスクとよく似た性質をもつダミーデータを作成し、これを用いる。ダミーデータとは、実データのリカーレンスと似た性質をもつ、機械的に作成されたリカーレンスを指す。例として、研究打合せとい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を考える。研究打合せは「2週間間隔で発生」、「土日祝日には行われぬ」、「および、「長期休暇には行われぬ」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う。具

体的には、著者らの所属する研究グループの36種類のリカーレンス [2] を対象に、それらのリカーレンスと似た性質を持つ36種類のダミーデータを作成し、実験に用いた。なお、ダミーデータの作成に際して、タスクの特徴である**予定名**は除いた。理由として、ダミーデータでは規則的なタスクを作成するため、曖昧な予定名を与えることが困難なためである。通常、タスクの予定名は人間が過去の記憶を頼りに入力することが多く、ユーザがすべてのタスクの予定名を意識して統一しているとも言い難いため、曖昧になりやすい。このことから、実際のカレンダーのタスクの予定名は曖昧さを持つが、ダミーデータでは評価に適した曖昧さを持つ予定名を生成することが困難であるため、今回の実験ではパラメータから除いた。今回の実験では予定名を用いないが、予定名とその他のパラメータは独立であるため、互いに影響を受けず、予定名を加えた場合でも、性能が向上することはあっても低下することはないと考えられる。

次に、リカーレンスの信頼性を加えた場合と加えない場合の比較について述べる。4.1節の評価観点に基づき、リカーレンスの信頼性の効果を明確にするために、以下の3つの手法でそれぞれ実験を行う。

手法1 (提案手法) : 距離1, 2, 3, 4, 5, 6を用いる
 つまり、日付に関する統計情報(曜日情報, 月情報, 週情報, 日付情報, および時間帯情報)に加え、リカーレンスの信頼性を用いた手法である。

手法2: 距離2, 3, 4, 5, 6を用いる
 日付に関する統計情報のみを用いる手法である。

手法3: 距離1を用いる
 リカーレンスの信頼性のみを用いる手法である。

手法1と手法2を比較することで、リカーレンスの信頼性を加えたことによる精度の向上を示せる。また、手法3によって、リカーレンスの信頼性のみでの分類精度を示せる。

なお、上記の3つの手法のそれぞれについて、分類先の候補として第1候補, 第2候補, および第3候補まで提示する。これは、タスクのリコメンドにおいて、必ずしも第1候補のみで正しく分類できる必要はないためである。なぜならば、ユーザは既存のタスクをどこに分類すべきかを吟味したり、すでに分類先のリカーレンスを作成済みかどうかについて吟味したりする時間を削減できるという意味で、ある程度絞られてさえいれば、必ずしも第1候補のみで分類できる必要はないためである。

4.3 評価方法

4.3.1 想定する状況

詳細な評価方法について述べる前に、まず、未整理のタスクを整理する状況を考える。未整理のタスクがある場合、リカーレンスは登録されるべきタスクが登録されてい

表1 未整理のタスクの分類精度

候補数	ランダムに選択した場合			先頭から選択した場合		
	1	2	3	1	2	3
手法1(提案手法)	0.583	0.667	0.722	0.528	0.639	0.778
手法2	0.500	0.556	0.639	0.556	0.611	0.750
手法3	0.389	0.389	0.389	0.139	0.139	0.139

ない不完全な状態である。この状態は、本来登録されているべきタスクがリカーレンスから抜け落ちている状態といえる。本実験では、リカーレンスからタスクがランダムに抜け落ちた状況を想定する。また、ランダムに抜け落ちる特殊例として、リカーレンスの先頭が抜け落ちた状況も想定する。ここで、リカーレンスの先頭とは、リカーレンスのタスクのうち実施日が最新の日付のタスクを指す。以上を考慮して、それぞれの状況に対して評価を行う。

4.3.2 実験手順

未整理のタスクを既存のリカーレンスに正しく分類できるかを評価する。具体的には、以下の手順で実験を行う。まず、実験データであるリカーレンス36種類を用意する。これを既存のリカーレンスとする。次に、各既存のリカーレンスからタスクを1つ選択し、選択したタスクをリカーレンスから外し、未整理のタスクとする。最後に、外した36個の未整理のタスクを提案手法を用いて分類する。分類した結果、元のリカーレンスに再度分類できているかを確認する。なお、リカーレンスから外すタスクを選択する際、4.3.1項の状況を考慮し、ランダムに選ぶ場合とリカーレンスの先頭から選ぶ場合の2パターンで実験を行う。なお、 $k=10$ 近傍を候補のオブジェクトとした。

4.4 実測評価

評価結果を表1に示す。なお、表の候補数1, 2, 3はそれぞれ、分類先の候補として第1候補, 第2候補, および第3候補まで提示したことを示す。また、ランダムのパターンは複数回試行した結果、すべて同様の傾向を示したため、1パターンのみ記載している。

まず、手法1(提案手法)と手法2を比較すると、ランダムに選択した場合においては、すべての候補数で手法1(提案手法)が手法2の分類精度を10%程度上回る結果となった。このことから、ランダムに選択した場合においては、リカーレンスの信頼性が効果的に働き、より良い精度が得られたことがわかる。一方、先頭から選択した場合においては、候補数2と3の両方で手法1(提案手法)が手法2の分類精度を2.8%上回ったが、候補数1では2.8%下回る結果となった。実験では、36種類のリカーレンスからそれぞれ1つ分のタスクを未整理にし、分類対象としたため、 $2.8\% (= \frac{1}{36})$ はタスク1つ分の差となる。このため、先頭から選択した場合においては、手法1(提案手法)と手法2にそれほど差がないことがわかる。ここで、手法3(リカーレンスの信頼性のみ)の分類精度に注目する。分類精度が

向上したランダムに選択した場合には、手法3はすべての候補数で38.9%である一方、分類精度が向上しなかった先頭から選択した場合には、手法3はすべての候補数で13.9%であった。このことから、先頭から選択した場合において、リカーレンスの信頼性が効果的に働いておらず、手法1(提案手法)と手法2が同程度の結果となったことがわかる。この理由として、先頭から選択した場合、元々のリカーレンスの信頼性が高いことが原因と考えられる。つまり、リカーレンスの両端からタスクを1つ除いたとしても、残ったタスクは抜けのない完全なリカーレンスのままであるため、未整理のタスクを正しいリカーレンスに分類しても信頼性が向上しないためであると考えられる。

次に、各手法共通の特徴として、候補数を増やせば、分類精度が向上することについて述べる。これは当然の結果であるが、タスクの分類をリコメンド機能として利用する場合において、実用性を判断する際に必要となる。手法1において、候補数1と3を比べると、ランダムに選択した場合 $72.2 - 58.3 = 13.9\%$ 、先頭から選択した場合 $77.8 - 52.8 = 25.0\%$ 分類精度が向上している。手法1では72.2から77.8%、つまり、5回に4回は適切なリコメンドができる。また、手法2では63.9から75.0%、つまり、4回に3回は適切なリコメンドができ、ある程度候補数を増やすことで、ユーザにとって十分なリコメンドが提供できることがわかった。また、手法3において、すべての候補数で同じ精度となる理由は、手法3ではパラメータが1つしかなく、かつ、リカーレンスの信頼性は同じリカーレンスに属するタスクですべて同じ値をとるため、 $k=10$ 近傍の範囲に同じリカーレンスに属するタスクしかないためであると考えられる。

5. おわりに

本稿では、まず、作業発生の規則性を扱うカレンダーシステムの課題として、カレンダー情報の整理における課題について述べた。具体的には、システムの提供する機能を使ってカレンダー情報が整理されなかった場合、Inbox内の未整理のタスクが増え、整理の手間が増大する可能性があるという課題である。この対処として、Inbox内のタスクをより効率的に整理する方法としてタスクの整理先のリコメンド機能を挙げ、これを実現するために、未整理のタスクの分類手法を提案した。この分類手法では、タスクの周期性に着目し、発生周期に関する自己相関を示す指標であるリカーレンスの信頼性を用いることでタスクを分類する。また、評価では、分類のパラメータとしてリカーレンスの信頼性を加えた場合と加えない場合で比較し、加えない場合と比べ、分類精度が約10%向上することを示した。また、分類結果に関して、ユーザにリコメンドする数を1から3に増やすことで、分類精度が最大25%向上することを示した。

今後の課題として、実運用されているカレンダーに登録されたタスクに対する評価がある。また、評価において、タスクが分類先のリカーレンスにおいて先頭の場合、元々のリカーレンスの信頼性が高いため、未整理のタスクを正しいリカーレンスに分類しても信頼性が向上せず、分類のパラメータとして有効に働かないということがわかった。これに関して、リカーレンスの先頭の場合、リカーレンスの信頼性が維持されるかどうか指標になると考えられる。この指標をパラメータとしてうまく取り入れるという課題がある。

謝辞 本研究の一部は、科学研究費補助金・基盤研究(C)(課題番号: 26330224)による研究費を得て実施した。ここに記して謝意を示す。

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—グループウェアの再考, ITR White Paper (オンライン), 入手先 (http://www.itr.co.jp/library/public/ITR_WhitePaper/ITR_WP_C10090023.pdf) (参照 2015-01-14).
- [2] 吉井英人, 北垣千祐, 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. 543-552 (2015).
- [3] 富浦洋一, 田中省作, 日高遠: 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).
- [4] 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).
- [5] 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).
- [6] 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).
- [7] 北垣千祐, 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).
- [8] 三原俊介, 乃村能成, 谷口秀夫, 南 裕也: 作業発生の規則性を扱うカレンダーシステムの評価, 情報処理学会論文誌, Vol. 54, No. 2, pp. 630-638 (2013).
- [9] 三原俊介, 乃村能成, 谷口秀夫: 作業発生の規則性を扱うカレンダーシステムの実現, 情報処理学会研究報告, Vol. 2011-DPS-149, No. 10, 第149回マルチメディア通信と分散処理研究発表会 (DPS), pp. 1-6 (2011).
- [10] Levenshtein, V. I.: Binary codes capable of correcting deletions, insertions, and reversals, *Soviet physics doklady*, Vol. 10, No. 8, pp. 707-710 (1966).