

特 別 研 究 報 告 書

題 目

作業発生の規則性を用いた作業予測システムの検討

指導教員

報 告 者

吉井 英人

岡山大学工学部 情報工学科

平成 24 年 2 月 9 日 提出

要約

オフィス環境において周期的に発生する作業を扱うモデルとして、作業発生の規則性というモデルが提案されている。このモデルが扱う作業の履歴から作業の周期性を抽出できれば、将来の作業発生の予測に有用である。

作業予測の手法の1つとして、時系列に並べた作業の履歴から作業発生間隔の平均日数を求め、予測に用いることが考えられる。しかし、この手法では、オフィス環境において発生する作業をうまく予測することができない。なぜならば、勤務体系や習慣から、土日や祝日には作業が発生しないことと長期休暇や年末年始といった期間には作業が発生しないことが考えられるからである。

そこで、本論文では、勤務体系や習慣を考慮した作業予測の手法を提案し、その手法を用いて将来の作業発生を予測する作業予測システムを設計した。

まず、オフィス環境において周期的に発生する作業の例として、著者が所属する研究グループで行われた打合せの発生について分析した。この結果、発生する曜日に偏りがあること、作業が発生しない期間があること、時間帯に偏りがあること、および年間における差異は大きくないことが分かった。そこで、分析の結果を考慮した手法として、「周期性の継承と発生確率を考慮した作業予測」と「年を単位とした周期性の継承による作業予測」の2つの手法を提案した。周期性の継承と発生確率を考慮した作業予測は、曜日の発生割合、作業が発生しない期間、および時間帯の情報を用いて発生間隔に修正を加える手法である。年を単位とした周期性の継承による作業予測は、去年の発生間隔を用いて作業を予測する手法である。この2つの手法の内、周期性の継承と発生確率を考慮した作業予測についてさらに検討し、必要な情報として繰返情報を定義した。また、提案手法を詳細化した予測アルゴリズムについて述べた。この予測アルゴリズムに従って将来の作業発生を予測し、予測結果を提示する作業予測システムを設計し、システムの概要、動作内容、および各部の詳細について述べた。

目次

1	はじめに	1
2	作業発生の規則性を用いた作業予測	3
2.1	作業発生の規則性	3
2.1.1	作業発生の規則性とは	3
2.1.2	作業発生の規則性を表現するモデル	3
2.2	オフィス環境におけるリカーレンスの例	5
2.3	事例分析	6
2.4	作業予測の手法	8
2.4.1	手法の提案	8
2.4.2	周期性の継承と発生確率を考慮した作業予測	10
2.4.3	繰返情報	11
2.4.4	予測アルゴリズム	12
3	作業予測システムの設計	15
3.1	概要	15
3.2	作業履歴 DB	16
3.2.1	概要	16
3.2.2	データ形式	17
3.3	繰返情報抽出部	17
3.3.1	概要	17
3.3.2	繰返情報の表現形式	18
3.3.3	繰返情報の抽出方法	23
3.4	予測部	25
3.5	UI 部	25
4	おわりに	26

謝辞	27
参考文献	28

目 次

2.1	作業発生の規則性を適用した例	4
2.2	周期性の継承による作業予測	5
2.3	2009 年度, 2010 年度の打合せの発生履歴	6
2.4	周期性の継承と発生確率を考慮した作業予測	8
2.5	年を単位とした周期性の継承による作業予測	9
2.6	予測アルゴリズム	14
3.1	作業予測システムの全体像	16
3.2	作業発生の規則性のデータ形式	17

表 目 次

2.1	曜日ごとの発生頻度	6
2.2	開始時刻の発生頻度	7
2.3	作業時間の発生頻度	7
2.4	繰返情報	12

第 1 章

はじめに

オフィス環境において計画を立てる際は、過去に発生した作業を参考にすることが多い。これは、オフィス環境における作業には繰り返し発生している作業が多く、過去の作業の発生から将来の作業の発生をある程度推測できるからである。たとえば、約2週間に1回といった間隔で発生している進捗報告会議は、最後に発生した日付から約2週間後に発生することが推測できる。このことから過去の履歴を用いることで、将来発生する作業を予測することができると考えられる。過去の履歴を利用する研究はこれまでもいくつか存在する [1][2]。また、仕事引継ぎの観点から過去の履歴の利用法を探る研究も現れてきている [3]。しかしながら、これらの研究で提案されているモデルは、複雑であり、一般的に普及しているとは言い難い。

オフィス環境において周期的に発生する作業を扱うモデルとして作業発生の規則性というモデルが提案されている [4][5]。このモデルは、シンプルかつ既存のカレンダシステムが扱うモデルと親和性の高いモデルである。このモデルは、繰り返し発生している同様の作業を履歴として扱っている。このモデルが扱う作業の履歴から作業の周期性を抽出できれば、将来の作業発生を予測することができると考えられる。また、将来の作業発生予測は、計画の立案や作業の引継ぎに有用だと考えられる。このため、本論文では、作業の履歴を用いて将来の作業発生を予測する手法について検討する。

オフィス環境において周期的に発生する作業の例として進捗報告会議を考える。この作業は、約2週間に1回発生する、上司の予定や仕事の進捗具合によって発生する日付が前後する、および発生する曜日に偏りがあるといった特徴を持つ。作業発生を予測する手法の1つとして、時系列に並べた作業の履歴から作業発生間隔の平均日数を求め、予測に用いる手法が考えられる。しかし、この手法では、進捗報告会議の発生をうまく予測することができな

い。なぜならば、勤務体系や習慣から、土日や祝日には作業が発生しないことと長期休暇や年末年始といった期間には作業が発生しないことが考えられるからである。

そこで、本論文では、勤務体系や習慣を考慮した作業予測の手法を提案する。まず、オフィス環境において周期的に発生する作業の例として、著者が所属する研究グループで行われた打合せについて分析し、その結果について述べる。次に、分析の結果を考慮した作業予測の手法を新たに提案し、その手法について述べる。さらに、提案手法に必要な情報について検討し、提案手法を詳細化した予測アルゴリズムについて述べる。そして、予測アルゴリズムに従って将来の作業発生を予測し、予測結果を提示する作業予測システムについて述べる。

第 2 章

作業発生の規則性を用いた作業予測

2.1 作業発生の規則性

2.1.1 作業発生の規則性とは

オフィス環境における作業の発生には、ある程度決まった規則性が存在する。たとえば、進捗報告会議や忘年会といった作業は、発生に関してそれぞれ「約 2 週間に 1 回」や「毎年 12 月下旬」といった曖昧な周期を持つ。また、それぞれには、会場の予約や議事録の作成といった事前、事後の作業がともなう。このように作業発生にはある程度決まった周期性と関連性があると考えられる。この作業発生に関わる作業の周期性と関連性を組み合わせたものを「作業発生の規則性」と呼ぶ。

2.1.2 作業発生の規則性を表現するモデル

作業発生の規則性には、以下の 4 つの要素がある。

(1) 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。また、個々のタスクは開始時刻による順序関係を持つ。

(2) 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。

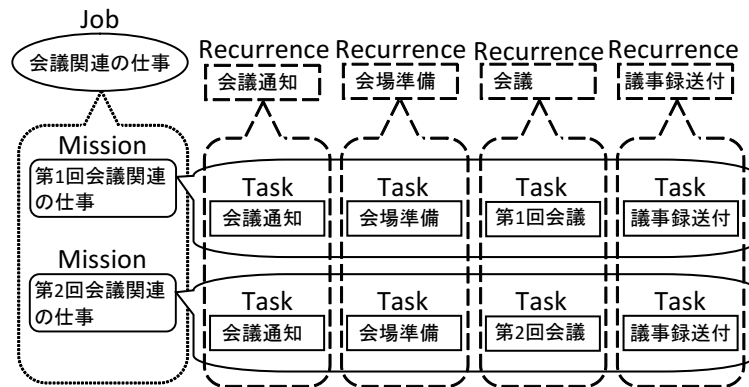


図 2.1 作業発生の規則性を適用した例

(3) 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。ミッションは複数の時間に分かれて実行される大きな粒度での作業を表現する。

(4) ジョブ (Job)

ジョブはミッションを要素とする集合である。ジョブはミッションにおけるリカーレンスであり、繰り返し発生している同様のミッションを1つの集合とする。

図 2.1 は、定例会議の例に4つのモデルを適用したものである。まず、「会議通知」、「会場準備」、「第1回会議」、および「議事録送付」はタスクとする。次に、「会議通知」、「会場準備」、「第1回会議」、および「議事録送付」の発生は関連していることを表現するために、これらのタスクを要素とするミッション「第1回会議関連の仕事」を定義する。同様に、「第2回会議関連の仕事」を定義する。さらに、各「会議通知」や各「会場準備」は曖昧な周期で発生する同様のタスクであることを表現するために、それぞれリカーレンス「会議通知」、「会場準備」、「会議」、および「議事録送付」を定義する。そして、「第1回会議関連の仕事」と「第2回会議関連の仕事」が曖昧な周期で発生する同様のミッションであることを表現するために、これらのミッションを要素とするジョブ「会議関連の仕事」を定義する。

リカーレンスを構成するタスクは、時系列に並べたときにある程度決まった周期性を持っていると考えられる。この周期性を用いれば、将来発生するタスクを予測することができると考えられる。このため、将来の作業発生を予測するために作業発生の規則性のモデルの内、タスクとリカーレンスを用いる。以降の節では、オフィス環境におけるリカーレンスの例について述べる。

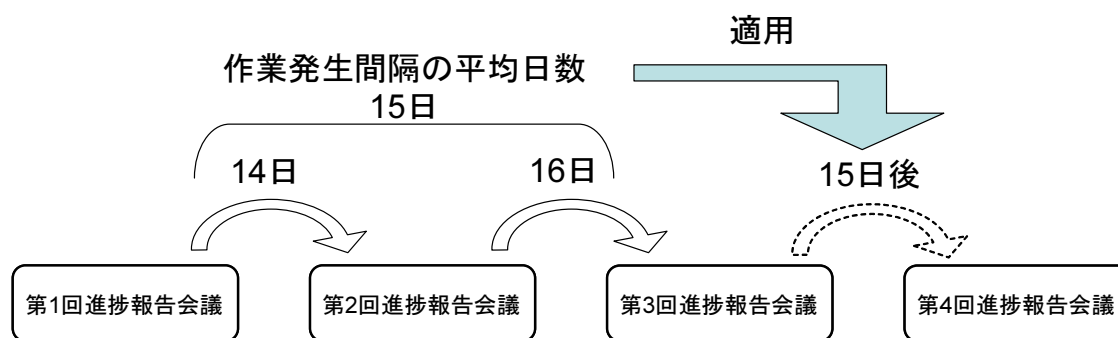


図 2.2 周期性の継承による作業予測

2.2 オフィス環境におけるリカーレンスの例

オフィス環境におけるリカーレンスの例として進捗報告会議を考える。進捗報告会議は、仕事や研究の進捗状況を報告するためのタスクである。発生する日付について、以下の3つの特徴を持つ。

- (1) 約2週間に1回発生する
- (2) 上司の予定や仕事の進捗具合によって発生する日付が前後する
- (3) 発生する曜日に偏りがある

これらの事実から、過去の発生履歴(リカーレンス)を利用して次のタスク発生を推測する場合、リカーレンス内のタスク発生日付の間隔の平均を求め、平均をそのまま次のタスク発生の推測に利用する手法が考えられる。この様子を図2.2に示す。しかしながら、この手法では、進捗報告会議の発生をうまく推測することができない。なぜなら、勤務体系や習慣から、以下のことが仮定されるからである。

- (1) 土曜日、日曜日、および祝日には発生しない
- (2) 長期休暇や年末年始には発生しない

図2.2に従った単純な推測では、15日後と推測した「第4回進捗報告会議」が夏季休暇といった長期休暇に容易にぶつかってしまう。さらに、長期休暇をはさんだ場合の作業発生間隔も周期として数えられるので、平均値が不当に大きくなってしまう。したがって、過去の発生日付の間隔を考慮するだけでは、不十分であることが分かる。

オフィス環境において上記のような特徴を持つ作業は多いと考えられる。このため、ここで取り上げた進捗報告会議の発生を予測する手法について検討する。

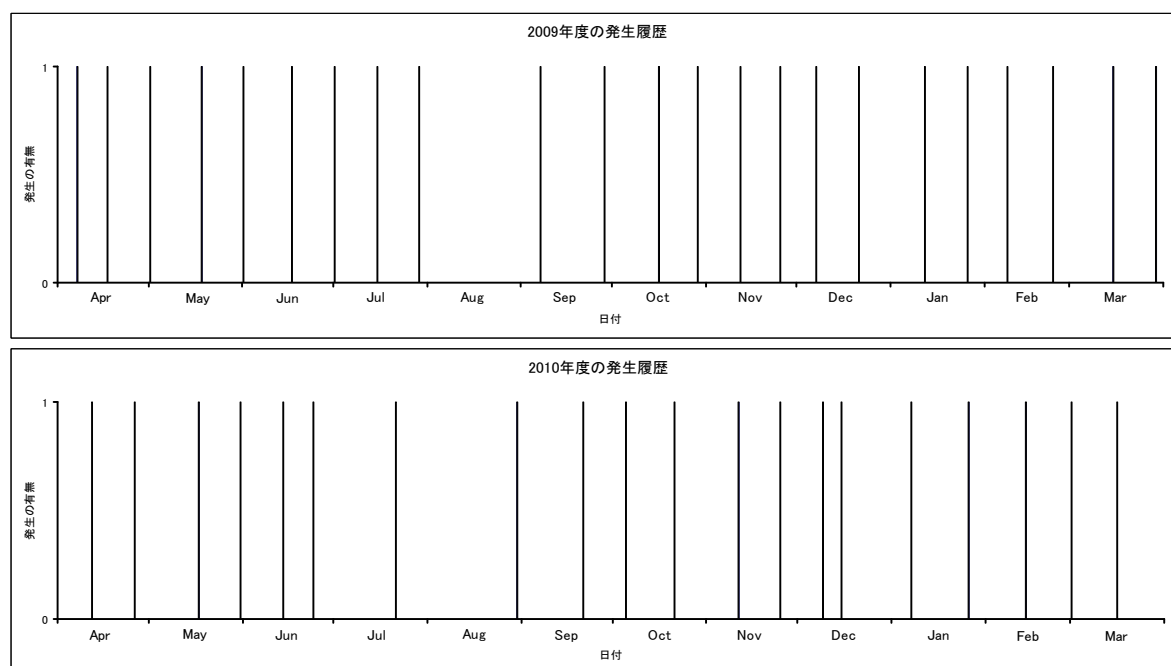


図 2.3 2009 年度，2010 年度の打合せの発生履歴

表 2.1 曜日ごとの発生頻度

曜日	日	月	火	水	木	金	土	祝日	合計
発生回数 (回)	0	16	7	9	7	4	0	0	43
発生割合 (%)	0	37	16	21	16	10	0	0	100

2.3 事例分析

図 2.3 は，オフィス環境における実際の作業発生の履歴を図示したものである．具体的には，著者が所属する研究グループでの打合せについて，2009, 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 で表現しており，前節で示した進捗報告会議の具体的な事例となっている．

図 2.3 のデータを分析した結果，以下の 4 つのことが分かった．

(1) 発生する曜日に偏りがある

作業が発生する曜日には，偏りがある．これは，他のタスクの曜日が固定されていたり，日付決定の際に 2 週間という周期を意識した意思が働いているからだと考えられる．曜日ごとの発生頻度を表 2.1 に示す．表中では，祝日も曜日の 1 つとしている．表

表 2.2 開始時刻の発生頻度

通番	開始時刻	発生回数(回)	累積割合(%)
1	13:00	12	28
2	10:00	9	49
3	09:30	3	56
4	15:00	3	63
5	09:00	2	67
6	13:10	2	72
7	14:00	2	77
8	10:25	1	79
9	13:30	1	81
10	14:15	1	84
11	14:30	1	86
12	14:35	1	88
13	14:20	1	91
14	15:15	1	93
15	16:10	1	95
16	16:15	1	98
17	16:20	1	100

表 2.3 作業時間の発生頻度

通番	作業時間(分)	発生回数(回)	累積割合(%)
1	120	12	28
2	160	4	37
3	165	4	47
4	100	3	53
5	180	3	60
6	145	2	65
7	150	2	70
8	170	2	74
9	65	1	77
10	75	1	79
11	95	1	81
12	90	1	84
13	105	1	86
14	110	1	88
15	125	1	91
16	130	1	93
17	135	1	95
18	185	1	98
19	240	1	100

2.1 によると、月曜日が全体の約 40%を占めることが分かる。また、土日や祝日にはまったく発生していないことが分かる。

(2) 作業が発生しない期間が存在する

お盆、夏期休暇、年末年始といった作業が発生しない期間が明確にあらわれている。具体的には、7/30-8/29、また 12/22-1/6 が各年度に共通して作業が発生しない期間と観測される。

(3) 時間帯に偏りがある

時間帯を開始時刻と作業時間の組み合わせであると考えた場合、作業が発生する時間帯はある程度決まっている。開始時刻、作業時間の発生頻度を表 2.2、表 2.3 に示す。これによると、開始時刻は、発生回数の多い上位 3 つで全体の 50%以上を占めていることが分かる。また、作業時間は、発生回数の最も多い 120 分の発生回数が 2 番目に発生回数の多い 160 分の発生回数の 3 倍であることが分かる。

(4) 年間における差異は大きくない

年単位で発生を系列を比較してみると、1 年を周期とした相似形を観測できる。つまり、作業自身は、2 週間周期で発生しているタスクと考えているが、1 年を周期として見直すと、2009 年度の第 1 回会議と 2010 年度の第 1 回会議が 1 つのリカーレンスを

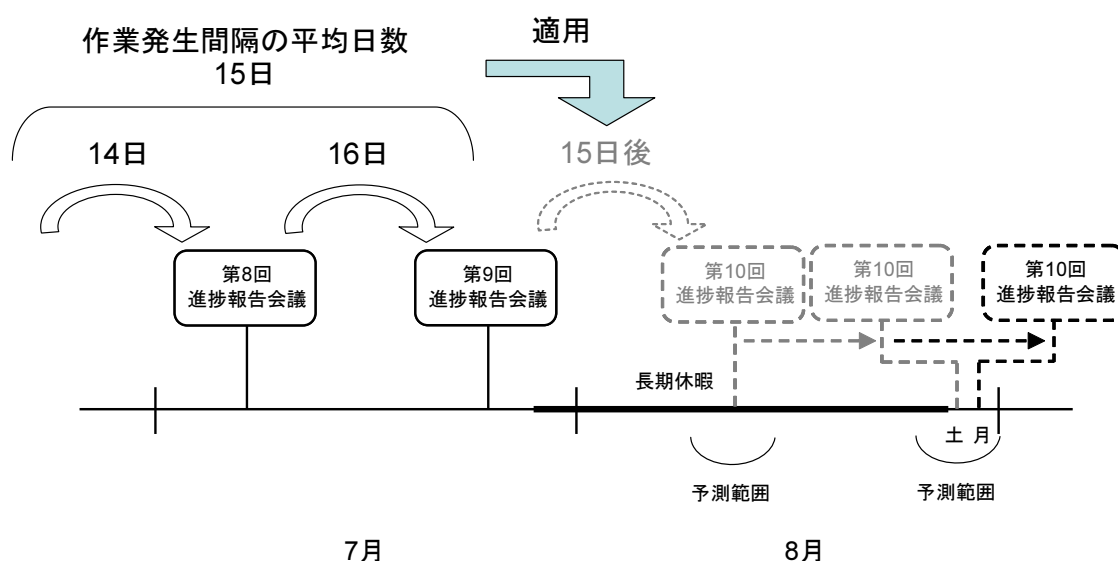


図 2.4 周期性の継承と発生確率を考慮した作業予測

形成していると考えられる。この観点から、1年という周期で各回の会議が発生していると考え、長期休暇にも同様の周期を見い出すことができる。

2.4 作業予測の手法

2.4.1 手法の提案

作業予測の手法として、以下の2つが考えられる。

(提案手法 1) 周期性の継承と発生確率を考慮した作業予測

この手法は、個々の作業発生間隔は、次の作業の発生にも適用されるという考えに基づく手法に修正を加えたものである。前節の事例でいうと曜日ごとの発生割合、作業が発生しない期間、および時間帯の情報を用いて、2週間に1回の発生間隔に修正を加える手法である。2.2節で述べた進捗報告会議の例に、この手法を当てはめると図 2.4 のようになる。図 2.4 は、「第 10 回進捗報告会議」の予測の様子を示したものである。この手法では、最後に発生した作業の日付に作業発生間隔の平均日数を加えた日付を基準日とし、基準日と作業発生間隔のばらつきから予測範囲を求める。予測範囲内の日付で基準日に近く、曜日の発生割合の高い日付を発生日付として予測する。図 2.4 では、作業発生間隔の平均日数は 15 日であるため、「第 9 回進捗報告会議」の発生日付か

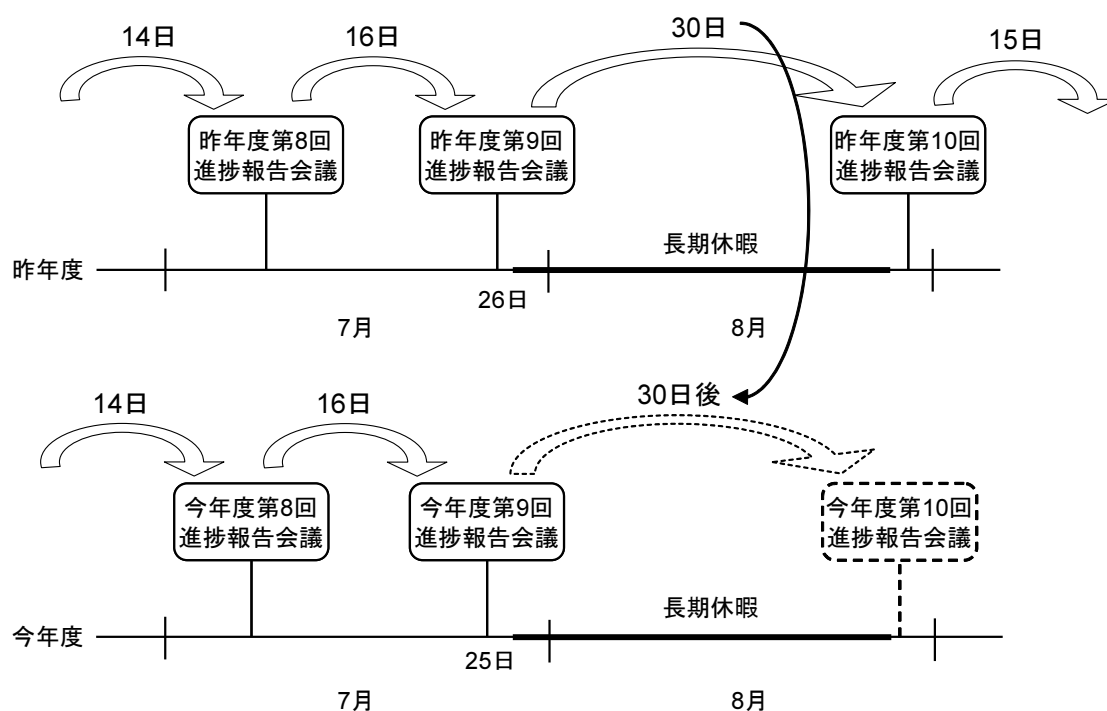


図 2.5 年を単位とした周期性の継承による作業予測

ら 15 日後を基準日とし，基準日と作業発生間隔のばらつきから予測範囲を求めている．しかし，予測範囲が作業が発生しない期間内であるため，基準日を作業が発生しない期間の最終日の翌日までずらし，再度予測範囲を求める．基準日は，土曜日であり作業が発生しない曜日であるため，予測範囲内で発生割合の高い月曜日に修正し，その日付を「第 10 回進捗報告会議」の発生日付と予測する．また，上記の方法により求めた発生日付に時間帯の情報を加えて，発生日時を予測する．

(提案手法 2) 年を単位とした周期性の継承による作業予測

前節で示したように，リカーレンスを持つタスクは，その発生日付において年を単位とする相似形を持っていることが分かる．たとえば，年度初回に発生する作業の時期は，毎年 4 月の 10 日前後で変わらない．また，2 週間程度とした作業発生間隔は繁忙期には短くなるという傾向がある．この原因が，特別研究報告書締切などのイベントに影響を受けていることを考えると，年を単位とした周期性の継承の妥当性が理解できる．2.2 節で述べた進捗報告会議の例に，この手法を当てはめると図 2.5 のようになる．図 2.5 は，「今年度第 10 回進捗報告会議」の予測の様子を示している．この手法では，昨年の同様の時期に発生した作業の作業発生間隔を用いて予測する．図 2.5 で

は、「今年度第 10 回進捗報告会議」を予測するために、前回の「今年度第 9 回進捗報告会議」の発生日付である 7 月 25 日を基準にして、昨年との対応点を考える。具体的には、昨年の 7 月 25 日から最も近い会議とその発生日付を求め、「昨年度第 9 回進捗報告会議」(7 月 26 日)を得る。「昨年度第 9 回進捗報告会議」と「昨年度第 10 回進捗報告会議」の間隔 30 日が今年度にも適用できると考えて、「今年度第 9 回進捗報告会議」から 30 日後を「今年度第 10 回進捗報告会議」の発生日付と予測する。

これらの手法を組み合わせることで、作業予測の精度を向上できると考えられる。以降では、周期性の継承と発生確率を考慮した作業予測の手法について、具体的な方針と作業履歴から抽出すべき情報について検討する。

2.4.2 周期性の継承と発生確率を考慮した作業予測

周期性の継承と発生確率を考慮した作業予測の方針についてさらに検討する。2.3 節より、作業発生間隔の平均日数以外に、以下の 3 つを考慮することが考えられる。

(1) 曜日を考慮する

2.3 節の結果から作業が発生する曜日には偏りがあることが分かる。具体的には、発生する割合の高い曜日や全く発生しない曜日がある。曜日ごとの発生割合を求め、予測に用いることにより、作業が発生する割合の高い曜日や作業が発生しない曜日を考慮して予測することができる。

(2) 作業が発生しない期間を考慮する

2.3 節の結果から長期休暇や年末年始といった作業が発生しない期間があることが分かる。作業の発生を 1 年単位で見た場合、作業が発生しない期間は、毎年同様であると考えられる。1 年の内、作業が発生しない期間を考慮することにより、作業が発生しない期間に予測されることを防ぐことができる。

(3) 時間帯を考慮する

時間帯は、開始時刻と作業時間の組み合わせであると考えられる。この場合、2.3 節の結果から開始時刻には、発生回数の多いものが複数存在することが分かる。また、発生回数の違いから発生することが多い開始時刻は、ある程度決まっていると考えられる。このため、発生回数の多い開始時刻の上位 3 つを予測に用いる。作業時間に関しては、発生回数の多い作業時間がほぼ決まっており、その作業時間の近辺でばらつきがあるこ

とが分かる。また、作業時間は、実際に作業をするまで分からないことが考えられる。このため、発生回数の最も多い作業時間を予測に用いる。

以上のことを考慮することにより、予測の精度を上げることができると考えられる。

周期性の継承と発生確率を考慮した作業予測では、発生日付と時間帯を別々に求める。まず、最後に発生した作業の日付、作業発生間隔、曜日ごとの発生割合、および作業が発生しない期間の情報から作業の発生日付を予測する。次に、時間帯を求める。求めた発生日付と時間帯を組み合わせて発生日時を予測する。発生日付と時間帯は、それぞれ発生する確率が高いものを求める。発生する確率が高い日付と時間帯は、それぞれ以下のものであると考えた。

(1) 発生する確率の高い日付

最後に発生した作業の日付に作業発生間隔の平均日数を加えた日付を基準日とする。また、作業発生間隔のばらつきから予測範囲を求める。予測範囲内の日付で、基準日に近く、曜日の発生割合の高い日付を発生する確率の高い日付とする。ただし、作業が発生しない期間を除く。

(2) 発生する確率の高い時間帯

発生回数の多い時間帯の上位 3 つと最も発生回数の多い作業時間を組み合わせた 3 つの時間帯のパターンとする。

2.4.3 繰返情報

周期性の継承と発生確率を考慮した作業予測に必要な情報として、繰返情報を定義する。繰返情報は、以下の 5 つからなる。表 2.4 は繰返情報についてまとめたものである。

(1) 発生間隔情報

発生間隔情報は、作業発生間隔の平均日数と標準偏差である。

作業発生間隔のばらつきを表わす情報として標準偏差を用いる。

(2) 曜日情報

曜日情報は、日月火水木金土の 7 つの曜日に祝日を加えた 8 つの発生割合である。

祝日を 1 つの曜日として考え、祝日に発生した場合は、曜日に関係なく祝日に発生したとして扱う。

(3) 例外情報

例外情報は、1 年の内の長期休暇や年末年始といった作業が発生しない期間である。

表 2.4 繰返情報

繰返情報	詳細
発生間隔情報	作業発生間隔の平均日数と標準偏差
曜日情報	日月火水木金土の 7 つの曜日に祝日を加えた 8 つの発生割合
例外情報	1 年の内の長期休暇や年末年始といった作業が発生しない期間
時間帯情報	発生回数の多い開始時刻の上位 3 つと最も発生回数の多い作業時間を組み合わせた 3 つの時間帯のパターン
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

(4) 時間帯情報

時間帯は、開始時刻と作業時間の組み合わせで表現する。

時間帯情報は、発生回数の多い開始時刻の上位 3 つと最も発生回数の多い作業時間を組み合わせた 3 つの時間帯のパターンである。

(5) 最終作業情報

最終作業情報は、リカーレンス内の直近の作業に関する情報であり、以下の 4 つである。

(A) 作業名

(B) 開始時刻

(C) 終了時刻

(D) リカーレンスの系列を示す識別子

2.4.4 予測アルゴリズム

前節で定義した繰返情報と 2.4.1 項で述べた周期性の継承と発生確率を考慮した作業予測に基づいて、予測アルゴリズムを考察する。このアルゴリズムは、最後に発生した作業から次の作業の発生を予測するものである。まず、日付を求め、求めた日付に時間帯を付加して発生日時を予測する。ここでは、予測した日付を予測日、予測した日時を予測日時とする。最終作業情報と予測日時により次の作業の発生を予測する。予測日と予測日時の求め方を以下に示す。

(1) 予測日

繰返情報の発生間隔情報を用いて基準日と予測範囲を求め、基準日からの日数差、曜日情報、および例外情報を用いて予測範囲内の日付に重み付けする。重み付けの値が最も高い日付を予測日とする。

(2) 予測日時

予測日に繰返情報の時間帯情報を付加したものを予測日時とする。

上記の方法で、予測を行う予測アルゴリズムを図 2.6 に示す。以下、アルゴリズムの説明において用いる語句の説明である。

(1) 作業発生間隔の平均日数 ($r_{average}$)

繰返情報の 1 つである発生間隔情報の作業発生間隔の平均日数である。基準日を求めるために用いる。

(2) 作業発生間隔の標準偏差 ($r_{standard_deviation}$)

繰返情報の 1 つである発生間隔情報の作業発生間隔の標準偏差である。予測範囲内の日付 (候補日) を求めるために用いる。

(3) 時間帯情報 (r_{time_slot})

繰返情報の 1 つである。予測日とこの情報を用いて予測日時を求める。

(4) 最終作業情報 ($r_{latest_occurrence}$)

繰返情報の 1 つである。予測日時とこの情報を用いて予測結果を求める。

(5) 最終作業日 (d_{last})

最後に発生した作業の日付であり、繰返情報の最終作業情報の開始時刻から求められる。

(6) 基準日 (d_{pivot})

予測範囲を求める際の基準となる日付である。初期値は、最終作業日に発生間隔情報の平均日数を加えた日付である。

(7) 候補日 ($d_{candidate}$)

予測範囲内の日付であり、予測日の候補となる日付である。初期値は、基準日と発生間隔情報の標準偏差を用いて求められる。基準日も候補日の 1 つである。

(8) 予測日 (d_{max})

候補日リストの中で最も高い評価値を持つ候補日である。これを予測日とする。

(9) 候補日リスト ($D_{candidate}$)

候補日のリストである。

```

1   $d_{pivot} := d_{last} + r_{average}$ 
2  loop:
3     $D_{candidate} := (d_{pivot} - r_{standard\_deviation}, \dots, d_{pivot} + r_{standard\_deviation})$ 
4    for each  $d$  in  $D_{candidate}$ 
5      begin
6         $E(d) :=$  if  $d \leq d_{last}$  then 0
7                  else if  $d$  が例外情報の期間内である then 0
8                  else  $(100 - ((100 / (r_{standard\_deviation} + 1)) * |d_{pivot} - d|)) * d$  の曜日の発生割合
9      end
10   if  $Max(E) = 0$  then
11     begin
12        $d_{pivot} := d_{pivot} + r_{standard\_deviation} + 1$ 
13        $D_{candidate}$  を空にする
14     goto loop
15   end
16    $d_{max} :=$   $Max(E)$  を持つ  $d_{candidate}$ 
17   return [ $d_{max}$ ,  $r_{time\_slot}$ ,  $r_{latest\_occurrence}$ ]

```

図 2.6 予測アルゴリズム

(10) 評価値 (E)

候補日に付与される重み付けの値である。この値が高いほど、発生する確率が高い日付である。

第 3 章

作業予測システムの設計

3.1 概要

繰返情報を用いて作業発生を予測する作業予測システムを設計する。作業予測システムは、将来発生する作業を予測し、予測結果をユーザに提示する。ユーザは、UI(ユーザインタフェース)部から予測結果の閲覧と修正を行うことができる。作業予測システムの全体像を図 3.1 に示す。以下、作業予測システムの各部の構成と動作内容について述べる。

作業予測システムの各部の構成を以下に示す。

(A) 作業履歴 DB

作業データを保存する DB である。作業データは、作業発生の規則性を扱うモデルに沿ったデータ形式で保存されている。

(B) 繰返情報抽出部

繰返情報抽出部は、作業履歴 DB から作業データを受け取り、作業データを解析して 2.4.3 項で示した繰返情報を抽出する。

(C) 予測部

予測部は、繰返情報抽出部から繰返情報を受け取り、2.4.4 項で示したアルゴリズムに従って予測を行う。

(D) UI(ユーザインタフェース)部

UI 部は、予測結果をユーザに提示する。ユーザは、UI 部から予測結果の修正を行うことができる。

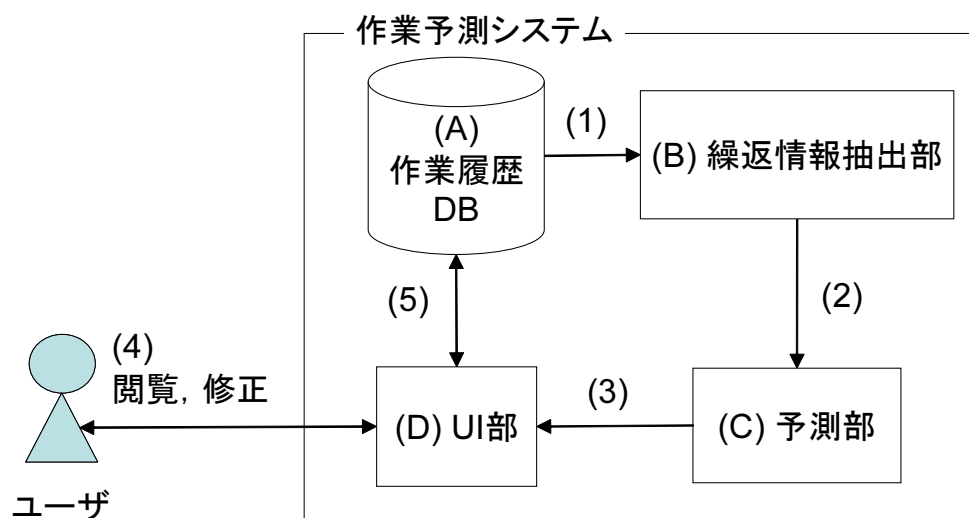


図 3.1 作業予測システムの全体像

次にシステムの動作内容を説明する。図 3.1 の矢印に付記している番号は、以下で述べる動作内容の説明と対応している。

- (1) 繰返情報抽出部 (B) は、作業履歴 DB(A) に保存されている作業データを受け取り、繰返情報を抽出する。
- (2) 予測部 (C) は、抽出された繰返情報を受け取り、将来発生する作業を予測する。
- (3) UI 部 (D) は、予測された結果を受け取り、それをユーザに提示する。
- (4) ユーザは、予測された結果を閲覧し、適切なものに修正する。
- (5) 作業履歴 DB(A) は、修正された作業データを受け取り、作業履歴として保存する。

3.2 作業履歴 DB

3.2.1 概要

作業履歴 DB は、作業データを保存する DB である。作業データは、作業発生の規則性を扱うモデルに沿ったデータ形式で保存されている。具体的には、リカーレンスやタスクに対応した作業の名前や作業発生の時刻を保存している。これらの情報は、典型的には、カレンダーシステムのデータベースから抽出可能である。

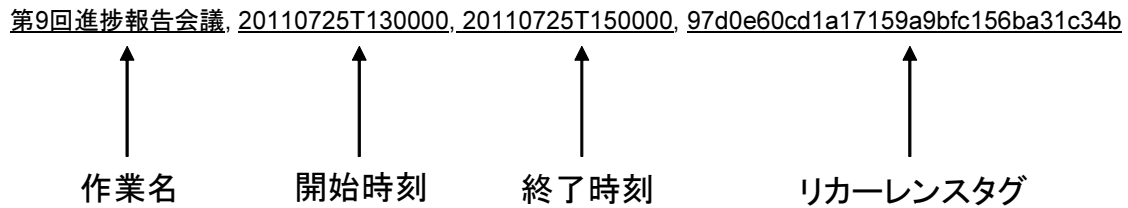


図 3.2 作業発生の規則性のデータ形式

3.2.2 データ形式

データ形式は、作業発生の規則性の要素であるタスクとそれが属するリカーレンスを表現している。図 3.2 に作業データの例を示す。作業データは情報として、以下の 4 つを持つ。

(1) 作業名

ユーザが識別する際に利用する任意長の文字列である。カレンダーシステムにおけるサマリ情報と同等である。

(2) 開始時刻

作業の開始時刻を文字列で表現する。表現形式は ISO8601[6] に準じた LocalTime とする。

(3) 終了時刻

作業の開始時刻と同様に終了時刻を表現する。

(4) リカーレンスタグ

任意長の文字列であり、作業発生の規則性におけるリカーレンスを表現している。つまり、当該タスクがリカーレンスに属している場合のリカーレンスを識別する ID である。

3.3 繰返情報抽出部

3.3.1 概要

繰返情報抽出部は、作業履歴 DB の出力である作業データを入力とする。また、繰返情報抽出部は、作業データを解析し、抽出した繰返情報を出力とする。

3.3.2 繰返情報の表現形式

2.4.3 項で述べた繰返情報をシステムで扱うために，繰返情報の表現形式を決める．以下，繰返情報の名前，用途，型，詳細，および例について述べる．

(1) 発生間隔情報

(A) 名前

cycle

(B) 用途

基準日と候補日を求めるために用いる．

(C) 型

型：発生間隔情報の配列

発生間隔情報の配列は，以下に示す表記で定義する．

cycle	= [cycle_average, cycle_standard_deviation]
cycle_average	= 1 以上の整数
cycle_standard_deviation	= 0 以上の整数

(D) 詳細

発生間隔情報の配列は，cycle_average と cycle_standard_deviation の 2 つの要素を持つ配列である．cycle_average は，作業発生間隔の平均日数を表す情報であり，値は 1 以上の整数とする．cycle_standard_deviation は，作業発生間隔の標準偏差を表す情報であり，値は 0 以上の整数とする．

(E) 例

作業発生間隔の平均日数が 15 日で標準偏差が 3 の場合の例を以下に示す．

例：[15,3]

(2) 曜日情報

(A) 名前

day_of_the_week_ratio

(B) 用途

曜日ごとの発生確率の違いを考慮するために用いる．

(C) 型

型：曜日情報の配列

曜日情報の配列は，以下に示す表記で定義する．

```

day_of_the_week_ratio = [sun_ratio,mon_ratio,tue_ratio,wed_ratio,
                        thu_ratio,fri_ratio,sat_ratio,hol_ratio]

sun_ratio              = ratio
mon_ratio              = ratio
tue_ratio              = ratio
wed_ratio              = ratio
thu_ratio              = ratio
fri_ratio              = ratio
sat_ratio              = ratio
hol_ratio              = ratio
ratio                  = 0 から 1 までの実数

```

(D) 詳細

曜日情報の配列は，8つの要素を持つ配列である．この8つの要素とは，日曜日から土曜日までの通常の7つの曜日と祝日のそれぞれの発生割合である．祝日とは，国民の祝日，振替休日，および国民の休日を合わせたものである．祝日に発生した場合は，曜日に関係なく祝日に発生したとして扱う．発生割合は，0以上1以下の実数であり，すべての曜日の発生割合の合計値は1である．

(E) 例

曜日の発生割合がそれぞれ，月曜日が0.37，火曜日が0.16，水曜日が0.21，木曜日が0.16，金曜日が0.1，および土日と祝日が0の場合の例を以下に示す．

例：[0,0.37,0.16,0.21,0.16,0.1,0,0]

(3) 例外情報

(A) 名前

exceptive_period

(B) 用途

長期休暇や年末年始といった作業が発生しない期間を考慮するために用いる．

(C) 型

型：例外情報の配列

例外情報の配列は，以下に示す表記で定義する．

```
exceptive_period = ["start_date/end_date"]
start_date       = full_date
end_date         = full_date
full_date        = date_fullyear date_month date_mday
date_fullyear    = 4 桁の整数
date_month       = 01 から 12 までの 2 桁の整数
date_mday        = 01 から 31 までの 2 桁の整数
(date_month の値により，01 から 28, 01 から 29,
または 01 から 30 までに変換することがある)
```

(D) 詳細

例外情報の配列は，例外期間を格納した配列である．例外期間は，開始日と終了日を/(スラッシュ)で結んだ文字列で表す．例外期間が複数ある場合は，,(コンマ)を区切りとして記述する．開始日と終了日は，年を表す 4 桁の整数，月を表す 01 から 12 までの 2 桁の整数，および日を表す 01 から 31 までの 2 桁の整数を繋げたものである．

(E) 例

開始日が 2011 年 7 月 30 日で終了日が 2011 年 8 月 29 日である場合の例を以下に示す．

例：["20110730/20110829"]

(4) 時間帯情報

(A) 名前

time_slot

(B) 用途

作業の時間帯を求めるために用いる．

(C) 型

型：時間帯情報の配列

時間帯情報の配列は，以下に示す表記で定義する．

```
time_slot          = [time_slot_pattern,time_slot_pattern,time_slot_pattern]
time_slot_pattern  = ["start_time",working_seconds]
start_time         = time
time               = time_hour time_minute time_second
time_hour          = 01 から 23 までの 2 桁の整数
time_minute        = 01 から 59 までの 2 桁の整数
time_second        = 01 から 60 までの 2 桁の整数
(60 の場合は繰上げされる)
working_seconds    = 整数
```

(D) 詳細

時間帯情報の配列は，時間帯のパターンを要素として持つ配列である．時間帯のパターンは，それぞれ第 1 要素に開始時刻，第 2 要素に作業時間を持つ配列である．開始時刻は，時を表す 00 から 23 までの 2 桁の整数，分を表す 00 から 59 までの 2 桁の整数，および秒を表す 00 から 60 までの 2 桁の整数をつなげたものである．作業時間は，整数であり，単位は秒である．

(E) 例

開始時刻が 9 時 30 分，10 時，および 13 時であり，作業時間が 7200 秒 (2 時間) の場合の例を以下に示す．

例：`[["093000", 7200],["100000",7200],["130000",7200]]`

(5) 最終作業情報

(A) 名前

`latest_occurrence`

(B) 用途

リカーレンス内の直近の作業の情報を扱う．

(C) 型

型：最終作業情報のハッシュ

最終作業情報のハッシュは，以下に示す表記で定義する．

```
latest_occurrence = {"name" => work_name,  
                     "start_date_time" => start_date_time,  
                     "end_date_time" => end_date_time,  
                     "recurrence_tag" => recurrence_tag}
```

work_name = 作業名を表す文字列

start_date_time = date "T" time

end_date_time = date "T" time

recurrence_tag = リカーレンスを表す文字列 (32 桁の英数字)

date = date_value

date_value = date_fullyear date_month date_mday

date_fullyear = 4 桁の整数

date_month = 01 から 12 までの 2 桁の整数

date_mday = 01 から 31 までの 2 桁の整数
(date_month の値により, 01 から 28, 01 から 29,
または 01 から 30 までに変わることがある)

time = time_hour time_minute time_second

time_hour = 00 から 23 までの 2 桁の整数

time_minute = 00 から 59 までの 2 桁の整数

time_second = 00 から 60 までの 2 桁の整数
(60 の場合は繰上げされる)

(D) 詳細

最終作業情報のハッシュは、キーが作業名、開始時刻、終了時刻、およびリカーレンスタグであり、バリューがそれぞれの値である。キーとして以下の 4 つの文字列がある。

- (a) "name"(作業名)
- (b) "start_date_time"(開始時刻)
- (c) "end_date_time"(終了時刻)
- (d) "recurrence_tag"(リカーレンスタグ)

それぞれのキーにおいて、バリューの形式が異なる。以下に、それぞれのキーに対応するバリューの値を示す。

(a) "name"の場合

作業名を表す文字列である。(例："第9回進捗報告会議")

(b) "start_date_time"の場合

作業の開始時刻を表す値である。開始時刻は、年月日と時刻を T で結んだものである。年月日は、西暦を表す 4 桁の整数と月を表す 01 から 12 までの 2 桁の整数と日を表す 01 から 31 までの 2 桁の整数を繋げたものである。時刻は、時を表す 00 から 24 までの 2 桁の整数と分を表す 00 から 59 までの 2 桁の整数を繋げたものである。(例："20110725T130000")

(c) "end_date_time"の場合

作業の終了時刻を表す値である。開始時刻と同様の形式である。
(例："20110725T150000")

(d) "recurrence_tag"の場合

リカーレンスを表現する文字列であり、リカーレンスの ID である。
(例："97d0e60cd1a17159a9bfc156ba31c34b")

(E) 例

最後に発生した作業の情報の例を以下に示す。

```
例：{ "name"           => "第9回進捗報告会議",  
      "start_date_time" => "20110725T130000",  
      "end_date_time"   => "20110725T150000",  
      "recurrence_tag"  => "97d0e60cd1a17159a9bfc156ba31c34b" }
```

3.3.3 繰返情報の抽出方法

作業履歴 DB から受け取った作業データを解析して、リカーレンスごとに繰返情報を抽出する。以下、繰返情報の抽出方法について述べる。

(1) 発生間隔情報の抽出方法

発生間隔情報は、作業発生間隔の平均日数と標準偏差の情報である。作業データの開始時刻から作業の日付を求め、作業日とする。まず、個々の作業発生間隔の求め方を以下に示す。作業データは n 個とする。

$$y_i = x_{i+1} - x_i \quad (i = 1 \dots n - 1)$$

x_i : 個々の作業日

y_i : 個々の作業発生間隔

以下、作業発生間隔を用いて、作業発生間隔の平均日数と標準偏差を抽出する方法について述べる。

(A) 作業発生間隔の平均日数の抽出方法

作業発生間隔の平均日数の抽出方法を以下に示す。作業データは n 個とする。

$$\bar{y} = \frac{1}{n-1} \sum_{i=1}^{n-1} y_i$$

(\bar{y} は、小数点以下を切り捨てる)

\bar{y} : 作業発生間隔の平均日数

(B) 作業発生間隔の標準偏差の抽出方法

作業発生間隔の標準偏差の抽出方法を以下に示す。作業データは n 個とする。

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} (y_i - \bar{y})^2}$$

(s は、小数点以下を切り上げる)

s : 作業発生間隔の標準偏差

上記の方法で求めた作業発生間隔の平均日数と標準偏差を配列に格納し、発生間隔情報とする。

(2) 曜日情報の抽出方法

曜日情報は、日曜日から土曜日までの通常の7つの曜日と祝日の発生割合を配列に格納したものである。作業データの開始時刻から曜日の情報を抽出する。作業データが n 個の場合の各曜日の発生割合の求め方は、以下のようになる。

$$\text{曜日の発生割合} = \frac{\text{曜日の発生回数}}{n}$$

曜日ごとに発生割合を求め、それぞれの曜日の発生割合を日月火水木金土祝日の順で配列に格納し、曜日情報とする。

(3) 例外情報の抽出方法

例外情報とは、長期休暇や年末年始といった作業が発生しない期間である。例外情報は、手動で入力する。

(4) 時間帯情報の抽出方法

時間帯情報は、発生回数の多い開始時刻の上位 3 つと発生回数の最も多い作業時間を組み合わせた 3 つの時間帯のパターンである。作業データの開始時刻、終了時刻から、発生回数の多い開始時刻の上位 3 つと発生回数の最も多い作業時間を抽出する。発生回数が同じ開始時刻、作業時間があつた場合は、それぞれ発生 of 新しい開始時刻、作業時間を抽出する。抽出した開始時刻、時間帯の組をそれぞれ配列に格納し、時間帯情報とする。

(5) 最終作業情報の抽出方法

リカーレンスの中で、直近の作業データから以下の 4 つの情報を抽出する。

- (A) 作業名
- (B) 開始時刻
- (C) 終了時刻
- (D) リカーレンスタグ

これらの情報をハッシュに格納し、最終作業情報とする。

3.4 予測部

予測部は、繰返情報抽出部の出力である繰返情報を入力とし、予測した作業データを出力とする。この際に、2.4.4 項で示したアルゴリズムを用いる。

3.5 UI 部

UI 部は、予測部の出力である予測結果、作業履歴 DB に保存されている作業データ、またはユーザにより修正された作業データを入力とする。また、予測結果、作業履歴 DB に保存されている作業データ、またはユーザにより修正された作業データを出力とする。

第 4 章

おわりに

本論文では、オフィス環境における作業の例として、著者の所属する研究グループで行われた打合せを分析した。この結果、発生する曜日に偏りがあること、作業が発生しない期間があること、時間帯に偏りがあること、および年間における差異は大きくないことが分かった。分析の結果をふまえて、予測の手法として「周期性の継承と発生確率を考慮した作業予測」と「年を単位とした周期性の継承による作業予測」の2つの手法を提案した。この2つの手法の内、周期性の継承と発生確率を考慮した作業予測についてさらに検討し、必要な情報として繰返情報を定義した。また、提案手法を詳細化した予測アルゴリズムについて述べた。さらに、作業予測システムを設計し、システムの動作概要、各部の構成、繰返情報の表現形式、および繰返情報の抽出方法について述べた。

残された課題として、年を単位とした周期性の継承による作業予測の詳細についての検討と予測結果の提示方法や予測の誤りを修正する仕組みについての検討がある。また、作業予測システムの実装と評価がある。

謝辞

本研究を進めるにあたり，懇切丁寧なご指導をしていただきました乃村能成准教授に心より感謝の意を表します。また，研究活動において，数々のご指導やご助言を与えていただいた谷口秀夫教授，山内利宏准教授，後藤佑介助教に心から感謝申し上げます。

また，日頃の研究活動において，お世話になりました研究室の皆様に感謝いたします。

最後に，本研究を行うにあたり，経済的，精神的な支えとなった家族に感謝いたします。

参考文献

- [1] 安部田 章, 松並 勝, 碓崎 賢一, “スケジュール情報の共有・再利用に着目した協調作業支援システム,” 情報処理学会研究報告. [グループウェア], Vol. 95, No. 67, pp. 7-12, 1995.
- [2] 山根 隼人, 長尾 確, “AcTrec : 行動履歴を用いた個人行動支援,” 情報処理学会第 66 回全国大会講演論文集, Vol. 66, No. 3, pp. 115-116, 2004.
- [3] 斉藤 典明, 金井 敦, 赤埴 淳一, “知識蓄積・継承のためのスケジュールデータ構成法,” 情報処理学会研究報告. [グループウェアとネットワークサービス (GN)], Vol. 2012-GN-82, No. 19, pp. 1-7, 2012.
- [4] 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提案,” マルチメディア通信と分散処理ワークショップ論文集, Vol.2010, No.11, pp. 215-220, 2010.
- [5] 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,” 研究報告マルチメディア通信と分散処理 (DPS) , Vol.2011-DPS-149, No.10, pp. 1-6, 2011.
- [6] International Organization for Standardization, 8601, “Data elements and interchange formats – Information interchange – Representation of dates and times,” 2004.