

計画立案の行動に着目したカレンダー情報の ダミーデータ生成手法の提案

吉田 尚史¹ 乃村 能成¹

概要: カレンダー情報を扱う手法の研究やアプリケーションの開発において、カレンダー情報をテストケースとして用意したいという要求がある。しかし、実際に使用されているカレンダー情報の収集はプライバシーの保護や守秘義務の観点から困難である。そこで、本稿では、カレンダー情報のダミーデータを生成する手法を提案する。提案手法では、複数の繰り返し発生する予定の集合と単発で発生する予定を組み合わせ、カレンダー情報のダミーデータを生成する。また、繰り返し発生する予定を生成する場合、計画立案の行動に着目して予定を生成する。カレンダー利用者のモデルを想定し、想定したモデルの特徴を反映したカレンダー情報のダミーデータを生成する。

キーワード: カレンダー情報, ダミーデータ

1. はじめに

カレンダー情報を利用してスケジュール管理を円滑にしたリ、次の行動や予定の発生を予測したりする研究がある。たとえば、複数人の予定を考慮し、新たな予定の登録可能な時間を可視化する手法 [1] や GPS データと組み合わせた行動予測手法 [2] が存在する。また、独自の機能をもったカレンダーシステム [3] を開発する場合もある。これらの手法やシステムを評価したりテストしたりする際に問題となるのは、テストケースとしてもっともらしいカレンダー情報を作成することである。たとえば、文献 [1] の手法の評価では、複数人の予定を考慮した空き時間探索において、各人の予定の密度によって探索の精度が異なると考えられる。そこで、この精度を測るベンチマークとして、各人の予定の密度が異なる複数のカレンダー情報をデータセットとして用意しなければならない。この際、単にランダムに予定を配置したカレンダー情報を作成したのでは、実際の利用状況に則した適切な評価を期待できないという問題がある。そこで、本稿では、評価やテストの意図に合わせて調整可能なダミーのカレンダー情報を作成する手法について提案する。

カレンダー情報に限らず、ソフトウェアテストを目的としてダミーデータを用意することがある。ダミーデータとは、実データの統計情報、あるいはランダムな値を利用し

て作成される人工的なデータである。ダミーデータの例として、名前、住所、生年月日、性別、および電話番号といった個人情報をランダムまたは統計情報を基に生成するサービス [4] やクレジットカードの妥当な番号をランダムに生成するサービス [5] が存在する。また、教育の分野において、学生の成績や行動のダミーデータを生成し、教育支援システムのテストケースを用意する手法 [6] が存在する。しかし、これらのデータ生成手法は、まったくのランダムか、実際の利用者からの実データに基づいているため、カレンダー情報のダミーデータ生成に同様の手法を適用することは難しい。なぜなら、まったくのランダムなデータでは、前述の通り実際の利用に則した評価を期待できないといえ、また、実データを利用する場合は、プライバシーや守秘義務に関わる情報を多数含むため、収集できる情報の量が限られる。その結果、少量の実データしか取得できず、データに偏りが生じやすいため、統計手法に基づいてダミーデータを作成することも難しい。

これに対処するためには、実データに頼らず、ある傾向を想定したダミーデータの自動生成が必要となる。カレンダー情報のダミーデータは、実データから個人情報や守秘義務に関わる情報を隠蔽し、かつ、実データのもつ特徴を失わないように生成する必要がある。そこで、我々は人が予定を立てる際の行動に着目した。本稿では、まず、カレンダー情報のダミーデータ、求められる要求、およびダミーデータ生成時の問題を述べる。次に、カレンダー情報の特徴を表す要素を抽出し、一般化することで、カレンダー情報を

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

パラメータかすることを試みる。その後、パラメータを利用したダミーデータの生成手法について述べる。最後に、提案手法を実装し、パラメータが有効に機能しているかの確認を行う。

2. カレンダー情報のダミーデータ

2.1 ダミーデータに求められる要求

1章にて、まったくのランダムなデータでは、実際の利用に則した評価やテストが難しいと述べた。また、評価やテストの内容によってダミーデータに求められる内容が異なるため、評価やテストの意図に合わせて調整可能なダミーデータが必要であると述べた。評価やテストの意図をうまく反映させるためには、ダミーデータに以下の要素を反映させるべきであると考えられる。

(1) カレンダー利用者のモデルを設定できること

カレンダー利用者のモデルとは、会社員や学生といった職業、予定の数、および予定の種類といったカレンダー利用者の特徴を表現するものである。評価やテストによっては、特定のユーザを想定したカレンダー情報が必要である。たとえば、予定が多く多忙か、「会議」といった定例の予定が多いか、あるいは「出張」といった不在時間が長いかなどを反映したダミーデータが必要である。

(2) カレンダーの予定登録に関する時系列を再現できること

実際のカレンダーでは、予定が登録された日と実施される日が異なる。その差異は、予定によって大きく異なる。たとえば、多数が集まる会議は、実施より数ヶ月前から日付が確定するのに対して、少人数の会議では、直前という場合もある。講義の予定のように、年度の初めにほぼ一括して決定される予定もある。この差異を考慮するかしないかで、「ある日付時点で見るカレンダー」は、異なるものとなる。したがって、ダミーデータは、上記のような計画立案の手順を再現できなければならない。

2.2 ダミーデータ生成時の問題

カレンダー情報のダミーデータを作成する場合、ダミーデータ利用者自身がデータを作成してしまうと、作成者の主観による偏りが生じるという問題が考えられる。つまり、カレンダー情報は職業や組織内の立場によって特徴が異なるため、自身のカレンダー情報と異なる特徴をもつデータを作成することは困難と考える。これには、2つ問題があると我々は考える。1つは、他の職業やユーザがどのようなカレンダー情報を持っているのかを知ったり想像したりすることが困難なことであり、もう1つは、そもそもカレンダーが持つ特徴を表現する要素とは何か明らかになっておらず、適切な記法もないことである。そのため、これまで、「多数のユーザ」から「実データ」を提供してもらう必

要があった。

そこで、後者の問題に対処するために、カレンダー情報の特徴を表す要素とは何かをまず定義して、パラメータとして表現することを考える。そのパラメータを用いることで、意図した特徴を持つカレンダー情報を表現可能であり、表現手法が得られることで、客観的で流通可能な情報としてカレンダー情報のダミーデータを扱えると考えられる。また、カレンダー情報のパラメータのみを協力者から受け取ったり、それらをライブラリ化したりできる。あるいは、実在しないユーザを想定したり、あるユーザがもう少し忙しくなった場合を想定したシミュレーションを行ったりも可能であると考えられる。

そこで、次章ではカレンダー情報の特徴を表す要素を抽出することを試みる。

3. カレンダー情報の特徴を表す要素

3.1 調査方針

カレンダー情報の特徴を表す要素とは何かを考えるために、実在のカレンダーを調査する。

カレンダーに登録される予定は、繰り返し発生する予定と単発で発生する予定の2種類に分けられる。以降では、繰り返し発生する予定の集合をリカーレンスと呼ぶ。たとえば、約1カ月に1回の周期で会議を行う場合、「第1回会議」、「第2回会議」、「第3回会議」という一連の予定の集合を「会議」というリカーレンスとする。本稿では、カレンダー情報の特徴を表す要素として、リカーレンスの特徴を表す要素に着目する。多くの予定は、何らかの周期(再び同様の予定がいつかは発生する)がある[7]ため、リカーレンスの特徴を表す要素を抽出することで、それらの集合としてカレンダー情報の特徴の大部分を定義できると考えるからである。

リカーレンスの特徴を表す要素は、人が予定を立てる際に意識する要因を一般化することで抽出できるものと考えられる。たとえば、「約2週間に1回実施しよう」や「土日祝日、盆正月には行わない」、「ある程度時期が近づいてから実施日を決めよう」といった要因である。これらの要因は、「特定の周期で発生する」や「平日か休日かに影響される」、「どの程度前に予定が確定するか」といった特徴を表現する要素として抽出できる。こうした要素を実在のリカーレンスの調査から抽出する。

一方、単発で発生する予定は、リカーレンスに比べて予測不可能な要因で発生する。そこで、単発で発生する予定は、リカーレンスに割ってランダムに入る予定として単純化し、予定名と実施時間のみを特徴として与えることとする。

3.2 調査対象と方法

調査対象は我々の所属する大学の研究室で利用している

表 1 調査対象のリカーレンス一覧

通番	種別	リカーレンス名
1	研究ミーティング	ミーティング 1
2	(約 2 週間に 1 回)	ミーティング 2
3	進捗報告会	進捗報告会 1
4	(約 1 カ月に 1 回)	進捗報告会 2
5		進捗報告会 3
6	年中行事 (半期に 1 回)	誕生日会
7	年中行事	研修会
8	(約 1 年に 1 回)	忘年会 1
9		忘年会 2
10		送別会 1
11		送別会 2
12		新人歓迎会
13		暑気払い
14		卒業アルバム写真撮影
15		M2 中間発表
16		B4 中間発表
17		防災訓練
18		M1 論文紹介
19		M2 論文紹介
20		卒論発表
21		修論発表
22		オープンキャンパス
23		研究室大掃除
24		大学の編入試験
25		大学の推薦入試
26		大学の卒業式
27	レクリエーション	レクリエーション
28	(1 年に約 3 回)	スポーツ大会
29	講義	講義 1
30	(約 1 週間に 1 回 (半期))	講義 2
31		講義 3
32		講義 4
33		講義 5
34	恒例行事	大学の前期入試
35	(決まった日付)	大学の後期入試
36		大学の入学式

共有カレンダーに含まれる 36 種類のリカーレンスである。表 1 に 36 種類のリカーレンスを示す。各リカーレンスに属する予定を立てる際に意識する固有の要因を列挙する。たとえば、表 1 の通番 1 の「ミーティング 1」というリカーレンスは「2 時間程度」、「約 2 週間間隔」、「平日のみ」、「カレンダー外の予定と重複しない」、「長期休暇には行わない」、および「特別研究報告書・修士論文の締切前には発生間隔が短くなる」といった固有の要因が存在する。その後、列挙した各リカーレンス固有の要因を一般化する。たとえば、「2 週間間隔」や「1 カ月間隔」といった、各リカーレンス固有の要因は「ある発生間隔をもつ」と一般化する。一般化したそれぞれの要因がカレンダー情報の特徴を表す要素と考える。また、複数の要素を組み合わせることでリカーレンスの特徴を表現できると考える。

3.3 調査結果

カレンダー情報の特徴を表す要素を表 2 に示す。これらの要素は調査対象の各リカーレンスがもつ固有の要因を一

表 2 カレンダー情報の特徴を表す要素

通番	要素名	説明
1	予定名	予定の名前
2	実施時間	何時間実施するか
3	発生間隔	周期が存在する
4	曜日	特定の曜日に影響される
5	休日	土日祝日といった休日に影響される
6	月	何月かに影響される
7	週	月の第何週目に影響される
8	日付	日付に影響される
9	非発生間隔	予定が発生しない期間が存在する
10	カレンダー外の予定	カレンダーには登録されない予定に影響される
11	順序関係	他の予定との間に順序関係が存在する
12	同時発生	他の予定と関連して同日に発生する
13	締切	何らかの締切が存在し、影響される
14	生成タイミング	予定がいつ生成されるか

般化した要因である。表 2 の通番 9 から 14 は予定を立てる際の行動に関わる要因であり、カレンダーに登録されている予定を後から見ただけでは、判断できない。このような要因を考慮することで、実際の予定を立てる過程を再現できる。

ここで、表 2 の通番 10 の「カレンダー外の予定」は、生成するカレンダーには存在しない予定の影響を受けるかどうかである。たとえば、「上司の出張」という予定はカレンダーには登録されないが、「会議」という予定は「上司の出張」と同日には発生しないと考える。また、表 2 の通番 14 の「生成タイミング」は、大まかな時期が決まっていて、その時期が近づくと生成する場合と次回の予定を逐次的に生成する場合の 2 種類に分類する。たとえば、「忘年会」といった予定は、実施する約 2 カ月前になると生成される。このため、時期が近づくと生成するという特徴をもつ。一方、「会議」といった予定は、その予定を行った際、他の予定を確認しながら、次回の実施日を決定する。このため、逐次的に生成するという特徴をもつ。

4. ダミーデータ生成手法

4.1 設定するパラメータ

4.1.1 リカーレンスごとに設定するパラメータ

リカーレンスごとに設定するパラメータは表 2 の 14 個の要素の組み合わせである。表 3 に表 2 の各要素に対応するパラメータ名を示す。ここで、表 3 の通番 14 の TIMING は、時期が近づくと生成する場合には「approaching」、次回の予定を逐次的に生成する予定には「successively」と設定する。表 3 のパラメータを組み合わせることによって、リカーレンスの特徴を表現する。たとえば、「月例会議」というリカーレンスの場合、表 3 の通番 1, 2, 3, 5, 9, 10, および 14 のパラメータを図 1 のように設定する。このような設定によって、特定のリカーレンスを表現する。また、複数のリカーレンスを組み合わせ、カレンダー利用者のモデルを表現する。

表 3 リカーレンスごとに設定するパラメータ

通番	要素名	パラメータ名
1	予定名	SUMMARY
2	実施時間	DURATION
3	発生間隔	INTERVAL
4	曜日	WDAY
5	休日	HOLIDAY
6	月	MONTH
7	週	MONTHWEEK
8	日付	DATE
9	非発生期間	VACATION_TERM
10	カレンダー外の予定	AVOID_HIDDEN_EVENTS
11	順序関係	ORDER
12	同時発生	SIMULTANEOUS
13	締切	DEADLINE
14	生成タイミング	TIMING

SUMMARY: 月例会議
DURATION: 120
INTERVAL: 30
HOLIDAY: false
AVOID_HIDDEN_EVENTS: true
VACATION_TERM:
- 8/10-8/14
- 12/25-1/3
TIMING: successively

図 1 リカーレンスに設定するパラメータの例

4.1.2 カレンダー情報に全体に対して設定するパラメータ

リカーレンスとは関係なく設定されるべきパラメータとして、4つのパラメータを考える。表 4 に4つのパラメータを示し、以下で各パラメータについて説明する。

(1) 生成範囲

ダミーデータを生成する日付の範囲を設定する。

(2) 1カ月当たりの単発で発生する予定の割合

単発で発生する予定が1カ月当たり存在している数を表し、この値に従って、ランダムな予定を生成する。たとえば、月に2回程度不定期に出張が発生する場合にこの値を「2」に設定する。

(3) 1年当たりのカレンダー外の予定の割合

カレンダー外の予定が存在する割合を設定する。たとえば、カレンダーには表れないが、自身が影響を受ける上司の予定がこれにあたる。「上司の出張」といった、生成するカレンダーには登録されない予定の年間の数を設定する。これらは、AVOID_HIDDEN_EVENTSが設定されたリカーレンスから参照される。

(4) 就業時間帯

予定を登録できる時間帯を設定する。このパラメータは予定を立てる際の行動に関わるパラメータである。このパラメータを設定することで、1日に登録できる時間を超過してしまう予定は生成しないで、別の日に生成することで、1日に予定が偏って生成されることを防止する。

設定するパラメータの例を図 2 に示す。図 2 の例では、生成範囲は2015年度、1カ月当たりの単発で発生する予定の割合は月に2件、1年当たりのカレンダー外の予定の割合

表 4 カレンダー情報全体に対して設定するパラメータ

通番	要素名	パラメータ名
1	生成範囲	RANGE
2	1カ月当たりの単発で発生する予定の割合	RANDOM_EVENTS_PER_MONTH
3	1年当たりのカレンダー外の予定の割合	HIDDEN_EVENTS_PER_YEAR
4	就業時間帯	WORKING_HOUR

RANGE: 20150401-20160331
RANDOM_EVENTS_PER_MONTH: 2
HIDDEN_EVENTS_PER_YEAR: 50
WORKING_HOURS: 10:00-17:00
RECURRENT:

-
SUMMARY: ミーティング1
DURATION: 120
INTERVAL: 14
HOLIDAY: false
AVOID_HIDDEN_EVENTS: true
VACATION_TERM:
- 8/10-8/14
- 12/25-1/3
DEADLINE:
- 2/5
TIMING: successively

-
SUMMARY: ミーティング2
DURATION: 120
INTERVAL: 14
HOLIDAY: false
AVOID_HIDDEN_EVENTS: true
VACATION_TERM:
- 8/10-8/14
- 12/25-1/3
TIMING: successively

図 2 カレンダー情報全体に設定するパラメータの例

は50件、就業時間帯は10時から17時となっている。また、RECURRENT以下にリカーレンスごとのパラメータの組み合わせを記載している。

4.2 カレンダー情報のダミーデータ生成手順

4.2.1 方針

提案手法ではダミーデータの生成を行う際、複数のリカーレンスからなるデータと単発で発生する予定のデータを組み合わせる。生成手法の方針を以下に述べる。

(方針 1) 優先度の高い予定から生成

予定に優先度を設定し、その優先度の高い順に予定を生成する。実際のカレンダー利用者の行動では、優先度の高い予定が後から生成される場合、優先度の低い予定を別の日付にずらすことがある。このため、その行動と同等の結果を再現するために優先度の高い予定から先に生成することとする。

(方針 2) 単発で発生する予定をリカーレンスに属する予定よりも優先

単発に発生する予定は外部の人と関わる予定や突発的に発生する予定である場合が多く、実施日の変更が難しい場合が多いと考えられる。このため、リカーレンスに属する予定よりも優先度が高いと考える。

(方針 3) 実施日を決定する過程を再現

リカーレンスの周期は曖昧である場合が多く、実際に予定の実施日を決定する際には、複数の要素を考慮し

ている。たとえば、2週間に1回という周期であっても、「お盆や年末年始には発生しない」や「祝日には発生しない」といった特徴をもっている場合、2週間に1回ではない場合も存在する。リカーレンスに属する予定のデータを生成するためには、実際に予定の実施日を決定する過程を再現し、逐次的にデータを生成する。

(方針4) 単発で発生する予定やカレンダー外の予定はランダムに生成

単発で発生する予定には決まった規則や特徴が存在しないため、パラメータとして設定したRANDOM_EVENTS_PER_MONTHの値に従って、予定をランダムに生成する。また、カレンダー外の予定も同様にパラメータのHIDDEN_EVENTS_PER_YEARの値に従って、予定をランダムに生成する。

4.2.2 予定の優先度

(方針1)に示した通り、ダミーデータを生成する際、優先度の高い予定を先に生成しなければならない。予定の優先度はさまざまな要因によって決定されるが、ここでは、生成する予定の優先度を以下の3つとした。

(高) 単発で発生する予定

(中) 時期が近づくと生成されるリカーレンスの予定

(低) 逐次的に生成されるリカーレンスの予定

また、同じ優先度に属する予定間では、発生間隔が長いものを優先する。以下にこれらの理由を説明する。

(理由1) 単発で発生する予定は実施日の変更が困難

単発で発生する予定は定期的には関わらない外部の人と関わる予定であったり、突発的に発生したりするため、実施日の変更が難しい場合が多く、優先度は高いと考える。たとえば、不定期に発生する「出張」という予定は、出張先の場所や人の予定も関係するため、実施日の変更は難しく、優先度は高い。

(理由2) 逐次的に生成される予定は他の予定を確認しながら実施日を決定

逐次的に生成される予定は予定の生成時に既に決定している予定を確認し、すでに別の予定が決定している場合は重複しないように実施日を決定するため、優先度は低いと考える。

(理由3) 周期の長い予定は短い予定より実施日の変更が困難

周期の長い予定は短い予定よりも先に実施日が決定したり、変更できる日付の候補日が少なかったりするため、優先度は高いと考える。

以上の理由から、予定の優先度を決定する際は、生成する予定の種類に応じた優先度を設定し、その中で周期の長い予定を優先するように設定する。

4.2.3 ダミーデータ生成手順

4.2.2 項の優先度順に予定を生成する。カレンダー情報の

ダミーデータの生成手順を以下に示す。

(手順1) 単発で発生する予定を生成

表4の通番2のRANDOM_EVENTS_PER_MONTHに設定した値に従って、単発で発生する予定をランダムに生成する。具体的には、生成範囲の各日付に対して、RANDOM_EVENTS_PER_MONTHに応じた確率で予定を生成し、カレンダーに追加する。また、予定名は「単発で発生する予定」とし、実施時間は就業時間帯内に収まる時間からランダムに設定する。

(手順2) カレンダー外の予定を生成

表4の通番3のHIDDEN_EVENTS_PER_YEARに設定した値に従って、生成するカレンダーには表示されない予定をランダムに生成する。具体的には、(手順1)と同様に生成範囲の各日付に対して、HIDDEN_EVENTS_PER_YEARに応じた確率で予定を生成する。生成した予定はカレンダーに追加しないがAVOID_HIDDEN_EVENTSを設定した他の予定はこれらの予定が生成された日付には生成されないとする。

(手順3) 時期が近づいた場合に生成されるリカーレンスの予定を生成

表3の通番14のTIMINGが「approaching」のリカーレンスの予定を生成する。生成する順番は4.2.2項の(理由3)で示した通り、周期の長いリカーレンスから生成する。また、生成する各予定の実施日を決定する手順は次節に示す。

(手順4) 逐次的に生成されるリカーレンスの予定を生成

TIMINGが「successively」のリカーレンスの予定を生成する。(手順2)と同様に、生成する順番は4.2.2項の(理由3)で示した通り、周期の長いリカーレンスから生成する。各予定の実施日を決定する手順は次節に示す。それぞれのリカーレンスについて、1件の予定を生成後、次の予定を生成するタイミングを保持する。生成範囲の日付を1日づつ順番に確認し、予定を生成するタイミングのリカーレンスがあれば、そのリカーレンスの予定を生成し、次の予定を生成するタイミングを上書きする。生成範囲内の日付を最後まで確認し終わると、ダミーデータの生成を終了する。

4.2.4 リカーレンスに属する予定の実施日の決定手順

4.2.3 項の(手順3)と(手順4)で用いる、リカーレンスに属する予定の実施日を決定する際の手順を以下に示す。

(手順1) 生成範囲のすべての日付に評価値を付与

リカーレンスに設定されたパラメータに従って、日付に評価値を付与する。たとえば、「休日には発生しない」というパラメータを与えた場合、土曜日や日曜日には低い評価値を与える。また、「12月に発生する」というパラメータを与えた場合、12月1日から12月31日に高い評価値を与える。

(手順2) 1つ前の予定から「発生間隔」付近の日付に加点

前回の予定の日付から「発生間隔」で設定された日数分先の日付を基準日とし、基準日から前後数日に評価値を加算する。たとえば、「発生間隔」が14日である場合、前回から14日後を基準日とし、基準日の前後3日に加算する。

(手順3) 評価値の合計が最も高い日付を実施日として選択

(手順1)と(手順2)で与えた評価値が最も高い日付を候補とする。候補が複数存在する場合、基準日から近い日付を優先する。このとき、パラメータの「就業時間帯」に予定が生成できる空きがない場合は、その日付の評価値を下げ、(手順3)を繰り返す。

(手順4) 実施日に予定を生成

その日付に予定を生成し、カレンダーに追加する。選択した実施日が生成範囲外だった場合、生成を終了する。そうでなければ、(手順2)に戻り、次の予定を生成する。

以上の手順でリカーレンスに属する予定の実施日を決定し、予定を生成する。この手順で決定した実施日は単純に発生間隔ごとに繰り返すのではなく、リカーレンスの特徴を反映している。

5. パラメータの有効性の確認

5.1 確認対象

提案手法を実装し、ダミーデータの生成を行った。生成したダミーデータを確認し、他の予定との重複を考慮するパラメータが有効に働いているか確認する。確認対象とするパラメータは以下の2つである。

WORKING_HOURS: 就業時間帯

生成するカレンダー情報に対して設定するパラメータであり、予定を生成できる時間を設定している。予定の実施日の決定時に、この時間内に生成できない場合は、別の実施日に変更する。

AVOID_HIDDEN_EVENTS: カレンダー外の予定

生成するリカーレンスに対して設定するパラメータであり、上司の出張といったカレンダーには登録されない予定に実施日の選択が影響されるかを表す。影響されるリカーレンスの予定は、カレンダー外の予定が入っている日付を選択する確率が下がる。

これらのパラメータが有効に働いていない場合、「就業時間帯」を超える予定が生成されたり、「カレンダー外の予定」と重複して予定が生成されたりする。このため、上記の2つのパラメータを対象にパラメータが有効に働いているか確認する。

5.2 確認に用いるダミーデータ

対象の2つのパラメータがそれぞれ有効に働いているか確認するために、以下の3種類のダミーデータを生成する。

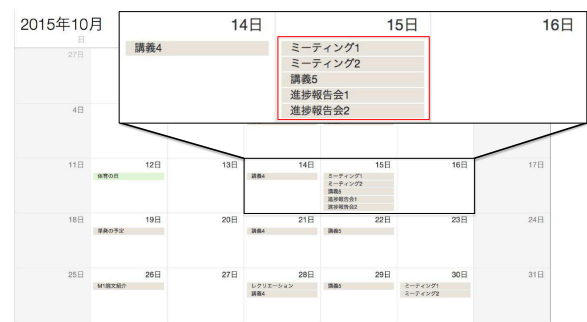


図3 比較1:「就業時間帯」を考慮しないダミーデータ



図4 比較1:「就業時間帯」と「カレンダー外の予定」を考慮したダミーデータ

- (1) 「就業時間帯」「カレンダー外の予定」を考慮
- (2) 「就業時間帯」を考慮しない
- (3) 「カレンダー外の予定」を考慮しない

生成するダミーデータのモデルは3.1節で調査を行った、我々の研究室内の共有カレンダーである。パラメータの設定は、生成範囲は2015年度、就業時間帯は10時から17時までの7時間、単発で発生する予定の割合は月に2件、およびカレンダー外の予定は年間50件とした。生成するリカーレンスの種類は36種類でそれぞれに調査した特徴を表すパラメータを設定した。生成したダミーデータを用いて、次節でパラメータが有効に働いているか確認する。

5.3 各パラメータの有効性の確認と結果

5.3.1 「WORKING_HOURS」の有効性

「就業時間帯」を考慮しないダミーデータを図3に、「就業時間帯」と「カレンダー外の予定」を考慮したダミーデータ(提案手法)を図4にそれぞれ示す。両者の10月15日に注目すると、図3では、5件の予定が生成されており、合計時間は9時間半で、設定した「就業時間帯」を超えている。一方、提案手法では、3件の予定が生成されており、合計時間は5時間半で、設定した「就業時間帯」に収まっている。時間を超過してしまった予定について確認すると、提案手法では「ミーティング1」が翌日の10月16日、「ミーティング2」が前日の10月14日に生成されている。このため、WORKING_HOURSのパラメータが有効に働き、予定を別の日付に分散できている。



図 5 比較 2: 「カレンダー外の予定」を考慮しないダミーデータ

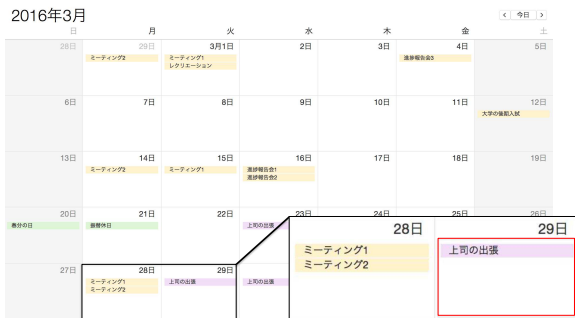


図 6 比較 2: 「就業時間帯」と「カレンダー外の予定」を考慮したダミーデータ

5.3.2 「AVOID_HIDDEN_EVENTS」の有効性

「カレンダー外の予定」を考慮しないダミーデータを図 5 に、「就業時間帯」と「カレンダー外の予定」を考慮したダミーデータ (提案手法) を図 6 にそれぞれ示す。また、両者にはカレンダー外の予定を「上司の出張」という予定名で表示している。両者の 3 月 29 日に注目すると、図 5 は、「上司の出張」と「ミーティング 1」が重複している。「ミーティング 1」が属するリカーレンスは「カレンダー外の予定と重複しない」という特徴をもつため、この特徴を満たせていない。一方、提案手法では、「上司の出張」が存在する日付には予定が生成されていない。重複していた「ミーティング 1」を確認すると、提案手法では前日の 3 月 28 日に生成されている。このため、AVOID_HIDDEN_EVENTS のパラメータが有効に働き、カレンダー上には存在しない予定との重複を回避できている。

6. おわりに

本稿では、想定したカレンダー利用者のモデルを再現するダミーデータ生成手法を提案した。まず、テストケースとしてカレンダー情報を利用する場合の問題を示した。具体的には、まったくのランダムなデータでは実際の利用に則した評価が困難なことから実データを利用する場合、プライバシーや守秘義務に関わる予定や情報が含まれているため、収集が困難なことである。次に、カレンダー情報の特徴を表す要素を調査した。我々の所属する大学の研究室の共有カレンダーを対象に予定を立てる際に意識する要因を一般化

し、14 種類の要素を抽出した。その後、ダミーデータを生成手順を示した。提案手法では、予定の優先度の設定や実施日の決定手順によって、予定を立てる際の行動を再現する。最後に、提案手法を実装し、他の予定との重複を考慮するパラメータについて、そのパラメータが有効に働いているかどうか確認した。

今後の課題として、生成したダミーデータの定量的な評価がある。設定したパラメータが有効に働いていることを示したが、生成したダミーデータと想定したカレンダー情報のモデルとの類似度や生成できるダミーデータの多様性を評価する必要がある。また、リカーレンスの特徴を個別に設定するのではなく、職業、役職、および 1 か月や 1 日当たりの予定数といった情報から、パラメータの組み合わせを生成する必要がある。これを実現し、ライブラリ化すると、想定したカレンダー利用者を表現しやすくなると考える。

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。

参考文献

- [1] Faulring, A. and Myers, B. A.: Availability bars for calendar scheduling, *In Proceedings of the Conference on Human Factors in Computing Systems Extended Abstracts (CHI)*, ACM, pp. 760-765 (2006).
- [2] 西野正彬, 中村幸博, 武藤伸洋, 阿部匡伸: あいまいな表現を含むスケジューラデータと GPS データとの時間的共起関係を利用した行動予測手法の検討, 電子情報通信学会技術研究報告, Vol. 109, No. 450, pp. 73-78 (2010).
- [3] 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).
- [4] kazina: なんちゃって個人情報, kazina.com (オンライン), 入手先 (<http://kazina.com/dummy>) (参照 2016-10-12).
- [5] iGoPayGo Soft Gen: Credit Card Generator, iGoPayGo Soft Gen (online), available from (<https://names.igopaygo.com/credit-card>) (accessed 2016-10-12).
- [6] Marciel, M., Michelinakis, F., Fanou, R. and Muñoz-Merino, P. J.: Enhancements to Google Course Builder: Assessments Visualisation, YouTube Events Collector and Dummy Data Generator, *XV Simposio Internacional de Tecnologas de la Informacin y las Comunicaciones en la Educacin (SINTICE 2013)*, pp. 6-13 (2013).
- [7] 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, pp. 1-6 (2011).