

期間同士の関係の基づくあいまいな時間データの精緻化

関野 樹 (国際日本文化研究センター)

概要: 人文学研究において散見される「14世紀～16世紀」といった始点や終点があいまいな期間を処理するための新たな手法が提案された。この手法は、期間同士の関係に基づいて始点や終点がとり得る範囲を絞ることによって、データのあいまいさを低減(精緻化)させるとともに、複数の期間同士の関係における矛盾やデータの誤りの検出も可能にする。これらの手法を実データに適用するため、データの構築における自然言語処理技術の必要性や Linked Data を用いたデータ提供方法について検討が加えられた。

キーワード: 時間情報, あいまいなデータ, 推論, 矛盾検出, Linked Data

Ambiguous temporal data refining based on relations between time intervals

Tatsuki Sekino (International Research Center for Japanese Studies)

Abstract: This study proposed novel methods to process ambiguous temporal data such as “14th – 16th century” often seen in humanities researches. The method enables to reduce ambiguity of the ambiguous temporal data (refinement) and to detect contradiction in relations between the temporal data. In order to apply these methods to practical data, this study examined necessity of natural language processing in data construction and possibility to use Linked Data technology for data distribution.

Keywords: Temporal Information, Ambiguous Data, Reasoning, Contradiction Detection, Linked Data

1. はじめに

時間は、さまざまな情報に含まれる最も一般的なデータの一つである。しかしながら、実際には精度が可視化や解析などの各種処理に必要な水準に満たないあいまいな時間データが数多く存在する。これは、人の手で記述された過去のデータを扱う機会が多い人文分野でより顕著である。

あいまいな時間データの多くは、期間の始点・終点や時間点が定まらず、範囲として示される。例えば、あるできごとが起きた時間が「14世紀～16世紀」、「江戸中期」といった形で表現される。一方で、時間データを計算機で扱おうとする場合、一般的には期間の始点・終点や時間点は時間軸上の点であることを前提としている。したがって、より正確な結果を得るには、時間範囲として表現されたこれらの期間の始点・終点や時間点の幅をできるだけ絞り込む必要がある。

あいまいなデータを計算機上で扱うために、これまでにも数多くの研究がなされてきた[1]。しかしながらこれらの研究は、ファジー集合[2,3]やラフ集合[4,5,6]を用いることにより、あいまいなデータをあいまいなまま処理するためのものであり、データのあいまいさそのものを低減させる手法ではない。これに対し、Sekino(2021)[7]は、期間同士の関係に着目することにより、時間データのあいまいさを低減させる手法を提案した。例えば、ある二人の人物の生存期間を考えると、この二人が母と子の関係であれば、子の生存期間が母の生存期間より早く始まることはあり得ない。したがって、子の生存期間があいまいで、その始点が

時間範囲として示されていたとしても、母の生存期間より前の部分は削除することができる。このように、2つのあいまいな期間同士の関係が既知である場合、その関係に基づいて、その始点や終点を示す時間範囲の幅を狭める、つまり、あいまいさを低減するための操作(以降、「精緻化」と呼称する)が可能となる。本研究では、Sekino(2021)[7]のあいまいな期間の緻密化の考え方を、その原理や定式化の過程を踏まえながら手法として確立させるとともに、時間データの誤りやデータ間の矛盾を検出するなどの精緻化操作を応用した新たな手法を検討する。さらに、これらの手法に適用可能なデータの構築・提供方法についての課題を整理し、手法を実データへ適用するための道筋をつける。

2. あいまいな期間の精緻化

2-1. あいまいな期間の表現

本研究では、始点または終点が不明確な期間をあいまいな期間とする。あいまいな期間では、本来時間点として表されるべきこれらの端点が時間範囲として表され、端点の真値はこれらの時間範囲内に存在する(図1)。両端点を表す時間範囲が分かれている場合と重なっている場合に分けて考えると、両時間範囲が重ならない場合は(図1上図)、両時間範囲の間は、両端点がどの位置に確定しようとも当該期間に確実に含まれる範囲である(確定期間)[8]。一方、両時間範囲を含む期間全体は、端点確定後の期間に含まれる可能性がある範囲である(可能期間)。見方を変えれば、

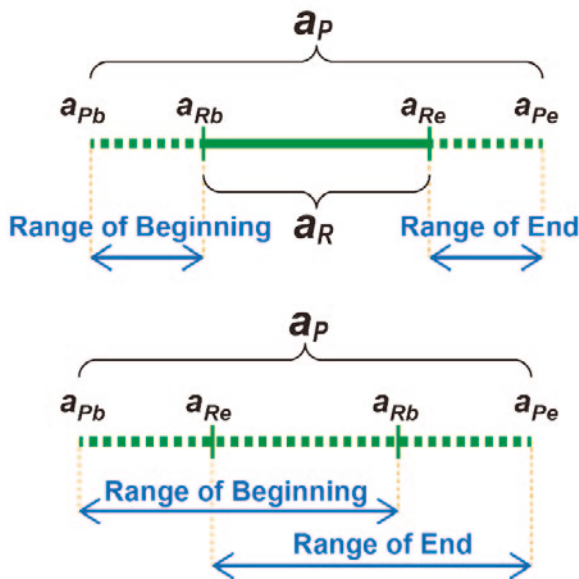


図 1 あいまいな時間の表現[8]. 上図: 始点と終点の時間範囲が重ならない場合, 下図: 両範囲が重なる場合.

Figure 1 Representation of uncertain time intervals [8]. Upper figure: When the time ranges of the beginning and end points do not overlap, Lower figure: When both ranges overlap.

可能期間の外側は、両端点がどの位置に確定しようとも当該期間に含まれることはない。ここで、あいまいな期間 a の確実期間と可能期間をそれぞれ a_R および a_P とすると、始点の範囲は可能期間の始点 (a_{Pb}) から確実期間の始点 (a_{Rb})、終点の範囲は確実期間の終点 (a_{Re}) から可能期間の終点 (a_{Pe}) となる。 $a_{Rb} > a_{Re}$ となると、両端点の時間範囲が重なった状態となり (図 1 下図)、確実期間 (a_R) は存在しない。

あいまいな期間が精緻化されるということは、両端点を表す時間範囲が縮小することである。たとえば、あいまいな期間 a が精緻化されて始点の時間範囲が縮小するという事は、可能期間の始点 (a_{Pb}) の値が大きくなる (図 1 では右に動く) か、確実期間の始点 (a_{Rb}) の値が小さくなる (図 1 では左に動く) ことである。終点の範囲も同様に、確実期間の終点 (a_{Re}) の値が大きくなるか、可能期間の終点 (a_{Pe}) の値が小さくなる。結果として、精緻化により可能期間 (a_P) は狭まり、確実期間 (a_R) は広がる。上述のとおり両端点の時間範囲が重なった場合には確実期間は存在しないが、精緻化によって a_{Rb} が小さくなる、または a_{Re} が大きくなることによって $a_{Rb} \leq a_{Re}$ となり (両範囲が重ならなくなり)、精緻化前には存在しなかった確実期間 (a_R) が新たに出現することも起こり得る。

2-2. 精緻化操作の考え方

2つのあいまいな期間 a と b を両者の関係に基

づいて精緻化する。 a と b の関係は Allen の時間関係[9]で表現する (図 2)。たとえば、 a と b の端点が確定した際に Allen の **during** 関係であること (以降、 a **during** b と表記) が既知であるとする、図 2 のとおり、確定後の a の始点と終点は、 b の内側に位置しなければならない。したがって、 a の両端点の範囲のうち、 b の可能期間 (b_P) の外側 (つまり、 b がどのように確定しようとも b には含まれない範囲) にはみ出した部分は削除することができる。図 3 上図の例では、結果として、 a の可能期間の始点 (a_{Pb}) を b の可能期間の始点 (b_{Pb}) まで移動させることができる (図 3-(1))。一方、図 3 上図の a の可能期間の終点 (a_{Pe}) は、すでに可能期間 (b_P) の内側にあるので、移動することはできない。さらに、 a **during** b であるならば、 a の両端がどのように確定しようとも、 a の確実期間 (a_R) は b に含まれていなければならない。 b から見れば、 b の両端がどのように確定しようとも、 b は a の確実期間 (a_R) を包含することが確実である。したがって、 b の確実期間 (b_R) を a_R の範囲まで広げることができる。図 3 上図では、結果として、 b の確実期間の終点 (b_{Re}) を b の確実期間の終点 (a_{Re}) まで移動させることができる (図 3-(2))。



図 2 Allen の時間関係. 期間同士の関係が 13 通りに分類される。 **equals** 関係以外は逆の関係が存在する (例: **before** 関係と **after** 関係は互いに逆の関係)。

Figure 2 Allen's time interval algebra which classify relations between time intervals into thirteen types. Each Allen's relation except of the **equals** relation has the inverse relation (e.g., the **before** and the **after** relations are the inverse relation each other).

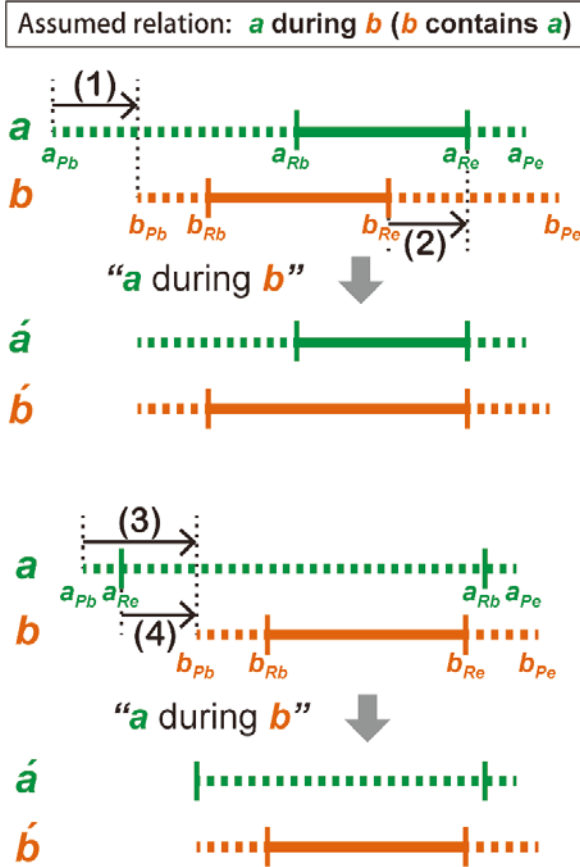


図3 あいまいな期間の精緻化の基本概念。実線および点線は、それぞれ確実期間および可能期間を示す。あいまいな期間 a と b が両者の関係 a during b に基づいて精緻化され、それぞれ、 \hat{a} と \hat{b} を得る過程を示す。上図と下図は、それぞれ、 a が確実期間(a_R)を持つ場合と持たない場合を示す。

Figure 3 Basic concept of refining uncertain time intervals. The solid and the dotted lines represent the reliable and the possible time intervals, respectively. Uncertain time intervals a and b are refined based on the relation of a during b , and become \hat{a} and \hat{b} . The upper and the lower figures show the case where a has the reliable time interval (a_R) and the case where it does not, respectively.

同様の精緻化の操作は、 a の始点と終点の範囲が重なって確実期間が存在しない場合 ($a_{Rb} > a_{Re} \Rightarrow a_R = \emptyset$) でも可能である (図3下図)。図3(3)は、図3(1)と同様に、 b の可能期間 (b_P) の外側にある a の始点の範囲を削除し、 a の始点の時間範囲の始点 (a_{Pb}) を b の可能期間の始点 (b_{Pb}) まで移動させる操作を示している。また、図3下図では、 a の終点の時間範囲の始点 (a_{Re}) も b_P の外側にあるため、 a_{Re} を b_P の始点 (b_{Pb}) まで移動させることができる (図3(4))。

2-3. 精緻化操作の定式化

前項で示したあいまいな期間の精緻化を計算機上で実行するために、その操作手順を定式化する。

る。あいまいな期間のあいまいさが完全に解消されて \hat{a} と \hat{b} に確定した場合に、 \hat{a} と \hat{b} のAllenの関係 $\lambda \in \{\text{before, after, ...}\}$ が満たすべき条件を $P_\lambda(\hat{a}, \hat{b})$ とする。確定した場合に関係 λ となるためには、精緻化前の a と b は関係 λ について「可能な関係」[8]であるはずなので、

$$a = \{\hat{a} | \exists \hat{b} P_\lambda(\hat{a}, \hat{b})\}$$

である。あいまいな期間 a を期間 b との関係 λ に基づいて精緻化する操作は、 a が確定した場合の期間 $\hat{a} \in a$ のうち、すべての \hat{b} に対してこの条件を満たすことができない \hat{a} を取り除く操作である。したがって、精緻化後の a を \acute{a} とすると、

$$\begin{aligned} \acute{a} &= a - \{\hat{a} | \forall \hat{b} \neg P_\lambda(\hat{a}, \hat{b})\} \\ &= \{\hat{a} | \forall \hat{b} \exists \hat{b} P_\lambda(\hat{a}, \hat{b})\} \end{aligned} \quad (1)$$

となる。

次に、 \hat{a} の始点と終点をそれぞれ \hat{a}_b と \hat{a}_e 、 \hat{b} の始点と終点をそれぞれ \hat{b}_b と \hat{b}_e とし、式(1)の考えに基づいて各端点を満たすべき条件から具体的な精緻化操作を導く。例として、 $\lambda = \text{during}$ である場合、確定した期間の端点を満たすべき条件 $P_{\text{during}}(\hat{a}, \hat{b})$ は、**during** 関係の定義より

$$\hat{a}_b > \hat{b}_b \wedge \hat{a}_e < \hat{b}_e \wedge \hat{b}_b < \hat{b}_e \quad (2)$$

である。条件の第1項 $\hat{a}_b > \hat{b}_b$ に着目すると、精緻化の操作は、 \hat{a}_b ($a_{Pb} \leq \hat{a}_b < a_{Rb}$) の中からいずれの \hat{b}_b ($b_{Pb} \leq \hat{b}_b < b_{Rb}$) に対しても $\hat{a}_b > \hat{b}_b$ とならない \hat{a}_b を除く操作であり、

$$\{\hat{a}_b | a_{Pb} \leq \hat{a}_b < a_{Rb}\} - \{\hat{a}_b | \forall \hat{b}_b \neg (\hat{a}_b > \hat{b}_b)\}$$

となる。この結果、精緻化後の \hat{a}_b がとり得る値は以下の条件を満たす。

$$\forall \hat{a}_b \exists \hat{b}_b (\hat{a}_b > \hat{b}_b) \quad (3)$$

この条件を満たすには、 \hat{a}_b と \hat{b}_b のとり得る範囲が以下の関係を満たせばよい。

$$a_{Pb} > b_{Pb}$$

したがって、精緻化操作の前の状態が $a_{Pb} < b_{Pb}$ である場合は、少なくとも a_{Pb} を b_{Pb} の位置まで移動させることができる。一方、精緻化操作の前の状態が $a_{Pb} > b_{Pb}$ である場合は、既に式(3)を満たしているため、 a_{Pb} は動かすことができない。 $a_{Pb} = b_{Pb}$ の場合も同様に、 a_{Pb} は動かすことができない。精緻化操作後の a_{Pb} の値を \acute{a}_{Pb} とすると、これらの操作は、

$$\acute{a}_{Pb} = \max(a_{Pb}, b_{Pb})$$

と表現できる。

さらに、他の端点についても精緻化の操作を定式化する。式(2)の定義を \hat{a} と \hat{b} の始点と終点のす

すべての関係に整理しなおすと,

$$\hat{a}_b > \hat{b}_b \wedge \hat{a}_b < \hat{b}_e \wedge \hat{a}_e > \hat{b}_b \wedge \hat{a}_e < \hat{b}_e \quad (4)$$

となる. あいまいな期間 a を期間 b との Allen の関係 λ に基づいて精緻化し, \acute{a} を得る操作を $R(\lambda, b)(a)$ とし, 式(4)のそれぞれの項について精緻化の操作を導くと,

$$\acute{a} = R(\text{during}, b)(a) = \begin{cases} \acute{a}_{Pb} = \max(a_{Pb}, b_{Pb}) \\ \acute{a}_{Rb} = \min(a_{Rb}, b_{Pe}) \\ \acute{a}_{Re} = \max(a_{Re}, b_{Pb}) \\ \acute{a}_{Pe} = \min(a_{Pe}, b_{Pe}) \end{cases}$$

を得ることができる. 同様に, 残りの Allen の関係についての精緻化操作も導出すると, 表 1 のとおりとなる.

表 1 で定数が指定されている操作は, 精緻化により値が変化しないことを示す (例: **before** の, \acute{a}_{Pb} など). 図 3 で示したように, あいまいな期間 a を期間 b との関係により精緻化すると同様に, b を a との関係によりを精緻化し, \acute{b} を得る操作も可能である. この場合は, Allen の関係における逆の関係をあてはめて考える. つまり, a **during** b であれば b **contains** a であるので (図 2), 表 1 より,

$$\acute{b} = R(\text{contains}, a)(b) = \begin{cases} \acute{b}_{Pb} = b_{Pb} \\ \acute{b}_{Rb} = \min(b_{Rb}, a_{Rb}) \\ \acute{b}_{Re} = \max(b_{Re}, a_{Re}) \\ \acute{b}_{Pe} = b_{Pe} \end{cases}$$

となる. なお, 一組の a と b の精緻化において, a の精緻化操作と b の精緻化操作の順序は最終的な

結果 \acute{a} と \acute{b} に影響しないことが明らかになっている [7].

表 1 によると, \acute{a}_{Pb} に関する操作は, 定数となるもの以外は全て **max** 関数で表現されている. つまり, 精緻化により a_{Pb} が移動する場合は常に値の大きい方に移動することを示している. 一方, \acute{a}_{Rb} に関する操作は, 定数となるもの以外は全て **min** 関数で表現されており, 精緻化により a_{Rb} が移動する場合は常に値の小さい方に移動することを示している. 結果として, あいまいな期間 a の始点の範囲はこれらの操作によって縮小し, あいまいさが低減されていることが分かる. これは, 終点の範囲についても同様である. また見方を変えると, これらの操作により, あいまいな期間 a の可能期間 (a_P) は縮小し, 確実期間 (a_R) は拡大する. この点からも, これらの操作によってあいまいな期間のあいまいさが低減されることを見て取れる.

3. 精緻化操作の展開

3-1. 複数の関係に基づく精緻化操作

前節では単独の Allen の関係に基づいた精緻化操作を導出したが, 現実のあいまいな期間同士の関係では, 一組のあいまいな期間に複数の関係を想定することができる. たとえば, 実母と子の関係において, 子から見た場合の母親の生存期間は, 一般的には **overlapped-by** または, **during** である. さらに事故などの痛ましいケースによる **finishes** と **met-by** も含めれば, 4 通りの関係が想定できる (図 2 参照). 実父と子の関係では, 子の出生前に父が死亡するケースが想

表 1 あいまいな期間 a をあいまいな期間 b との Allen の関係 λ に基づいて精緻化する操作.

Table 1 Operations to refine uncertain time interval a based on Allen's relation λ with b .

λ	\acute{a}_{Pb}	\acute{a}_{Rb}	\acute{a}_{Re}	\acute{a}_{Pe}
before	a_{Pb}	$\min(a_{Rb}, b_{Rb})$	a_{Re}	$\min(a_{Pe}, b_{Rb})$
after	$\max(a_{Pb}, b_{Re})$	a_{Rb}	$\max(a_{Re}, b_{Re})$	a_{Pe}
during	$\max(a_{Pb}, b_{Pb})$	$\min(a_{Rb}, b_{Pe})$	$\max(a_{Re}, b_{Pb})$	$\min(a_{Pe}, b_{Pe})$
contains	a_{Pb}	$\min(a_{Rb}, b_{Rb})$	$\max(a_{Re}, b_{Re})$	a_{Pe}
overlaps	a_{Pb}	$\min(a_{Rb}, b_{Rb})$	$\max(a_{Re}, b_{Pb})$	$\min(a_{Pe}, b_{Pe})$
overlapped-by	$\max(a_{Pb}, b_{Pb})$	$\min(a_{Rb}, b_{Pe})$	$\max(a_{Re}, b_{Re})$	a_{Pe}
meets	a_{Pb}	$\min(a_{Rb}, b_{Rb})$	$\max(a_{Re}, b_{Pb})$	$\min(a_{Pe}, b_{Rb})$
met-by	$\max(a_{Pb}, b_{Re})$	$\min(a_{Rb}, b_{Pe})$	$\max(a_{Re}, b_{Re})$	a_{Pe}
starts	$\max(a_{Pb}, b_{Pb})$	$\min(a_{Rb}, b_{Rb})$	$\max(a_{Re}, b_{Pb})$	$\min(a_{Pe}, b_{Pe})$
started-by	$\max(a_{Pb}, b_{Pb})$	$\min(a_{Rb}, b_{Rb})$	$\max(a_{Re}, b_{Re})$	a_{Pe}
finishes	$\max(a_{Pb}, b_{Pb})$	$\min(a_{Rb}, b_{Pe})$	$\max(a_{Re}, b_{Re})$	$\min(a_{Pe}, b_{Pe})$
finished-by	a_{Pb}	$\min(a_{Rb}, b_{Rb})$	$\max(a_{Re}, b_{Re})$	$\min(a_{Pe}, b_{Pe})$
equals	$\max(a_{Pb}, b_{Pb})$	$\min(a_{Rb}, b_{Rb})$	$\max(a_{Re}, b_{Re})$	$\min(a_{Pe}, b_{Pe})$

定できるので、これに **after** が加わり 5 通りとなる。

このような複数の関係に基づく精緻化の操作については、Sekino (2021) [7] で検討がなされている。これらの関係は「または」の連結されていることから、最も変化の少ない操作を選択すればよく、 a と b は関係の集合を Λ_{ab} とすると、

$$\acute{a} = R(\Lambda_{ab}, b)(a) = \begin{cases} \acute{a}_{pb} = \min_{\lambda \in \Lambda_{ab}} (R(\lambda, b)(a_{pb})) \\ \acute{a}_{Rb} = \max_{\lambda \in \Lambda_{ab}} (R(\lambda, b)(a_{Rb})) \\ \acute{a}_{Re} = \min_{\lambda \in \Lambda_{ab}} (R(\lambda, b)(a_{Re})) \\ \acute{a}_{Pe} = \max_{\lambda \in \Lambda_{ab}} (R(\lambda, b)(a_{Pe})) \end{cases}$$

となる。

3-2. 複数の期間との関係に基づく精緻化

実際にあいまいな期間を確定させてゆく過程では、他の複数の期間との関係を用いることになる。例えば、人物の生存期間の推定では、両親や子供など人物同士の関係のほか、著書、書簡、絵画などの著作物の制作年、当該人物について書かれた文書や古記録の日付など、さまざまな状況証拠を積み重ねながら当該人物の生存期間を絞り込んでゆくことになる。この場合、どのような順番で精緻化操作を行うかが問題になるが、期間同士のそれぞれ関係に矛盾が無い限り、精緻化操作の順番は最終的な結果に影響しないことが明らかになっている[7]。

3-3. 期間同士の関係における矛盾の検出

精緻化操作を実施するには、 a と b が想定された Allen 関係について「可能な関係」[8]であることが前提である。これを用いて、期間同士のそれぞれ関係に矛盾があるかを検証することができる。たとえば、図4で示すあいまいな期間 a 、 b および c において、 a を関係 ***a overlapped-by b*** および ***a during c*** に基づいて精緻化することを考える。この場合、***a overlapped-by b*** が「可能な関係」である条件は

$$a_{Rb} > b_{pb} \wedge a_{pb} < b_{pe} \wedge a_{pe} < b_{Re} \\ \wedge a_{pb} < a_{pe} \wedge b_{pb} < b_{pe}$$

a ***during c*** が「可能な関係」である条件は

$$a_{Rb} > c_{pb} \wedge a_{Re} < c_{pe} \wedge c_{pb} < c_{pe}$$

である[8]。精緻化前は、 a 、 b および c はそれぞれこれらの条件を満たしており、「可能な関係」である。しかしながら、***a overlapped-by b*** に基づいて精緻化された \acute{a}_{ab} は、 a ***during c*** の条件のうち $a_{Rb} > c_{pb}$ を満たさず、 \acute{a}_{ab} ***during c*** は「可能な関係」とはならない。同様に、 a ***during c*** に基づいて精緻化された \acute{a}_{ac} は、***a overlapped-by b*** の条件のうち $a_{pb} > b_{pe}$ を満たさず、 \acute{a}_{ac} ***overlapped-by b*** は「可能な関係」とはならない。つまり、 a と b もしくは a と c の関係、または、 a 、 b もしくは c の値のいずれかに誤りがあることが示唆される。

図4の場合、関係 ***a overlapped-by b*** および a ***during c*** が正しければ、 $b_{pe} > c_{pb}$ でなければ両

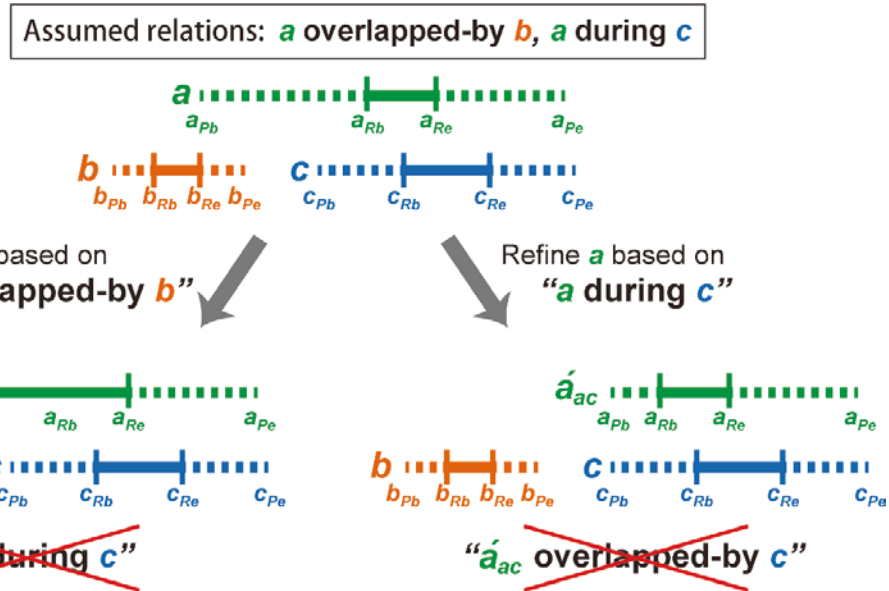


図4 精緻化操作を応用した期間同士の関係の矛盾検出。一方の関係に基づいてあいまいな期間 a を精緻化すると、もう一方の関係が成り立たなくなる。これにより、期間の値または想定された期間同士の関係に誤りがあることが分かる。

Figure 4 Detecting contradiction in relations between time intervals by applying the refinement operation. When the uncertain time interval a is refined based on one relation, the other relation cannot be satisfied. This shows that there are errors in the time intervals or relations between them.

関係は成立せず、 b または c の値に誤りがあると考えられる。一方、 a 、 b および c の値が正しく、 a **overlapped-by** b も正しければ、 a と c の関係は、**overlaps**、**contains** または **finished-by** のいずれかであるはずであり、 a **during** c とはならない。反対に、 a **during** c が正しければ、 a と b の関係は、**after** 以外あり得ない。

この精緻化操作と「可能な関係」の検証を組み合わせる矛盾を検出する手法は、文書や古記録における誤記や偽書の発見、作品の成立順序の推定など応用範囲は広い。しかしながら、すべての矛盾を検出できるのか、矛盾を引き起こしている関係や値を特定する普遍的な方法があるのか、対象となる期間が増えた場合の効率的な計算アルゴリズムを構築できるかなど、現時点では、実用化に向けたさまざまな検証、解決すべき課題も残されている。

4. 実用面での課題

4-1. 言語表現による期間同士の関係

本研究が提案するあいまいな期間の精緻化を実行するためには、期間同士の関係が Allen の関係として表現されていなければならない。一方で、通常、期間同士の関係は何らかの言語表現により表されている。したがって、「～の後に」や「～の間に」といった表現は、それぞれ Allen の **after** や **during** 関係などに置き換えなければならない。また、言語表現は Allen の時間関係ほど厳密ではない。つまり「～の後に」は、基準となるイベントの直後である **met-by** 関係を含む可能性もあるし、始点に着目して「～の後に」と表現しているのであれば、**overlapped-by** 関係を指している可能性もある。これらの違いは、文脈から読み取って適切な Allen の時間関係に置き換える、または可能性のある Allen の時間関係を最大限取り入れるなどの手順が考えられる。

「～の後に」や「～の間に」といった明示的な期間同士の関係だけでなく、暗黙的に期間同士の関係が表現される場合もある。前述の母と子の関係はその典型例であり、期間同士の関係としては示されていないものの、その関係から、それぞれの生存期間が **overlapped-by**、**during**、**finishes**、または **met-by** の 4 通りの Allen の関係（子から母を見た関係）であることが類推される。他にも、ある人物が何らかのイベントに参加したのであれば、その人物の生存期間とイベントの実施期間との関係は、一般的には **contains** 関係である（不幸にも、そのイベントが飛行機事故である場合なども想定すれば、**finished-by** などの関係も含まれる）。

言語表現から Allen の時間関係への置き換えは、定型化されたデータでは容易に実行可能である。後述のように、RDF (Resource Description Framework) では主語と目的語が期間または期間に関連付けられるリソースであれば、それらの関

係を表す述語を Allen の時間関係に置き換えれば良い。また、人名辞典のように記述されている関係が限られている場合も、比較的容易に Allen の時間関係を得ることが可能であろう。一方で、自由記述の文章を用いる場合は、その前段階として、期間およびそれらの期間同士の関係を表す言葉を抽出する仕組みが必要である。ただし、自由記述の文章には期間同士の関係につながる多様な表現が含まれている。さまざまな文章の活用は、時間情報を使った情報技術を発展させるためにも今後取り組むべき課題であり、自然言語処理分野との連携による技術の進展が不可欠である。

4-2. 期間長

期間や期間同士の関係だけでなく、「10 年間」などの期間長を表す言語表現も精緻化操作や矛盾検出の精度を向上させるために重要である。期間長の概念を導入することによって、人物の生存期間を扱う場合であれば寿命の概念を導入することが可能となり、精緻化をより進めることができる。「およそ 100 日間」や「2~3 年間」のようなあいまいな期間長については、あいまいな期間 [8] と同様の表現方法を提案した研究例があるものの [10]、精緻化操作への導入は今後の課題である。

4-3. Linked Data としてのデータの提供

本研究では、複数の期間を使って精緻化を繰り返すことにより、あいまいな期間がとり得る範囲を絞り込むことができること、また、期間同士の「可能な関係」の検証と精緻化を組み合わせることで、複数の期間同士の関係の矛盾や期間の範囲の誤りを検出できることが示された。こうした複数の期間を用いた操作は、RDF により表現された Linked Data との親和性が高い。

第一に、RDF は期間を表す主語と目的語の関係を経年同士の関係を表す述語により明確に表現することができる。Allen の時間関係を表す述語が用いられていれば、そのまま精緻化の操作を実行できるし、それ以外の述語が用いられている場合も、前述のように Allen の関係への置き換えが可能である。第二に、期間が HuTime オントロジー [11] に則って記述された場合、時間軸上の位置を直接比較できるユリウス通日 [12] の値を容易に得ることができる。これは、日付を指定した場合だけでなく、別のイベント等により階層的に期間の始点や終点を記述した場合も同様である。

一例として、以下の情報がある人物 X について考える。

- ・人物 P は人物 X の実母である。
- ・ある史料に人物 X が行事 Q に参加した記録がある。
- ・人物 R は人物 X の弟子である。

人物 P、R の生存期間および行事 Q の開催期間が

既知であるとする、それぞれの期間同士の関係は図5のとおりとなる。母子、参加、師弟などの関係は、手作業により、Allen の関係に置き換えられている(事故などを想定した一般的ではない関係は省略)。行事 Q は、その始点と終点の日付が HuTime の暦日リソース[13]により指定されており、そこから時間軸上の位置を表すユリウス通日を取得することができる。図中では示されていないが、人物 P および人物 R も同様の方法で時間軸上の位置をユリウス通日として取得することができる。

図5で示された人物 P, R および行事 Q のユリウス通日とこれらの人物 X との関係に基づいて、人物 X の生存期間の精緻化を実行することができる(図6)。人物 X の生存期間は、精緻化前には始点、終点がまったく定まっていなかったが、この精緻化により、生存期間に确实期間が現れていることを確認できる。

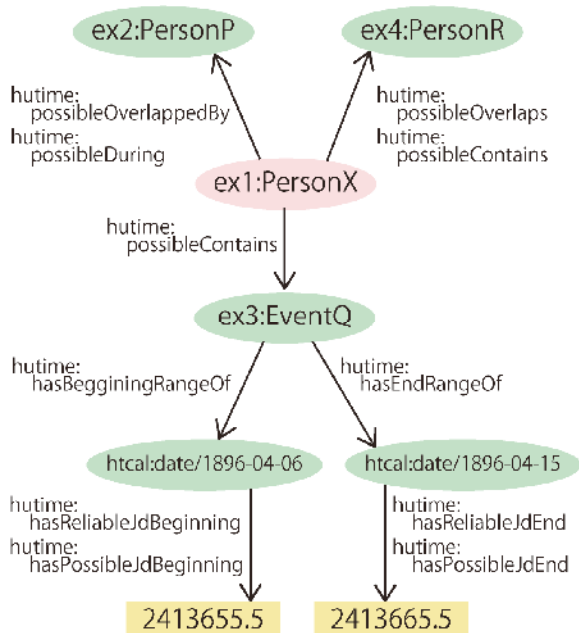


図5 期間同士の関係として表現された人物 X に関する情報。プレフィックス hutime:は http://resource.hutime.org/ontology/を、htcal:は http://datetime.hutime.org/calendar/101.1/を示す。プレフィックス ex1 から ex4 は、それぞれ異なる架空のドメインの URL を表す。人物 P および R も行事 Q と同様に暦日リソースを介してユリウス通日に紐づけられるが、図中では省略されている。

Figure 5 Information about the Person X represented as relationships between time intervals. Prefixes “hutime:” and “htcal:” represent “http://resource.hutime.org/ontology/” and “http://datetime.hutime.org/calendar/101.1/”, respectively. Prefixes “ex1” to “ex4” represent URLs of different example domains. Like Event Q, persons P and R are also linked to Julian Date values through the calendar date resources, but are omitted in this figure.

Assumed relations:
 Person X overlapped-by | during Person P
 Person X contains Event Q
 Person X overlaps | contains Person R

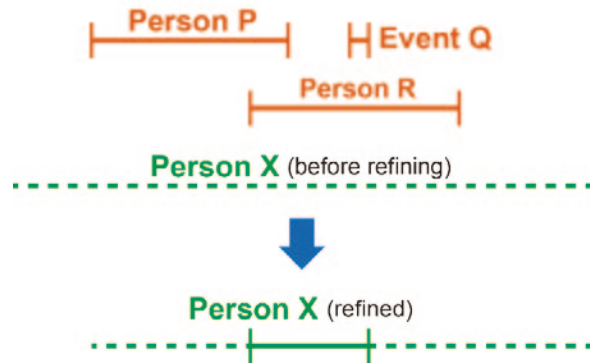


図6 人物 X の生存期間を図5に記された関連情報に基づいて精緻化する操作の概念図。精緻化前にはなかった确实期間が現れている。

Figure 6 Conceptual diagram of refining the lifespan of Person X based on the information shown in Figure 5. The reliable time interval appears that did not exist before refinement.

図5では、人物、行事を示すリソースに異なったプレフィックス (ex1 から ex4) が付されており、これらが異なるドメインから公開されているデータであることを表している。Lined Data の仕組みを活用することにより、さまざまな組織が公開するデータを使って、精緻化や矛盾検出などの解析が可能となる。さらに、それらの解析結果が公開されているデータに反映されれば、Linked Data の仕組みを通じて、それらのデータを用いた別の解析にも連鎖的に結果を反映させることが可能である。

5. まとめ

本研究では、Sekino (2021) [7]が提案したあいまいな時間データのあいまいさを低減させる「精緻化」について、その性質や定式化の過程を検証しながら計算機上で実行可能な手法として確立させた。さらに、この精緻化の操作を応用することにより、時間データの誤りやデータ間の関係の矛盾を検出する方法も新たに提案した。本研究が提案するこれらの手法は、歴史などの人文分野で特に有効である。たとえば、歴史人物や建築物の存在期間、史資料や文学作品の作成・成立年代、災害や戦乱の発生時期などの推定に用いるだけでなく、それらの時間データとしての健全性を検証することも可能になる。

これを実現するためには、精緻化の操作で参照される時間データを数多く利用できる環境が必要である。この点について本研究では自然言語による期間同士の関係の表現や Linked Data の活用など実データを扱うための課題が提示された。

本研究が提案するこれらの操作は、従来、歴史研究者などが史資料を突き合わせながら手作業で実施してきたものである。今後、利用可能な時間データが増え、これらの操作が広く活用されることにより、手作業では難しかった新たな事実の発見がもたらされると期待される。

謝辞

本研究は、JSPS 科研費 20H00017「多面的な時空間範囲の同定と記述法の開発－緯度・経度／年月日からの脱却」の助成を受けたものである。また、部分的に、JSPS 科研費 20H04381, 21H04376 の助成を受けている。

参考文献

- [1] Billiet, C. and De Tré, G.: The role of computational intelligence in temporal information retrieval: A survey of imperfect time in information systems, Troels Andreasen, T. et al.: Flexible Query Answering Systems 2015, pp. 41–56, Springer (2016).
- [2] Billiet C., Pons, J. E., Matthé, T., De Tré, G., and Capote, O. P.: Bipolar fuzzy querying of temporal databases, Christiansen, H. et al.: Flexible Query Answering Systems 2011. LNCS. Springer, pp. 66–71 (2011).
- [3] Nagypál, G. and Motik, B.: A fuzzy model for representing uncertain, subjective, and vague temporal knowledge in ontologies, Meersman, R. et al.: On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE, Springer, pp. 906–923 (2003).
- [4] Asmussen, K., Qiang, Y., De Maeyer, P., and Van de Weghe, N.: Triangular models for studying and memorising temporal knowledge, Proc. International Conference on Education, Research and Innovation, pp. 1849–1859 (2009).
- [5] Bassiri, A., Malek, M. R., Alesheikh, A. A., and Amirian, P.: Temporal relationships between rough time intervals, Gervasi, O. et al.: Computational Science and Its Applications – ICCSA 2009, Springer, pp. 543–553 (2009).
- [6] Qiang, Y., Delafontaine, M., Asmussen, K., Stichelbaut, B., De Tré, G., De Maeyer, P., and Van de Weghe, N.: Modelling imperfect time intervals in a two-dimensional space, Control and Cybernetics, Vol. 39, No. 4, pp. 983–1010 (2010).
- [7] Sekino, T: Refinement of uncertain temporal data based on relations between time intervals, Proc. 2021 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC) (2021).
- [8] Sekino, T.: Data description and retrieval using periods represented by uncertain time intervals. Journal of Information Processing, Vol.28, pp.91-99 (2020).
- [9] Allen, J.F.: Maintaining knowledge about temporal intervals. Communications of the ACM, Vol.26, No.11, pp.832–843 (1983).
- [10] Sekino, T.: Representation and comparison of uncertain temporal data based on duration. Proc. 2018 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC) (2018).
- [11] Sekino, T. HuTime Ontology to Represent Uncertain Time Intervals. Proc. Workshop “Ontologies for Linked Data in the Humanities”, Digital Humanities Conference 2019 (2019).
- [12] Dershowitz, N. and Reingold, E. Calendrical Calculations, Cambridge University Press (2007).
- [13] Sekino, T.: Basic linked data resource for temporal information, Proc. 2017 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC) (2017).