

空間配置課題における物体の向きを含む指示表現の解析

後藤 大希 西川 仁 徳永 健伸

東京工業大学 大学院情報理工学研究科

gotou.d.aa@m.titech.ac.jp, {hitoshi,take}@c.titech.ac.jp

1 はじめに

自然言語による対話中に出現する空間的な情報を扱うことは重要である。身近な例として、引っ越しの際に引っ越し業者が顧客から与えられた指示通りに家具を配置する作業を考える。このとき、以下のような対話が行われると考えられる。

顧客 冷蔵庫は流し台の隣に置いてください。

業者 これでもいいですか？

顧客 あ、こっち側に向くようにしてください。

顧客の指示に従って家具を配置するエージェントを想定すると、エージェントは与えられた指示を解釈することで物体を配置する位置や向きなどの空間的な情報を推定し配置を行う必要がある。与えられた指示に基づいて、物体を特定の方向に向けた状態で、特定の位置に配置する問題を本稿では空間配置課題と呼ぶ。

本稿では空間配置課題における指示を解析するために必要な情報を記述するアノテーションの枠組みを提案し、提案した枠組みに基づき実際の空間配置課題に関する対話に対してアノテーションを行う。さらに、対話コーパスにアノテーションした情報を用いて学習を行い、指示に対して自動でアノテーションを行う解析器を構築する。

2 関連研究

空間的な情報を扱うアノテーションの枠組みとして、Pustejovsky et al. (2011) は ISO-Space を提案した。ISO-Space は街や国土といった地理的な実体の間に存在する距離や方向などの空間的関係を記述することに加えて、それらの空間的関係の時間的な変化を記述することを目的としている。空間配置課題に関する情報を記述するために ISO-Space の枠組みは有用であるが、ISO-Space の枠組みでは地理的な実体およびそれらの空間的関係については記述できるものの、実体そのものがどの方向を向いているか記述することは想定されていないという問題点がある。詳しくは後述するが、空間配置課題を解決する手続きにおいて物体の向きに関して考慮することは重要である。本稿では ISO-Space の枠組みを物体の向きに関して記述できるように拡張することで、空間配置課題に関する情報を記述する枠組みを提案する。

空間的な情報に関するアノテーションが行われた対話コーパスとして、Tokunaga et al. (2012) による REX コーパスや Zariëb et al. (2016) による PentoRef が存在する。REX コーパスや PentoRef は参照表現の

解析に主眼を置いたコーパスであり、物体間の空間的関係に関するアノテーションが行われている。しかし、これらのコーパスでは物体の向きに関してアノテーションが行われていない。空間配置課題に関する解析を行うため、本稿では物体の位置および向きについてアノテーションを行ったコーパスを構築する。

3 空間配置課題

自然言語による命令が与えられたとき、与えられた命令に基づいて、物体を特定の方向に向けた状態で、特定の位置に配置する問題を空間配置課題と定義する。本稿では、物体の配置は二次元平面上で行われ、配置の対象となりうる物体は複数存在すると仮定する。また、物体を配置する指示を与える指示者と、指示を受け取って物体を配置する操作者との役割を分割する。空間配置課題を解決するための手続きは、以下の節に示す三つの手続きに分割できる。

3.1 物体の同定

操作者は与えられた指示に基づき、空間内に存在する物体の中から操作する物体を同定する必要がある。指示から物体を同定する問題は、マルチモーダルな状況設定における参照表現を解析する問題とみなすことができる。

3.2 位置の決定

操作者は操作対象の物体を配置する位置を決定する必要がある。これに関しても参照表現の解析を行うことで解決できると考えられるが、物体ではなく位置が参照される空間的参照表現を扱う必要がある。すなわち「三角形の隣」「ベッドの1メートル右の方」「部屋の真ん中」など、空間内の参照物体あるいは参照標識の位置を基準として対象の位置を参照する表現を解析する必要がある。

3.3 向きの決定

操作対象の物体を同定し、それを配置する位置を決定した後、操作者は物体を配置する向きを決定する必要がある。例えば「机を左の方に向けて」という指示が与えられた場合、操作者は机という物体が面する向きを決めなければならない。ここで、物体にはその物体を中心とする座標系 (intrinsic coordinate system) (Levinson, 2003) が存在する、すなわち物体には前面という部分が存在すると仮定する。このとき、操作者は物体のどの部分が前面であるかを推定し、また前面が適切な向きに面するように物体を配置する必要がある。

[小さい三角_{se1}] を [[正方形_{se2}] の [右_{sig1}]_{loc1}] に [置いて_{m1}]
 Spatial Entity (*se1*, type=TRIANGLE, mod=SMALL)
 Spatial Entity (*se2*, type=SQUARE)
 Location (*loc1*)
 Spatial Signal (*sig1*, type=DIRECTIONAL)
 Qualitative Spatial Link (*qsl1*, relType=RIGHT,
 trigger=*sig1*, figure=*se2*, ground=*loc1*)
 Motion (*m1*, motionClass=MOVE)
 Move Link (*ml1*, motion=*m1*, object=*se1*, goal=*loc1*)

図 1: 位置に関するアノテーションの例

[それ_{se3}] , [[直角_{p1}] が下になるように_{ds1}] [回して_{m2}]
 Spatial Entity (*se3*)
 Part (*p1*, partType=APEX, mod=RIGHT_ANGLE)
 Part Link (*pl1*, trigger=*p1*, source=*se3*)
 Direction Signal (*ds1*, dirType=ABSOLUTE,
 direction=LOWER)
 Direction Link (*dl1*, trigger=*ds1*, source=*p1*)
 Motion (*m2*, motionClass=ROTATE)
 Move Link (*ml2*, motion=*m2*, object=*se3*,
 dirSignal=*ds1*)

図 2: 向きに関するアノテーションの例

4 ISO-Space の拡張

2 節で述べたように, ISO-Space は空間内に存在する物体やそれらの空間的關係, 物体の動作による空間的關係の変化を記述するための枠組みである. 3 節で述べた空間配置課題を解決するための手続きに ISO-Space を用いる場合, 物体の同定および位置の決定に必要な情報は ISO-Space の枠組みに含まれるタグで記述することができる. しかしながら, ISO-Space では物体の向きに関する情報を記述することが考慮されていないため, 向きの決定に必要な情報を記述するためには ISO-Space を拡張する必要がある.

表 1 に物体の向きに関して拡張したタグおよび ISO-Space の枠組みの中で空間配置課題を解決する手続きに必要なと考えられる必要最小限のタグを示す. また, これらのタグを用いたアノテーションの例を図 1 と図 2 に示し, 3 節の内容と対応させて述べる.

4.1 物体のアノテーション

ISO-Space の枠組みに基づき, 物体の同定を行うための情報として, 物体へ言及する表現に Spatial Entity タグを付与する. Spatial Entity タグの属性として, 表現中に記述されている形容詞や対象となっている空間配置課題における物体の区分を記述する.

4.2 位置のアノテーション

ISO-Space の枠組みに基づき, 物体の位置に言及する表現に Location タグを付与し, 物体と位置の間に存在する空間的關係を示す表現に Spatial Signal タグを付与する. また, その空間的關係に関係するタグを Qualitative Spatial Link で関連付け, 属性としてその空間的關係の詳細を記述し空間配置課題における位置の決定の手続きに必要な情報をアノテーションする. さらに, 物体の動作に言及する表現に Motion タグを付与し, Motion Link タグによって操作対象の物体, 動作の内容, 物体を配置する位置を表現するタグを関連付ける.

4.3 向きのアノテーション

物体の向きに言及する表現に Direction Signal タグを付与し, その向きに面する操作対象の物体の前面に言及する表現に Part タグを付与する. 物体の前面と操作対象の物体を Part Link タグで, 物体の向きとその向きに面する物体の前面を Direction Link タグで関連付ける. これらのタグは本稿において追加したタグである. また, 物体の動作に言及する表現に Motion タグを付与し, Move Link タグによって操作対象の物体, 動作の内容, 物体を配置する向きを表現するタグを関連付ける. ISO-Space の枠組みでは Motion タグおよび Move Link タグは物体の位置に関する情報しか記述できなかったが, 本研究において物体の向きに関する情報も記述できるように拡張している.

図 2 では, 物体を配置する向きに言及する表現「直角が下になるように」に対して Direction Signal を付与し, 表現されている物体の向きの分類として属性 direction に LOWER を記述する. 「直角が下になるように」というテキストは絶対的な座標系に基づいた表現であるため, 物体の向きの型を ABSOLUTE と記述する. また, 「直角が下になるように」という向きに面する物体の一部分として「直角」に言及しているため, 物体の一部分を示す表現に Part タグを付与する. その物体の一部分が対象の空間配置課題においてどの区分に該当するかを Part タグの属性 partType に記述し, その表現に形容詞が含まれていれば属性 mod に記述する.

4.4 対話コーパスへのアノテーション

拡張した枠組みを評価するため, 空間配置課題に関する対話を収集した REX コーパス (Tokunaga et al., 2012) にアノテーションを行う. REX コーパスは, 二人の実験参加者が指示者と操作者に分かれ, 画面上のタングラム・パズルを解く実験中の発話を収集した対話である. 指示者と操作者は別の画面を見ており, ピースを組み立てる作業スペースは指示者と操作者の両者の画面で共有されている. 指示者はパズルの完成形を見ることができがパズルのピースを操作できず, 操作者はピースを操作できるが完成形を見られない. パズルを完成させるためには指示者が操作者にどのようにピースを配置するかを指示を与える必要があるため, この対話は空間配置課題を解決するための対話であるといえる.

指示表現の解析を行うため, 対話に含まれる指示者の発話をアノテーションの対象とする. 総対話数は 51, 指示表現を含む指示者の発話の数は 766 であった. その中でも物体の向きに関して言及している指示表現の数は 145 であり, これは全体の指示表現の 18.9% にあたる.

アノテーションを行った結果のタグの分布を表 2 に示す. 表に示した結果から, 空間的關係を示す Spatial Signal タグと物体の向きを示す Direction Signal タグの数が拮抗していることがわかる. これらのことから, 空間配置課題において物体の向きに関する情報は重要であるといえる.

表 1: アノテーションするタグ

テキストに付与されるタグ	概要
Spatial Entity	本質的に位置でなく、空間的關係によって同定される実体
Motion	ある物体の位置 または向き が変化する動作
Location	本質的に位置である、空間上に存在する実体
Spatial Signal	空間的關係に言及する表現
Direction Signal	ある物体の向きに言及する表現
Part	物体の一部分

実体間の関係を記述するタグ	概要
Qualitative Spatial Link	二つの物体の空間的關係
Move Link	動作によって位置 または向き が変化する物体と、その終着点となる位置または向きの關係
Part Link	ある物体と、その物体の一部分の關係
Direction Link	ある物体の向きと、その向きに面する物体の一部分の關係

* 拡張した要素には下線を引いている。

表 2: タグの分布

テキストに付与されるタグ	タグ数
Spatial Entity	722
Motion	744
Location	235
Spatial Signal	168
Direction Signal	144
Part	136

実体間の関係を記述するタグ	タグ数
Qualitative Spatial Link	472
Move Link	2,352
Part Link	186
Direction Link	199

* 拡張した要素には下線を引いている。

表 3: チャンクの抽出結果

タグ	CRF++			ルール			タグ数
	P	R	F	P	R	F	
Spatial Entity	.91	.88	.90	.38	.70	.49	722
Motion	1.0	.99	1.0	.88	.93	.90	744
Location	.89	.75	.81	.26	.68	.37	235
Part	.94	.78	.85	.40	.50	.44	136
Spatial Signal	.81	.65	.72	.44	.65	.52	168
Direction Signal	.98	.99	.99	.46	.41	.44	144

系列ラベリングの学習を行う素性としては、系列中のある形態素に対してその形態素とその前後各 2 形態素の表層、品詞、ひらがな・カタカナ・アルファベット (大文字/小文字)・数字・空白・その他の文字種、および前 2 形態素に割り当てられた IOB2 タグを用いる。また、これらの素性を用いた規則に基づいてタグの付与を行うルールベースのシステムを別途実装し、系列ラベリングを行う際のベースラインとして評価を行う。

5.2 チャンク間の関係抽出

チャンク間に存在する関係を決定するため、5.1 節の手続きによってタグが付与されたチャンクの組について、その組の間に存在する関係を表 1 中の関係を記述するタグのいずれかに分類する。

関係抽出の手法として、テキスト中のチャンクの組を全て抽出しそれらの関係を表 1 の関係に分類する手法と、Kim et al. (2016) によるテキスト中で関係の原因となりうるチャンクとそれ以外のチャンクの組だけを抽出しそれらの関係を表 1 の関係に分類する手法を用いる。前者を全抽出、後者を部分抽出と呼び、それぞれのシステムに 4.4 節で人手によりテキストにアノテーションを行ったコーパス (人手付与) と 5.1 節の手法により機械的にアノテーションを行ったコーパス (機械付与) を与え、スコアを比較する。全抽出においてチャンク間の関係が表 1 の関係のいずれにも当てはまらない場合、No Link と分類する。

表 5 に関係抽出に用いる素性を示す。素性の学習と分類は LinearSVC³ を用いて行う。

5 解析器の構築

提案した枠組みを用いたタグの自動付与システムの実現可能性を調べるため、表 1 に示したタグをテキストに割り当てるシステムを実装する。指示文が入力として与えられたとき、このシステムが行うタスクは以下の二つに大きく分けられる。

1. 入力文から命令の内容に関する表現を抽出し、表 1 のテキストに付与されるタグを割り当てる。このとき、入力文が区切られた意味的なまとまりをチャンクと呼ぶ。
2. タグが割り当てられた二つのチャンクの組に対して、それらの関係が表 1 の実体間の関係のうちどれに区分されるか分類する。

5.1 テキストへのタグの付与

入力として与えられた文からチャンクを抽出しタグを付与する手続きを、形態素の系列に Tjong et al. (1999) が提案した IOB2 タグを付与する系列ラベリングの問題として考える。入力テキストの形態素解析には日本語形態素解析器 MeCab¹ を用いて、系列ラベリングの学習と推定には CRF++² を用いる。

¹<http://taku910.github.io/mecab/>

²<https://taku910.github.io/crfpp/>

³<http://scikit-learn.org/stable/>

表 4: チャンク間の関係抽出の結果

関係	全抽出+人手付与			全抽出+機械付与			部分抽出+人手付与			部分抽出+機械付与			タグ数
	P	R	F	P	R	F	P	R	F	P	R	F	
Q. S. Link	0.31	0.30	0.31	0.45	0.20	0.29	0.83	1.00	0.91	0.89	0.56	0.69	472
Move Link	0.42	0.47	0.44	0.40	0.34	0.37	0.94	0.81	0.87	0.94	0.79	0.86	2,595
Part Link	0.33	0.59	0.42	0.34	0.44	0.38	0.93	0.77	0.84	0.96	0.57	0.71	228
Direction Link	0.38	0.50	0.43	0.33	0.41	0.37	0.98	0.98	0.98	0.99	0.95	0.97	202
No Link	0.68	0.62	0.65	0.73	0.77	0.75	-	-	-	-	-	-	5,913

表 5: 関係抽出に用いる素性

素性	素性の概要
tag pair	チャンクに付与されたタグの組
distance	二つのチャンク間の日本語における文字数
length	チャンクが含まれる命令文の文字数
N of spans	命令文に含まれるチャンクの数
pos	各チャンクの前後に出現する品詞の四つ組
case	各チャンクの直後に出現する助詞の格の組
dep	二つのチャンク間の係り受け数

5.3 結果と考察

タグのラベリング, 関係抽出の学習データとして 4.4 節のアノテーションを使い, 10 分割交差検定を行った. 表 3 にチャンク抽出を行った際の精度 (P), 再現率 (R), F 値 (F) を示す.

表 3 に示す通り, CRF++ を用いたチャンク抽出では高いスコアを記録している. この要因として, アノテーションを行った対話コーパスのドメインが極めて狭かったことが考えられる. 同じ枠組みをさらに広いドメインに適用した場合に, 今回と同様に高いスコアが記録されるかどうかを調査する必要がある.

チャンク抽出において誤りが起こる原因は大きく二つに分けられる. 一つは不十分なアノテーションのガイドラインによるもの, もう一つは話し言葉の形態素解析の誤りによるものである. 前者による誤りの数は 123, 後者による誤りの数は 239 であった.

4.4 節で行ったアノテーションでは, Spatial Entity タグを付与する際に空間的關係の中で物体に言及されている場合に限り記述を行っていた. しかし, 解析器は空間的關係の有無によらず物体へ言及する全ての表現に対してタグを付与したため, 偽陽性の誤りが生じていた. これを改善するためには, 他の物体との空間的關係の有無によらず物体へ言及する表現全てに Spatial Entity タグのアノテーションを行うようアノテーションのガイドラインを変更する必要がある.

表 4 に関係抽出の結果を示す. 各関係が引数として取りうるチャンクの種類は限定されるため, 部分抽出のシステムのスコアが高くなっている. 今回の枠組みにおいては空間配置課題を解決するために必要最小限のタグを用いているため, 表 1 に示したタグ以外の ISO-Space に含まれるタグを用いた場合にはよりスコアが低くなると考えられる. それらのタグを追加し, さらに大きいサイズのコーパスが広いドメインを対象としたコーパスに対する実験が必要である.

また, 部分抽出+人手付与のスコアと部分抽出+機械付与のスコアを比較すると, 特に Qualitative Spatial Link および Part Link の関係抽出で再現率が下がっている. これは表 3 で示したとおり Spatial Signal,

Part チャンクの抽出精度の低さに起因すると考えられる. Kim らの手法では, 関係の原因となるチャンクが正しく抽出されていない場合, そのチャンクを原因とする関係の関係抽出が行われない. Spatial Signal タグまたは Part タグが付与されるチャンクが正しく抽出されなかったことによりそれをトリガとする関係の抽出が行われず, 偽陰性の誤りが増加し再現率が低下したと考えられる. チャンク抽出の誤りが原因であるため, 再現率を上げるためには関係の原因となりうるチャンクを抽出する精度を改善する必要がある.

6 結論と今後の課題

本稿では, 自然言語理解における空間配置課題を定義し, 物体の位置および物体の向き両方に関する情報を記述できるアノテーションの枠組みを提案した. また, 実際にアノテーションを行って物体の向きに関する情報は無視できないものであることを示すとともにタグを機械的に付与する解析器を構築しその性能を評価した.

本稿では自然言語による指示をドメインとしたが, より現実的な課題を扱うためには視線や身振りなどのマルチモーダルな要素を考慮する必要がある. 非言語的な情報が空間配置課題の解決に与える影響の調査を行うことが今後の研究課題として考えられる.

参考文献

- James Pustejovsky, Jessica L. Moszkowicz, Marc Verhagen. Using ISO-Space for annotating spatial information. In *Proceedings of the International Conference on Spatial Information Theory*, 2011.
- Takenobu Tokunaga, Ryu Iida, Asuka Terai, Naoko Kuriyama. The REX corpora: A collection of multimodal corpora of referring expressions in collaborative problem solving dialogues. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, pp.422-429, 2012.
- Sina Zarriß and Julian Hough and Casey Kennington and Ramesh Manuvinakurike and David DeVault and Raquel Fernandez and David Schlangen. PentoRef: A corpus of spoken references in task-oriented dialogues. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pp.125-131, 2016.
- Levinson, Stephen C. *Space in language and cognition: Explorations in cognitive diversity*. Cambridge University Press, 2003.
- Bogyum Kim and Jae Sung Lee. Extracting Spatial Entities and Relations in Korean Text. In *Proceedings of Twenty-Sixth Conference on Computational Linguistics (COLING 2016)*, pp. 2389-2396, 2016.
- Erik F. Tjong and Kim Sang and Jorn Veenstra. Representing text chunks. In *Proceedings of 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL 1999)*, pp. 1730179, 1999.