

3D環境を利用したマルチモーダル対話システムにおける 領域オントロジー

A Domain Ontology for Multimodal Dialogue System using 3D Environment

スホルクワイク 亜蘭[†] 谷津 元樹[‡] 森田 武史[‡]
Alan Schalkwijk[†] Motoki Yatsu[‡] Takeshi Morita[‡]

[†] 青山学院大学大学院 理工学研究科

[‡] 青山学院大学 理工学部

[†] Graduate School of Science and Engineering, Aoyama Gakuin University.

[‡] College of Science and Engineering, Aoyama Gakuin University.

要旨

近年, Embodied AI などの現実世界や 3D 環境との相互作用を通じて学習を行い, 仮想ロボットを用いて 3D 環境で様々なタスクを行うことができるシステムの研究が行われている. しかし, これらにはユーザに対する質問や要求に答えるだけで対話が途切れてしまう一方のタスクが多い. 本研究では 3D の家庭内環境で, 仮想ロボットがユーザとの雑談を行いながら家を案内することができるタスク指向対話型システムを開発することを目的とする. 本発表では, 本目的を達成するために必要な領域オントロジーの構築について述べる.

1. はじめに

近年 Embodied AI という, 現実世界や 3D 環境で学習を行い, 仮想ロボットなどのエージェントを用いて, 目的の物や部屋に案内するオブジェクトナビゲーションや質問応答といった, 視覚認識や言語理解など幅広いスキルが要求されるタスクを行うことができる研究がみられる. 今まで視覚認識や言語理解は画像やテキストのパターン認識に焦点を置いた ‘Internet AI’ を用いて発展されてきたが, Embodied AI は環境内でロボットによるアクションができることに焦点を置いた新しい試みとなっている. 従来の Embodied AI には 3D 環境を利用した質問応答などのタスクが存在するが, ユーザの入力に仮想ロボットが応答するだけの一方の対話を行うものとなっている. 研究では, 仮想ロボットのセンシングした画像とユーザからのテキスト発話を用いて仮想ロボット自身からも 3D 環境に関する雑談を行うことによってユーザとロボットとの双方向の対話が可能となるマルチモーダル対話システムの開発を目指す.

2. 関連研究

AI Habitat[1]とは Embodied AI の研究をサポートするために開発されたシミュレーションプラットフォームのことであり, 3D 環境における仮想ロボットの学習を可能としている. Embodied AI のロボットを複雑な現実世界で直接学習をさせると, ロボットが学習することに適した環境の用意や, 学習に非常に時間がかかるといった様々な問題が生じる. そこで, Embodied AI を用いて 3D 環境を用いた現実的なシミュレーターで学習を行い, 学習した仮想ロボットのタスク能力を現実世界に移行することを可能とする. 本研究では, AI Habitat を用いて, 3D 環境及び仮想ロボットの構築を行う.

KnowRob[2]とはロボットのための知識処理システムのことであり, KnowRob では, タスクの詳細や物体などの多くの知識を領域オントロジーとして構築し, 様々な情報リソースを統合可能なセマンティックフレームワークを提供している. ロボットが行うタスクの中に, 異なる表現や語彙が含まれている場合, KnowRob のオントロジーを用いて, それを共通の表現になるように変換し, ロボットが意図したタスクを実行することができる. KnowRob では, 主に家庭内におけるタスクをロボットが実行することを目的としており, オントロジーには家具や部屋の名前といった, 家庭内に関する知識が多く含まれてい

る。本研究では、KnowRob オントロジーを基に、家庭内の3D環境に特化した3D環境オントロジーを構築し、それを対話に用いる。

3. 3D環境を利用した対話システム

本研究では、3D環境内に存在する家庭用対話ロボットに雑談をしながら、家を案内してもらう状況に特化したタスク指向型対話システムの開発を目的としている。対話方法としてはユーザからロボットに雑談や質問を行い、ユーザの対話内容に応じてロボットが家を案内しながら、対話ルールに基づいて応答するという流れになる。例えば、ユーザが「キッチンがどこにある？」という質問をした場合、ロボットは「こちらです。」と応答し、ユーザをキッチンに案内する。ロボットがキッチンにいることを認識した場合は、「好きな料理はありますか？」といった雑談をユーザに行い、双方向な対話を行う。このような対話を行う上で、ユーザとロボットは家庭内を共に行動していて、同じ景色（ロボットがカメラによってセンシングした画像）を見ているものとする。

本システムを実現するにあたって、3D環境と3D環境内における仮想ロボットの構築には関連研究で述べたHabitat Aiを用いる。3D環境の構築に必要な3DデータセットにはMatterPort3D[3]を用いる。MatterPort3Dには3Dの家庭内に存在する部屋や物に関するデータだけではなく、その部屋や物の名前や位置情報がセマンティックデータとして用意されている。3D環境に関するオントロジーをKnowRobで提供されているオントロジーを用いて構築し、そこからセマンティックデータに存在する家具や部屋の知識を参照することで、家庭内の環境に特化した対話を行う。

図1に本研究における対話システムの構成図を示す。AI HabitatとMatterPort3Dを用いて3D環境の構築および仮想ロボットの運用を行う。3D環境構築の際、MatterPort3Dに存在するセマンティックデータをKnowRobオントロジーに加え、3D環境オントロジーを構築する。3D環境オントロジーを対話システムが利用することにより、ユーザと仮想ロボットとの対話を支援する。対話システムの入力にはユーザの発話テキストと仮想ロボットがセンシングした画像を用いる。表2にMatterPort3Dにおけるセマンティックデータの一部を示す。表2においてcategoryは3D環境におけるオブジェクトの名前を、centerはx, y, zによるオブジェクトの中心位置座標を表す。図3に仮想ロボットが3D環境をセンシングしたRGB画像の例を示す。図3中のソファや照明などのオブジェクトには、表2に示すセマンティックデータが与えられている。

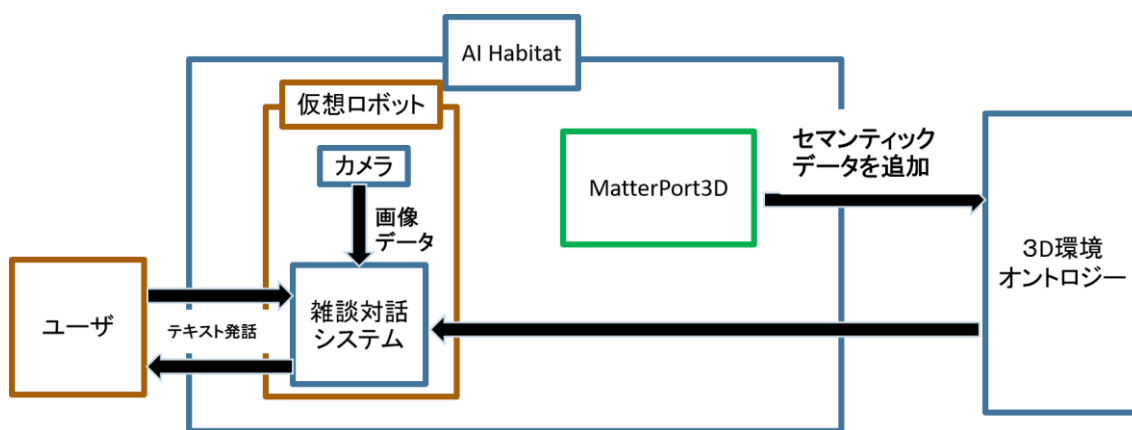


図1 本研究における対話システムの構成図

表2 MatterPort3Dにおけるセマンティックデータの一部

category: lounge	center: [0.619232 -1.02348 3.102932]
category: sofa	center: [0.738291 -1.845729 -8.5748394]
category: wall	center: [-0.217584 1.564728 5.485734]



図3 仮想ロボットが3D環境をセンシングしたRGB画像の例

4. 3D環境オントロジー

図4に3D環境オントロジーの構築手順を示す。図4の手順1では、MatterPort3Dに存在するすべてのセマンティックデータの部屋や物の名前を、オントロジー操作ライブラリである owlready2¹を用いて KnowRob オントロジーにおけるクラス名と照合を行い、KnowRob オントロジーに存在しないセマンティックデータに関するクラスを同定する。照合を行った結果、KnowRob オントロジーにはセマンティックデータに関する48個のクラスが存在しないことが判明した。KnowRob オントロジーにおけるクラス名と照合できなかったセマンティックデータにおけるオブジェクト名を元に、WordNet[4]と DODDLE-OWL[5]を用いてクラス階層の自動構築を行い、人手で KnowRob のオントロジーのクラス階層と自動構築したクラス階層を統合する。

図4の手順2では日本語での対話を行うために、追加したクラスや家具、部屋や料理といった家庭内における対話に用いられる可能性の高いクラスに優先的に日本語ラベルを追加した。図5に KnowRob オントロジーのクラス階層と、3D環境オントロジーにおけるクラス階層の一部を示す。図5の破線部分では足りないクラスと日本語ラベルを、実線部分では既存のクラスに日本語ラベルを追加している。

図4の手順3では、ロボットがユーザを案内するために必要な部屋と物の位置情報をオントロジーに登録するために、表2のようにセマンティックデータに位置情報として存在している x, y, z 座標を表すプロパティ（以下、座標プロパティ）をそれぞれ追加した。このプロパティを利用して、3D環境構築のために MatterPort3D のデータセットの1つを読み込んだ際、セマンティックデータに複数存在するオブジェクトのインスタンスを作成する。

図4の手順4では、作成したインスタンスに座標プロパティとその値を定義する。図6に物のインスタンスにおける座標プロパティの定義例の一部を示す。図6①は作成した家具のインスタンスの一部を、図6②はインスタンスの座標プロパティの定義例をそれぞれ示す。作成したインスタンスから座標プロパティ値を得ることによってロボットが場所を把握し、案内することが可能となる。

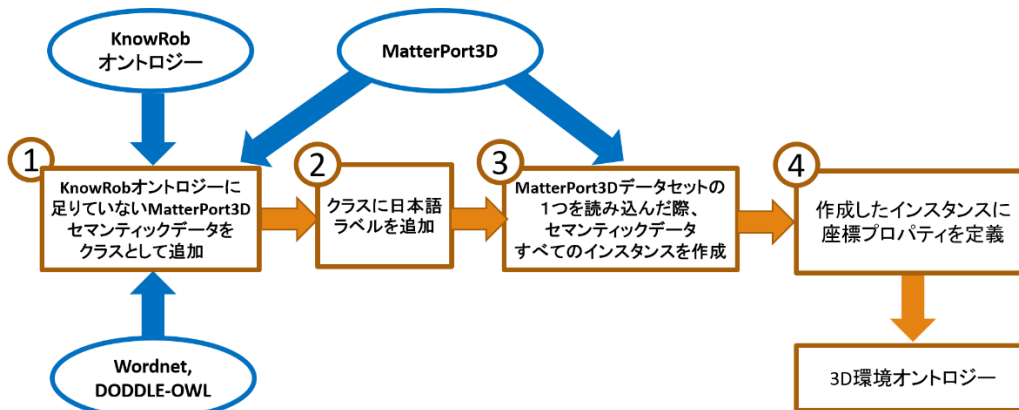


図4 3D環境オントロジー構築手順

¹ <https://owlready2.readthedocs.io/en/latest/intro.html>

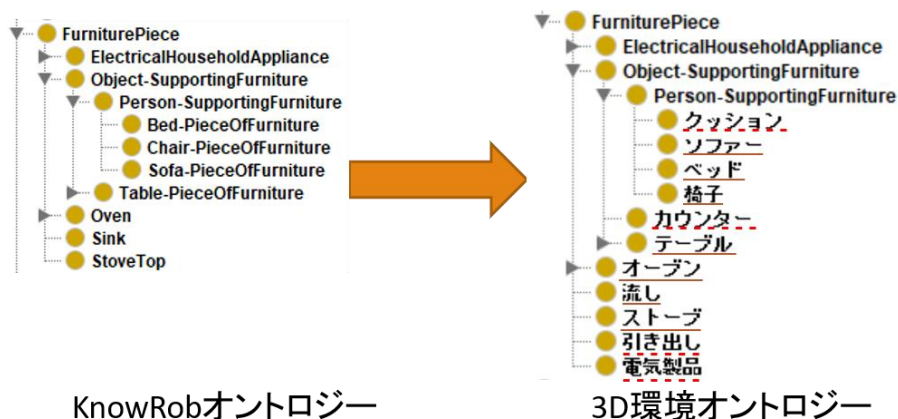


図5 Knowrob オントロジーと 3D 環境オントロジーにおけるクラス階層の一部

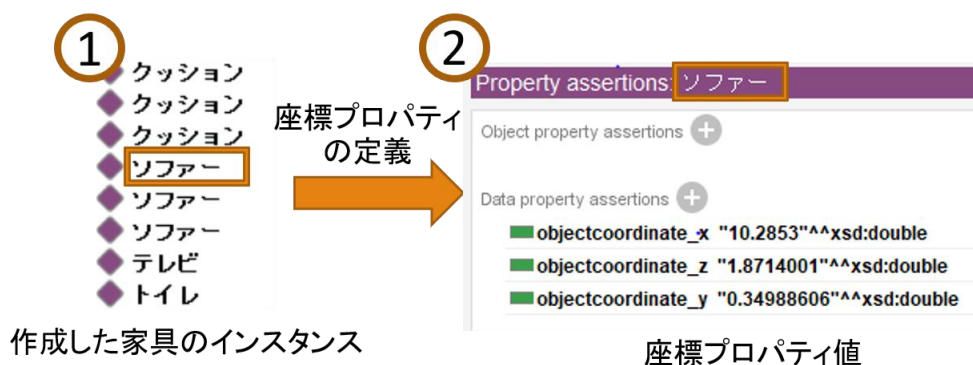


図6 物と部屋のインスタンスにおける座標プロパティの定義例

5. まとめ

本研究が目指している対話システムと、3D 環境内で家庭内に関する対話を行うために必要な 3D 環境オントロジーの構築の手法について述べた。今後、3D 環境オントロジーを用いて図1で示した対話システムの実現を目指す。

参考文献

- [1] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao¹, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, Dhruv Batra¹. **Habitat: A Platform for Embodied AI Research**. ICCV2019.
- [2] Moritz Tenorth, Michael Beetz, "Representations for robot knowledge in the KnowRob framework", In: Artificial Intelligence, Elsevier, 2015.
- [3] Chang, Angel and Dai, Angela and Funkhouser, Thomas and Halber, Maciej and Niessner, Matthias and Savva, Manolis and Song, Shuran and Zeng, Andy and Zhang, Yinda . Matterport3D. 3DV2017
- [4] Fellbaum, Christiane (2005). WordNet and wordnets. In: Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, Second Edition, Oxford: Elsevier, 665-670.
- [5] T. Morita, N. Fukuta, N. Izumi, T. Yamaguchi: DODDLE-OWL: Interactive Domain Ontology Development with Open Source Software in Java, IEICE Transactions on Information and Systems, Special Section on Knowledge-Based Software Engineering, Vol. E91-D No. 4 pp. 945-958 (2008. 4) DOI: [10.1093/ietisy/e91-d.4.945](https://doi.org/10.1093/ietisy/e91-d.4.945)