

家庭環境オントロジーと常識推論に基づく双方向型 対話ナビゲーションシステムの開発

Development of an interactive navigation system based on household
ontology and commonsense reasoning

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

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

[‡] 青山学院大学 理工学部

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

Abstract: In recent years, researchers from the fields of computer vision, language, graphics, and robotics have tackled Embodied AI research. Embodied AI can learn through interaction with the real world and virtual environments and can perform various tasks in virtual environments using virtual robots. However, many of these are one-way tasks in which the interaction is interrupted only by answering questions or requests to the user. In this study, we aim to develop a task-oriented interactive system using virtual household ontology and commonsense reasoning, in which a virtual robot can reason about the location of a guide while interacting with the user and guide the user around the house

1 はじめに

近年, Embodied AI[8]と呼ばれる, 現実世界や仮想空間で学習を行い, 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仮想空間における質問応答などのタスクの解決を試みる研究が行われている. これらのタスクを解決するためには, コンピュータビジョン, コンピュータグラフィックス, 自然言語処理, 人工知能, ロボティクスなど, 様々な分野の技術を統合する必要がある.

従来のコンピュータビジョンや自然言語処理の研究は, 画像やテキストのパターン認識に焦点を置いた‘Internet AI’により発展してきた. Embodied AI は, 仮想空間内で仮想ロボットが見る, 聞く, 話す, 移動するなどの動作ができることに焦点を置いている.

従来の Embodied AI 研究におけるオブジェクトナビゲーションや質問応答などのタスクは, ユーザからの具体的な要求に仮想ロボットが応答するだけの一方向の対話で解決できるものが多い. ユーザからの曖昧な要求に仮想ロボットが対応できるようにするためには, ユーザと仮想ロボットが複数回の対話を行い, ユーザの意図を推論する必要があると考えら

れる.

本研究では, Embodied AI 研究に知識推論の技術を導入し, 双方型対話ナビゲーションシステムを開発することを目的とする. 具体的には, 家庭環境オントロジーと常識推論に基づいて, 仮想ロボットがユーザと複数回の対話を繰り返しながらユーザの意図を推論し, ユーザが必要とする仮想家庭環境における場所を案内することが可能なタスクの解決を試みる.

2 関連研究

2.1 Vision and Dialog Navigation

Vision and Dialog Navigation (VDN)[1]とは人間が家庭環境で行われるような不特定な指示を行い, それに対して仮想ロボットが回答し, 案内するタスクのことである. VDNでは, 言語遠隔操作の家庭ロボットやオフィス用のロボットのように, なれない場所でどこに行けばわからない場合に, 人間の自動案内を可能とする仮想ロボットの学習を目的としている. また, VDNは Matter Port Room-2Room という写実的なシミュレーション環境を用いて, 案内に関する対話例とそれに対する仮想ロボットの行動(前進, 旋回

のデータセットを作成し、学習およびシミュレーションを行っている。

このタスクにおける主な対話例は“Go right”といった直接的な行動よりも“Go to the room with the bed”といった曖昧なものを対象としており、仮想ロボットがとらえている画像とユーザからの曖昧な発話から、案内場所への移動を行う。本研究では、VDNで想定されている対話例よりも、「お腹がすいた」といったより曖昧でかつ日常的に家庭内で行われるような発話から案内場所を推定し、案内することを目的としている。

2.2 常識知識に関する研究

常識知識に関する研究として COMET-ATOMIC[2]が挙げられる。COMET-ATOMICとは133万の英語で表現されたエンティティと事象に関する日常的な推論知識のタプルを持った常識的知識グラフのことである。

COMET-ATOMICは23種類の常識的関係に

基づいて、日常に関する事象や物体、さらにある事象に対する人間の行動や精神状態などの推論を可能とする。この知識グラフを用いて事象や物体と常識的関係の2つを入力として推論を行う言語モデルも作成している。

本研究では、この学習済みの言語モデルを用いてユーザの発話から案内場所の推論を行う。本研究では、「ある事象に対して人間が引き起こす事象」、「ある事象に対して人間が必要となる行動」、「ある物体が発見される可能性がある場所」の3種類の常識的関係を用いる。

3 提案システム

本節では、提案システムの構成、家庭環境オントロジーの構築手順、常識推論に基づく対話について述べる。

提案システムでは、ユーザからシステムに発話文の入力を行い、ユーザの発話内容から対話ルールに基づいて、仮想ロボットが応答し、仮想家庭環境を案内する。対話を行う上

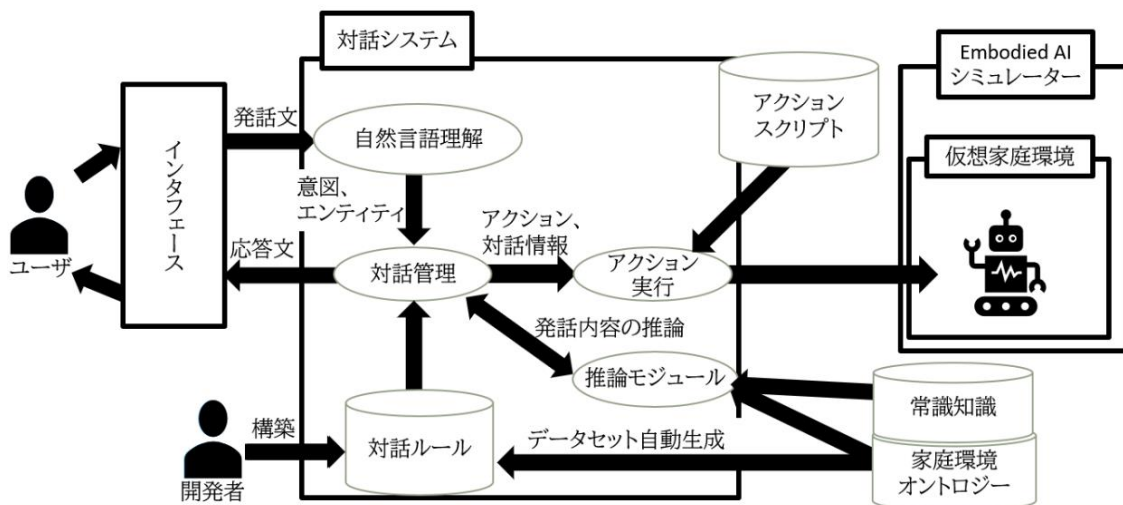


図1 システム構成

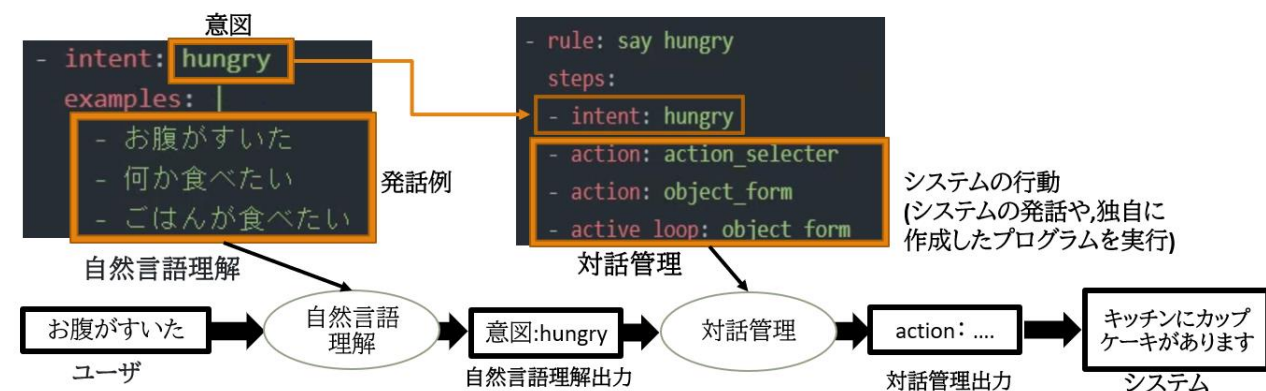


図2 対話システムの入出力例

で、ユーザと仮想ロボットは仮想家庭環境内を共に行動しており、同じ景色（仮想ロボットがカメラによってセンシングした仮想環境の画像）を見ているものとする。

3.1 システム構成

本研究におけるシステム構成図を図1に示す。最初にユーザとシステムとの会話はインタフェースを通してテキスト形式でやり取りされる。ユーザから発話文の入力が行われると、対話システムの中にある自然言語理解モジュールを通じて発話文の意図とエンティティが抽出される。その後、対話管理のモジュールにて、意図とエンティティと対話ルールに基づいてシステムの応答が決定する。また、対話ルールに基づいて仮想ロボットが案内を行う場合は、仮想ロボットを動かすためのアクションスクリプトを指定し、それを元に Embodied AI シミュレータ上でアクションが実行される。また、ユーザの発話内容が曖昧で案内場所の特定が困難な場合は、家庭環境オントロジーと常識推論を活用した推論モジュールを利用して発話内容から案内場所の推論を行う。

3.2 インタフェース

本研究ではインタフェースに Telegram¹を用いる。Telegram はインスタントメッセージアプリケーションであり、テキスト入力による会話を行うことができる。Telegram は、導入しやすく多くの人に利用されている点や、本研究に用いる対話フレームワークである Rasa [5] と連携が可能であるという点で、本研究の対話インタフェースとして利用する。

3.3 対話システム

本対話システムの構築に、タスク指向型対話システムの開発を目的としたオープンソースの対話フレームワーク Rasa [5] を用いる。

Rasa では図2のようにユーザが作成したある意図に対応する対話例をデータセットとして用意し、学習を行うことによって、ユーザの入力から意図やエンティティの抽出を行う

[Walk] <living_room> (1)

図3 アクションスクリプト例

ことが可能である。そして、対話ルールに基づいて対話を行う。対話ルールではユーザの入力から抽出した意図に対して、システムの出力内容などを決定することができる。さらに、対話ルールに基づいて Python で記述した特定のプログラムを実行することができる。

次に Rasa の選定理由について述べる。最新の対話システムのフレームワークに関する論文 [7, 9] を調査し、Rasa とその他のフレームワークとの比較を行った。その結果、Rasa は以下の点で優れていることが分かった。

- ・ ユーザの入力文から意図やエンティティの抽出が可能
- ・ 対話ルールやスロットに関する設定が可能
- ・ 対話ルールに基づいて独自に作成したプログラムを実行可能
- ・ 既存のインタフェースとの連携が可能
- ・ 開発規模が大きく、フレームワークに関するドキュメントが豊富であり、利用しやすい点、

以上より、本研究の対話システムの開発 Rasa を利用することにした。

3.4 Embodied AI シミュレータ

Embodied AI シミュレータには Virtual Home [6] を用いる。VirtualHome は仮想家庭内での活動をシミュレートすることを目的としたシミュレータである。VirtualHome の環境は知識グラフとして表されており、部屋や物とそれらに関する識別番号、位置座標、物の状態といったセマンティックデータがデータ名をキーとしそれに対応した値で表現された辞書形式で含まれている。本研究ではこの環境知識を用いて家庭環境オントロジーの構築を行う。

VirtualHome ではアクションスクリプトを用いて、特定のタスクを行うことが可能である。アクションスクリプトは図3のように仮想ロボットの行動内容、行動に移す対象の部屋や物、部屋や物の識別番号で構成されている。本研究において案内などの行う場合はこのアクションスクリプトを利用する。

最新の Embodied AI シミュレータに関する調査論文 [8] を参考に、VirtualHome と他の Embodied AI シミュレータとの比較を行った。その結果、環境内の物に関するセマンテ

¹ <https://telegram.org>

ティックデータが豊富、仮想ロボットによる特定の部屋や物への移動が容易である、仮想家庭環境を提供している、この3つの点において優れており、本研究に適していると考え、VirtualHome を利用することにした。

3.5 家庭環境オントロジー

図4に家庭環境オントロジーの構築手順を示す。図4の手順1では、VirtualHomeの知識グラフ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抽出する。手順2では、抽出したデータを DODDLE-OWL[3] と WordNet[4]を用いてオントロジーのクラスとして追加する。手順3では日本語の対話を行うために、追加したクラスに日本語ラベルを

追加した。次に手順4で、セマンティックデータすべてを追加したクラスのインスタンスとして生成する。手順5では作成したインスタンスにセマンティックデータである位置座標、識別番号、状態を格納するためのプロパティを定義する。さらに、案内を行いやすくするため、ある物がどこの部屋に存在しているかを表すためのプロパティを追加した。最終的に作成した家庭環境オントロジーのクラス数は239個、プロパティ数は6個、インスタンス数は356個となった。

3.6 発話内容の推論

家庭環境オントロジーと常識推論に基づくユーザの発話から場所案内までの流れを図5

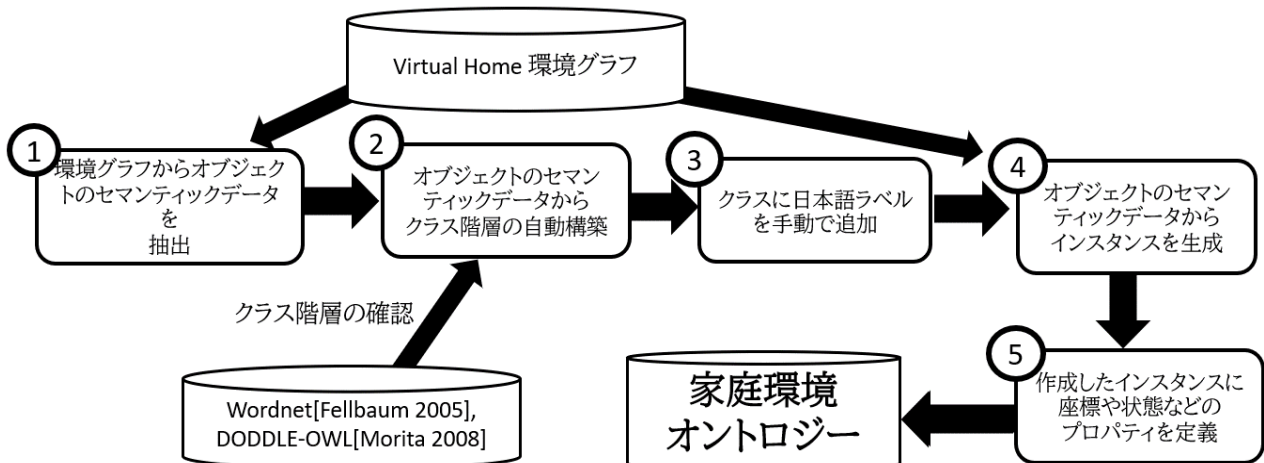


図4 家庭環境オントロジー構築手順

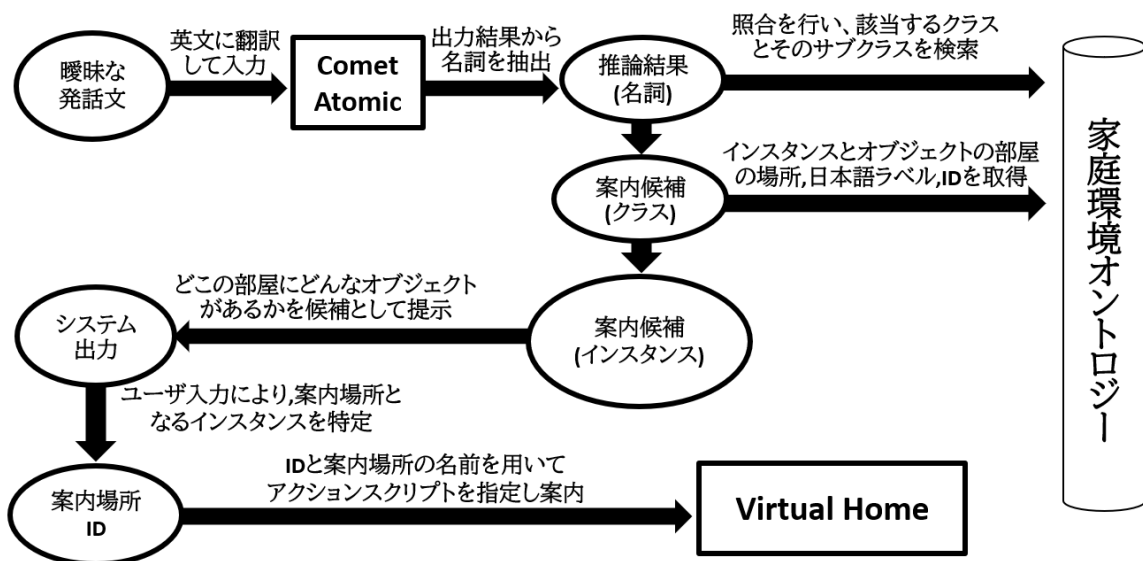


図5 家庭環境オントロジーと常識推論に基づくユーザの発話から場所案内の流れ

と具体例を用いて述べる。

ユーザから「お腹がすいた」といった曖昧な発話文が入力された場合、関連研究で紹介した COMET-ATOMIC を用いて推論を行うために、“I am hungry”と英文に翻訳する。次に翻訳した英文を COMET-ATOMIC に入力し、「ある事象に対して人間が引き起こす事象」、「ある事象に対して人間が必要となる行動」、「ある物体が発見される可能性がある場所」の3種類の常識的關係を用いてユーザの行きたい場所や行動などを推論し出力する。その結果、“go to kitchen”, “eat s food”といった出力結果が得られる。そして、出力結果から名詞の抽出を行い、“kitchen”や“food”などユーザの行きたい場所や欲しい物の名前を取得する。取得した名前を家庭環境オントロジーにおけるクラスとサブクラスと照合する。照合したクラスのインスタンスからその部屋の場所に関するプロパティ値と日本語ラベルを取得し、システムが「キッチンにカップケーキがあります」といった、ユーザにとどここの部屋にどんな物があるかを候補として提示する。そしてユーザがシステムとやり取りを行い、案内場所を絞る。案内場所が決定した場合は、そのインスタンスから識別番号を取得してアクションスクリプトを用いて案内を行う。

このようにして、ユーザの曖昧な発話の推論結果から、いくつか案内候補を提示し、さ

らにそこからユーザが選択することにより、双方向な対話を活用した案内を行うことが可能となる。

4 提案システムの実行例

Telegram を用いた提案システムの対話例を図 6 に示す。対話の中で「お腹がすきました」という入力が行われた場合、Rasa の自然言語理解コンポーネントによって入力文から、ユーザが空腹であるという意図を抽出する。抽出した意図と対話ルールに基づいて、推論モジュールを用いて、案内場所を推論し、ユーザとのやり取りを通じて案内場所の特定を行うプログラムを実行する。どこの部屋に何があるかを案内候補としてシステムに出力する。次に、ユーザが指定した物が違う部屋に複数存在していて、1 つに特定できていない場合は、システムがどちらの部屋にするか選択を促す。最終的に案内する部屋と物の場所を特定できた場合、図 7 のように VirtualHome を用いてシミュレーションを行い目的の場所に案内する。

5 評価実験の方針

タスク指向型対話システムに関する調査論文[7]によると、評価実験の手法として、以下の3つが挙げられる。

1. 自動評価

システム内のコンポーネントに対する自動評



図 6 提案システムの対話例

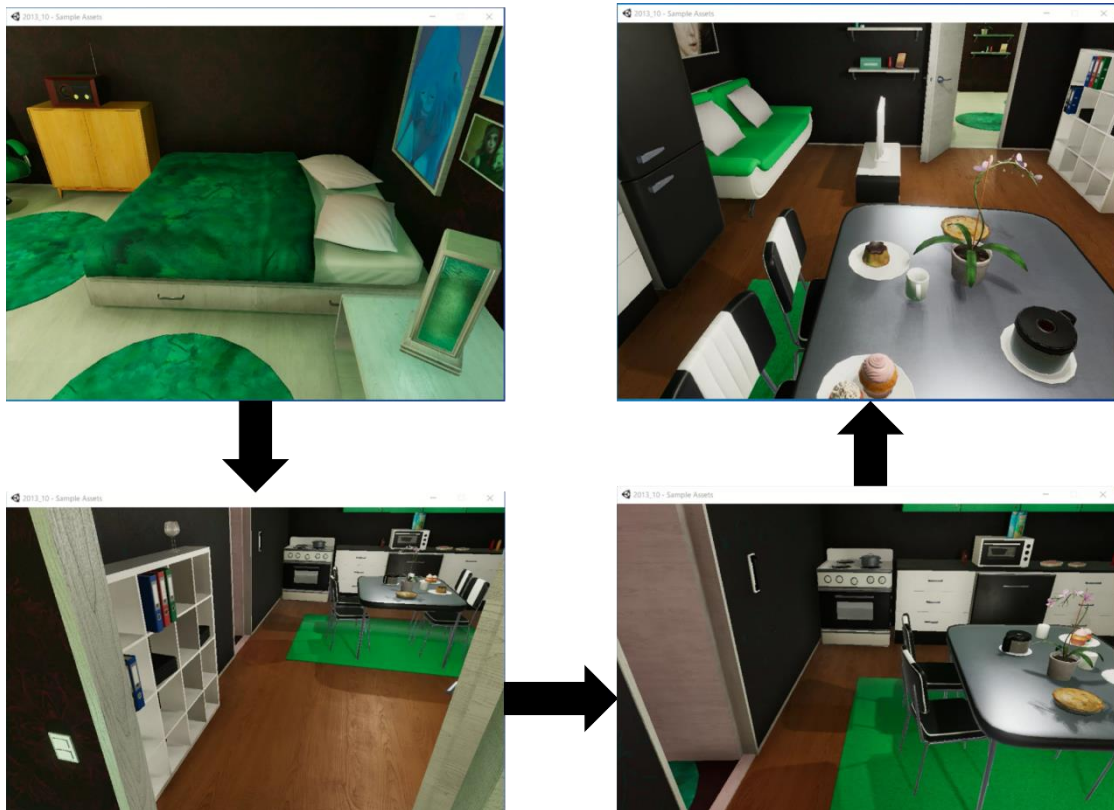


図7 シミュレーション結果

価値基準に基づいた評価方法

2. シミュレーションによる評価

ドメインの知識を活用してユーザシミュレータを構築し、人間のような会話を提供して評価を行う方法

3. 人間による評価

人間にシステムと対話を行う特定のタスクを完了してもらい、その体験について評価を行う方法

本対話システムでは、Embodied AI シミュレータ上でシステムとの対話に基づいて案内を行うため、案内が成功しているかどうかは人間による評価が必要になると考える。主な評価項目としては、案内が成功しているかどうかを表すタスク成功率、タスクが成功するまでの過程で不必要な対話があるかどうかを表す無関係ターン率、システムに対するユーザの満足度などが考えられる。

6 おわりに

本稿では、家庭環境オントロジーと常識推論に基づく双方向型対話ナビゲーションシステムの構成と実行例について述べた。

将来、仮想空間で学習された仮想ロボットを現実世界に反映することにより、家庭用の

家事ロボットや不動産における物件案内ロボットなどの開発に本研究を応用することが期待できる。

今後、対話システムに対話ルールや自然言語理解のデータセットを追加し、会話パターンを豊富にすることや、第5節で述べた評価実験を行う予定である。

対話ルールや自然言語理解のデータセットの拡張には、日本語日常会話コーパス(CEJC)[10]を用いる予定である。CEJCは、さまざまな場面における自然な日常会話を収集したコーパスである。本研究では、家庭内に関する対話を行うため、自宅での日常会話におけるコーパスを用いてデータセットの拡張を検討している。

参考文献

- [1]. J. Thomason, et al. Vision-and-Dialog Navigation. CoRL2019.
- [2]. D. Jena Hwangy, et al. (COMET-)ATOMIC2020: On Symbolic and Neural Commonsense Knowledge Graphs. AAAI2020.
- [3]. 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 DOI: 10.1093/ietisy/e91-d.4.945. 2018.
- [4]. C. Fellbaum. WordNet and wordnets. In: K. Brown et al. Encyclopedia of Language and Linguistics, Second Edition, Oxford: Elsevier, 665-670. 2005.
 - [5]. T. Bocklisch, et al, A. Nichol. Rasa: Open Source Language Understanding and Dialogue Management. NIPS 2017 Conversational AI workshop. 2017.
 - [6]. X. Puig, et al. "VirtualHome: Simulating household activities via pro-grams," Proc. the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8494–8502. 2018
 - [7]. Z. Zhang, et al. Recent Advances and Challenges in Task-oriented Dialog Systems. Sci. China Technol. Sci. 63, 2011-2027 (2020).
 - [8]. J. Duan, et al. A SURVEY OF EMBODIED AI: FROM SIMULATORS TO RESEARCH TASKS. CVIU2021
 - [9]. M. Burtsev, et al. DeepPavlov: Open-Source Library for Dialogue Systems. Proc. the 56th Annual Meeting of the Association for Computational Linguistics-System Demonstrations, pp. 122-127, 2018.
 - [10]. 小磯花絵, 居關友里子, 臼田泰如, 柏野和佳子, 川端良子, 田中弥生, 伝康晴, 西川賢哉. 「日本語日常会話コーパス」の構築. 言語処理学会第23回年次大会(NLP2017)予稿集.