



## معرفی عنوان پروژه

استاد پروژه: دکتر محمد اکبری  
استاد مشاور: دکتر مهدی قطعی  
نویسنده اول: امیرحسین کریمی

محمدرضا اردستانی  
کارشناسی علوم کامپیوتر  
۹۵۱۳۰۰۴

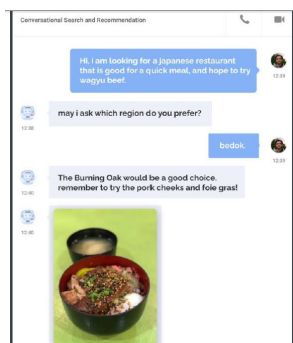
## ۱ شرح پروژه

### ۳.۱ تعریف مسئله

مسئله به این صورت است که یک شخص با سیستم در زمینه های اشاره شده گفتگو میکند تا به پاسخ خواسته خود برسد و در این حین سیستم دو کار عمده را نیز انجام میدهد: اول این که در زمان هایی که تشخیص بر recommend کردن داده شده است، سیستم یک عکس مرتبط با آن رستوران یا غذای خاص که در حال توصیه آن است به کاربر نشان میدهد. در بخش دوم، سعی میکند که مرتبط ترین عکس خود را به کاربر نشان دهد. به عنوان مثال وقتی کاربر درخواست یک رستوران با محوطه باز میکند و ما از رستوران مشخص شده عکس های زیادی داریم (از غذای آن رستوران، از مکان بیرونی، درونی، ...) ما عکسی به بیشترین ارتباط با این گفتگو فعلی را دارد، (در این مورد عکس مربوط به فضای بیرونی رستوران) را نشان می دهیم. مدل زبانی ای که استفاده خواهد شد SimpleTOD [4] است که از به روزترین مقاله های مرتبط به این زمینه است که از GPT-2 بهره گرفته است.

### ۴.۱ مثال

به عنوان مثال کاربر وارد سیستم میشود و درخواست خود را ارسال میکنم و هنگامی که سیستم در حالت پیشنهاد دادن می رود، یک عکس مرتبط را هم نمایش میدهد.



عکس ساخته شده از مجموعه داده mmconv

### ۱.۱ مقدمه

سه نوع مختلف چت بات،  $TOD^1$ ،  $CA^2$  و  $QA^3$ ، که در مقاله [1] به طور کامل مقایسه آن ها و جزئیات هر یک بررسی شده است، وجود دارد. هدف ما توسعه چت بات TOD هست، اما برخلاف نسخه های دیگر که تنها بر روی یک رسانه و یا یک زمینه کار میکنند، از سیستم چندرسانه ای-چندزمینه ای در این پروژه استفاده خواهد شد. دیتاست ای که ما استفاده خواهیم نمود mmconv [2] است. این دیتاست در ۵ زمینه مختلف (غذا، هتل، شبگردی، فروشگاه، مکانهای گردشگری) و همچنین دو رسانه مختلف (عکس و متن) پشتیبانی میکند.

### ۲.۱ انگیزه

انگیزه اصلی پاس کردن تست تورینگ در تمام زمینه ها بوده است، به طوری که یک شخص بتواند با یک سیستم در زمینه های مختلف صحبت کند بدون آن که بتواند تشخیص دهد طرف مقابل گفت و گو یک ربات میباشد، به همین خاطر شاخه چندرسانه ای-چندزمینه ای را انتخاب شده است. شرکت های متنوعی نیز، از جمله Nuance<sup>۴</sup>، در حال فعالیت هستند که زمینه همکاری های آینده را نیز برای این پروژه فراهم میکنند. همچنین در این پروژه درک عمیق تری از روش های یادگیری ماشین مانند ترنسفورمرها [3] در پردازش زبان به دست خواهیم آورد که میتواند در پروژه یادگیری عمیق و درک مطلب<sup>۵</sup> از آن بهره ببریم.

<sup>1</sup>Task Oriented Dialogue System

<sup>2</sup>Conversational Agent

<sup>3</sup>Question Answering

<sup>4</sup><https://www.nuance.com/index.html>

<sup>5</sup>[microsoft.com/en-us/research/project/machine-reading-comprehension/](https://microsoft.com/en-us/research/project/machine-reading-comprehension/)

## ۵.۱ معیارهای ارزیابی

در سیستم های پردازش زبان طبیعی، یکی از روش های بررسی عملکرد سیستم توسط انسان<sup>۶</sup> میباشد، که ما برای تست های اولیه از آن بهره خواهیم نمود. اما معیارهای ارزیابی نهایی برای بخش اول مسئله inform rate,success rate,and combined score خواهد بود. برای بخش دوم مسئله به صورت supervised و از معیار accuracy استفاده میکنیم. به این صورت که برای زمینه و نوع action که ما در مدل simpleTOD داریم یک لیبل هدف نظر میگیریم و عکس های موجود در دیتابیس را هم با همان لیبل گذاری میکنیم.

## ۶.۱ مراجع

### References

- [1] Jan Deriu, Alvaro Rodrigo, Arantxa Otegi, Guillermo Echegoyen, Sophie Rosset, Eneko Agirre, and Mark Cieliebak. Survey on evaluation methods for dialogue systems. *Artificial Intelligence Review*, 54(1):755–810, Jan 2021.
- [2] Lizi Liao, Le Hong Long, Zheng Zhang, Minlie Huang, and Tat-Seng Chua. Mmconv: An environment for multimodal conversational search across multiple domains. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '21*, page 675–684, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380379.
- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.
- [4] Ehsan Hosseini-Asl, Bryan McCann, Chien-Sheng Wu, Semih Yavuz, and Richard Socher. A simple language model for task-oriented dialogue, 2020.