

Sipmle Method For Dialogue State Tracking In MultiWOZ 2.4 Dataset

Phạm Ngọc Hải

27/12/2024

Abstract

```
<!--  
pandoc raw0.md -o RawManuscript.pdf --pdf-engine=xelatex  
-->
```

Introduction

Theo dõi trạng thái hội thoại (Dialogue State Tracking - DST) là một thành phần **then chốt** trong các hệ thống hội thoại hướng tác vụ (Task-Oriented Dialogue Systems).

Mục tiêu của DST là theo dõi và cập nhật trạng thái hội thoại theo thời gian thực, dựa trên các lượt tương tác giữa người dùng và hệ thống. Từ đó hỗ trợ cho việc đưa ra quyết định và tạo phản hồi phù hợp từ hệ thống.

Trạng thái này bao gồm những thông tin quan trọng như:

- **Ý định của người dùng (User Intent):** Mong muốn hoặc mục đích của người dùng (ví dụ: đặt vé máy bay, tìm nhà hàng, hỏi thời tiết).
- **Tham số liên quan (Slots/Entities):** Các chi tiết cụ thể liên quan đến ý định, như điểm đi, điểm đến, thời gian, hoặc các yêu cầu đặc thù khác.

Ví dụ Minh Họa

Người dùng: “Tôi muốn đặt vé máy bay từ Hà Nội đi Sài Gòn vào ngày mai.”

Trong trường hợp này, DST cần xác định:

- **Ý định:** Đặt vé máy bay.
- **Tham số liên quan:**
 - Điểm đi: Hà Nội.
 - Điểm đến: Sài Gòn.
 - Ngày đi: Ngày mai.

Sự phát triển của các tập dữ liệu hội thoại đa miền (multi-domain dialogue datasets) quy mô lớn như MultiWOZ 2.0 đã thúc đẩy đáng kể nghiên cứu về DST. Tuy nhiên, các phiên bản MultiWOZ ban đầu thường chứa nhiễu (noise) trong các chú thích trạng thái, **gây khó khăn** cho việc đánh giá hiệu suất mô hình một cách chính xác và công bằng. Để giải quyết vấn đề này, MultiWOZ 2.4 đã được giới thiệu như một phiên bản **tinh chỉnh** của MultiWOZ 2.1, tập trung vào việc sửa các lỗi chú thích trong tập validation và tập test. MultiWOZ 2.4 đã được chứng minh là mang lại **hiệu suất cao hơn** cho các mô hình DST tiên tiến, đồng thời tạo ra một môi trường đánh giá **chính xác và đáng tin cậy** hơn.

Mặc dù các phương pháp DST phức tạp đã đạt được những thành tựu đáng kể, việc phát triển một phương pháp **đơn giản và hiệu quả** vẫn là một **thách thức**, đặc biệt là khi xử lý các tập dữ liệu lớn và phức tạp như MultiWOZ 2.4.

Trong nghiên cứu này, chúng tôi đề xuất một phương pháp **đơn giản** cho DST trên tập dữ liệu MultiWOZ 2.4. Nghiên cứu này tập trung vào:

- **Giới thiệu một phương pháp DST mới:** Phương pháp này được thiết kế đơn giản, dễ hiểu và triển khai, đồng thời vẫn đảm bảo hiệu quả trong việc theo dõi trạng thái hội thoại trên tập dữ liệu MultiWOZ 2.4.
- **Đánh giá hiệu suất của phương pháp:** Chúng tôi sẽ đánh giá phương pháp được đề xuất trên tập dữ liệu MultiWOZ 2.4, sử dụng các chỉ số đánh giá tiêu chuẩn như độ chính xác mục tiêu chung (Joint Goal Accuracy) và độ chính xác slot (Slot Accuracy).
- **So sánh với các phương pháp DST hiện có:** Kết quả của phương pháp được đề xuất sẽ được so sánh với các phương pháp DST tiên tiến khác đã được đánh giá trên MultiWOZ 2.4 để chứng minh tính hiệu quả và khả năng cạnh tranh của phương pháp.

Nghiên cứu này hy vọng đóng góp một phương pháp DST **đơn giản, hiệu quả và dễ tiếp cận** cho cộng đồng nghiên cứu, giúp **giảm bớt rào cản** trong việc áp dụng DST vào các ứng dụng thực tế.

Related Work

Các Hướng Nghiên Cứu Chính

Mô Hình Dựa Trên Bản Thể (Ontology-Based Models)

- **Bản thể tĩnh (Static Ontology):** Giới hạn trong một tập hợp trạng thái cố định (ví dụ: FJST, TSCP).
- **Bản thể động (Dynamic Ontology):** Linh hoạt trong việc dự đoán các trạng thái mới (ví dụ: TRADE, TripPy, SOM-DST).

1. Phương Pháp Dựa Trên Ontology (Ontology-Based Methods):

<! --

* **TRADE (Wu et al., 2019):** Sử dụng cơ chế copy để tạo trạng thái hội thoại và xác

* **SOM-DST (Kim et al., 2020):** Áp dụng phương pháp ghi nhớ chọn lọc để cập nhật th

-->

- **Ý tưởng:** Các phương pháp này sử dụng một ontology chứa tập giá trị slot được xác định trước, từ đó dự đoán giá trị slot dựa trên danh sách các giá trị khả dĩ.
- **Ưu điểm:** Phù hợp với các hệ thống có miền xác định rõ ràng, đơn giản và dễ triển khai.
- **Nhược điểm:** Hạn chế trong việc mở rộng sang miền mới và xử lý các giá trị chưa biết.
- **Nghiên cứu cụ thể:**
 - **“Method_A Sequence-to-Sequence Approach to Dialogue State Tracking.pdf”:** Bài báo này đề xuất phương pháp Seq2Seq-DU, sử dụng BERT để mã hóa các câu nói và lược đồ, sau đó dùng bộ giải mã để tạo con trỏ đại diện cho trạng thái hiện tại của hội thoại. Bài báo cũng so sánh Seq2Seq-DU với các phương pháp DST hiện có trên nhiều tập dữ liệu khác nhau, bao gồm cả các phương pháp dựa trên bản thể.
 - **“Method_Adaptive Multi-Domain Dialogue State Tracking on Spoken Conversations.pdf”:** Bài báo này đề cập đến các phương pháp dựa trên bản thể tĩnh (Static Ontology-based) như một trong những hướng tiếp cận chính cho DST truyền thống.
 - **“Method_Continual Prompt Tuning for Dialog State Tracking.pdf”:** Bài báo này phân loại các phương pháp DST, bao gồm cả phương pháp dựa trên bản thể tĩnh, dựa trên việc sử dụng tập hợp ứng viên giá trị slot.
 - **“Method_Dialogue State Distillation Network with Inter-slot Contrastive Learning for Dialogue State Tracking.pdf”:** Bài báo này chia các phương pháp DST thành hai loại: dựa trên bản thể và không dựa trên bản thể.
 - **“Method_Dialogue State Tracking with Explicit Slot Connection Modeling.pdf”:** Bài báo này thảo luận về các phương pháp dựa trên bản thể tĩnh, sử dụng phân loại đa lớp trên tập hợp ứng viên giá trị slot được xác định trước.
 - **“Method_Domain State Tracking for a Simplified Dialogue System.pdf”:** Bài báo này đề cập đến việc sử dụng các mô hình ngôn ngữ được đào tạo trước như BERT trong các hệ thống hội thoại hướng đến mục tiêu, một số phương pháp trong số đó có thể sử dụng phương pháp dựa trên bản thể.
 - **“Method_Effective Sequence-to-Sequence Dialogue State Tracking.pdf”:** Bài báo này so sánh phương pháp được đề xuất với nhiều phương pháp cơ sở mạnh, bao gồm cả các phương pháp dựa trên bản thể tĩnh.
 - **“Method_Efficient Context and Schema Fusion Networks for Multi-Domain Dialogue State Tracking.pdf”:** Bài báo này thảo luận về các phương pháp DST dựa trên bản thể tĩnh và so sánh chúng với các phương pháp dựa trên từ vựng mở.
 - **“Method_End-to-End Multi-Domain Task-Oriented Dialogue Systems with Multi-level Neural Belief Tracker .pdf”:** Bài báo này phân loại các mô hình DST thành hai loại chính: dựa trên từ vựng cố định và dựa trên từ vựng mở. Mô hình được đề xuất trong bài báo này thuộc loại dựa trên từ vựng mở, nhưng bài báo cũng thảo luận về các mô hình dựa trên từ vựng cố định (tương đương với phương pháp dựa trên bản thể).
 - **“Method_From Machine Reading Comprehension to Dialogue State Tracking__ Bridging the Gap.pdf”:** Bài báo này đề cập đến các phương pháp

pháp DST truyền thống dựa trên bản thể tĩnh và thảo luận về những hạn chế của chúng. Bài báo cũng so sánh phương pháp được đề xuất với các phương pháp dựa trên bản thể tĩnh khác.

- **“Method_LUNA_ Learning Slot-Turn Alignment for Dialogue State Tracking.pdf”**: Bài báo này phân loại các phương pháp DST thành hai loại: phân loại và tạo. Các phương pháp phân loại thường dựa trên bản thể tĩnh.
- **“Method_Leveraging Slot Descriptions for Zero-Shot Cross-Domain Dialogue State Tracking.pdf”**: Bài báo này đề cập đến các phương pháp DST dựa trên bản thể tĩnh và thảo luận về việc sử dụng mô tả slot để cải thiện hiệu suất của chúng. Bài báo cũng so sánh phương pháp được đề xuất với các phương pháp dựa trên bản thể tĩnh khác.
- **“Method_Meta-Reinforced Multi-Domain State Generator for Dialogue Systems.pdf”**: Bài báo này thảo luận về các phương pháp DST dựa trên bản thể tĩnh và so sánh chúng với các phương pháp dựa trên từ vựng mở.
- **“Method_Multi-Domain Dialogue State Tracking based on State Graph.pdf”**: Bài báo này mô tả các phương pháp DST dựa trên bản thể tĩnh và so sánh hiệu suất của mô hình được đề xuất với các phương pháp dựa trên bản thể tĩnh khác.
- **“Method_Multi-Domain Dialogue State Tracking with Disentangled Domain-Slot Attention.pdf”**: Bài báo này thảo luận về các mô hình DST dựa trên bản thể và so sánh hiệu suất của mô hình được đề xuất với các phương pháp dựa trên bản thể khác.
- **“Method_Multi-domain gate and interactive dual attention for multi-domain dialogue state tracking.pdf”**: Bài báo này đề cập đến các mô hình dựa trên bản thể tĩnh và thảo luận về các ưu điểm và nhược điểm của chúng.
- **“Method_Non-Autoregressive Dialog State Tracking.pdf”**: Bài báo này phân loại các mô hình DST thành hai loại: dựa trên từ vựng cố định và dựa trên từ vựng mở. Các mô hình dựa trên từ vựng cố định tương đương với phương pháp dựa trên bản thể.
- **“Method_SAS_ Dialogue State Tracking via Slot Attention and Slot Information Sharing.pdf”**: Bài báo này tập trung vào việc chia sẻ thông tin slot trong DST dựa trên bản thể.
- **“Method_TripPy_ A Triple Copy Strategy for Value Independent Neural Dialog State Tracking.pdf”**: Bài báo này thảo luận về các phương pháp DST truyền thống dựa trên bản thể tĩnh và so sánh chúng với các phương pháp dựa trên từ vựng mở.

2. Phương Pháp Không Cần Ontology (Ontology-Free Methods):

- **Ý tưởng**: Trích xuất giá trị slot trực tiếp từ ngữ cảnh hội thoại, không cần tập giá trị xác định trước.
- **Ưu điểm**: Linh hoạt, xử lý được các giá trị chưa biết, phù hợp với các hệ thống mở miền.
- **Nhược điểm**: Khó kiểm soát chất lượng trích xuất giá trị slot.
- **Nghiên cứu cụ thể**:
 - **“Method_A Sequence-to-Sequence Approach to Dialogue State**

Tracking.pdf”: Bài báo này giới thiệu ba phương pháp không cần ontology: COMER, CREDIT, và SimpleTOD. Cả ba phương pháp đều coi DST như một bài toán tạo chuỗi và sử dụng các kiến trúc encoder-decoder để dự đoán chuỗi trạng thái dựa trên câu nói của người dùng.

- * COMER sử dụng mô hình encoder-decoder phân cấp dựa trên BERT.
- * CREDIT cũng sử dụng mô hình encoder-decoder phân cấp nhưng không sử dụng BERT.
- * SimpleTOD sử dụng mô hình chuỗi-chuỗi thống nhất dựa trên GPT-2.

- **“Method_Continual Prompt Tuning for Dialog State Tracking.pdf”**: Bài báo này đề cập đến các phương pháp không cần ontology như một cách để giảm sự phụ thuộc vào ontology và cải thiện khả năng khái quát hóa cho các giá trị chưa nhìn thấy. Một số phương pháp trong số này trích xuất giá trị từ ngữ cảnh hội thoại, trong khi một số khác tạo giá trị trực tiếp để xử lý các trường hợp giá trị bị thiếu trong ngữ cảnh.
- **“Method_Dialogue State Distillation Network with Inter-slot Contrastive Learning for Dialogue State Tracking.pdf”**: Bài báo này cũng chia các phương pháp DST thành hai loại: dựa trên bản thể và không dựa trên bản thể. Các phương pháp không dựa trên bản thể tạo ra các giá trị của slot từ ngữ cảnh hội thoại hoặc từ vệt.
- **“Method_Dialogue State Tracking with Explicit Slot Connection Modeling.pdf”**: Bài báo này thảo luận về các phương pháp không cần ontology, còn được gọi là các phương pháp từ vựng mở (open vocabulary), cho phép tạo giá trị slot chỉ với các slot mục tiêu.
- **“Method_Effective Sequence-to-Sequence Dialogue State Tracking.pdf”**: Bài báo này so sánh phương pháp được đề xuất với nhiều phương pháp cơ sở khác nhau, bao gồm cả các phương pháp không cần ontology.
- **“Method_Efficient Context and Schema Fusion Networks for Multi-Domain Dialogue State Tracking.pdf”**: Bài báo này thảo luận về các phương pháp DST không cần ontology, sử dụng kiến trúc encoder-decoder hoặc mạng con trở để tạo hoặc trích xuất giá trị trực tiếp cho mỗi slot từ ngữ cảnh hội thoại.
- **“Method_End-to-End Multi-Domain Task-Oriented Dialogue Systems with Multi-level Neural Belief Tracker .pdf”**: Bài báo này phân loại các mô hình DST thành hai loại chính: dựa trên từ vựng cố định và dựa trên từ vựng mở. Mô hình được đề xuất trong bài báo này thuộc loại dựa trên từ vựng mở (tương đương với phương pháp không cần ontology).
- **“Method_From Machine Reading Comprehension to Dialogue State Tracking_ Bridging the Gap.pdf”**: Bài báo này thảo luận về các phương pháp không cần ontology, bao gồm các phương pháp sử dụng cơ chế trở dựa trên attention để tìm vị trí bắt đầu và kết thúc của giá trị slot.
- **“Method_In-Context Learning for Few-Shot Dialogue State Tracking.pdf”**: Bài báo này đề xuất một khung học trong ngữ cảnh (ICL) cho DST zero-shot và few-shot, trong đó một mô hình ngôn ngữ lớn được đào tạo trước (LM) nhận một instance kiểm tra và một vài ví dụ làm đầu vào và giải mã trực tiếp trạng thái hội thoại mà không cần cập nhật tham số. Phương pháp này không cần ontology vì LM có thể học cách dự đoán giá trị slot từ ngữ cảnh.
- **“Method_LUNA_ Learning Slot-Turn Alignment for Dialogue**

State Tracking.pdf”: Bài báo này phân loại các phương pháp DST thành hai loại: phân loại và tạo. Các phương pháp tạo thường không cần ontology và tạo ra trạng thái hội thoại từ các câu nói bằng cách sử dụng kiến trúc seq2seq.

- **“Method_Meta-Reinforced Multi-Domain State Generator for Dialogue Systems.pdf”**: Bài báo này thảo luận về các phương pháp DST không cần ontology và sử dụng một bộ tạo trạng thái để tạo ra các giá trị slot từ ngữ cảnh hội thoại.
- **“Method_Multi-Domain Dialogue State Tracking based on State Graph.pdf”**: Bài báo này tập trung vào DST không cần ontology, trong đó các giá trị không được xác định trước và cần được trích xuất (tạo) trực tiếp từ đầu vào.
- **“Method_Multi-Domain Dialogue State Tracking with Disentangled Domain-Slot Attention.pdf”**: Bài báo này chủ yếu tập trung vào DST dựa trên bản thể. Tuy nhiên, bài báo cũng đề cập đến việc áp dụng phương pháp attention được đề xuất cho các mô hình tạo, vốn có thể không cần ontology.
- **“Method_Multi-domain gate and interactive dual attention for multi-domain dialogue state tracking.pdf”**: Bài báo này thảo luận về các mô hình không cần ontology, thường được gọi là mô hình từ vựng mở (open vocabulary), dựa vào việc tạo hoặc trích xuất giá trị slot từ lịch sử hội thoại và trạng thái hội thoại hiện có.
- **“Method_Non-Autoregressive Dialog State Tracking.pdf”**: Bài báo này đề xuất một phương pháp không cần ontology, sử dụng một bộ giải mã không tự hồi quy để tạo ra tất cả các token của trạng thái hội thoại mục tiêu cùng một lúc.
- **“Method_SAS_ Dialogue State Tracking via Slot Attention and Slot Information Sharing.pdf”**: Bài báo này tập trung vào việc chia sẻ thông tin slot trong DST dựa trên bản thể. Tuy nhiên, các kỹ thuật attention và chia sẻ thông tin slot cũng có thể được áp dụng cho các phương pháp không cần ontology.
- **“Method_TripPy_ A Triple Copy Strategy for Value Independent Neural Dialog State Tracking.pdf”**: Bài báo này đề xuất một phương pháp không cần ontology, sử dụng chiến lược sao chép ba lần dựa vào dự đoán span cũng như cơ chế bộ nhớ.
- **“Method_Zero-Shot Transfer Learning with Synthesized Data for Multi-Domain Dialogue State Tracking.pdf”**: Bài báo này tập trung vào việc học chuyển giao zero-shot cho DST đa miền. Mặc dù bài báo không đề cập rõ ràng, nhưng các phương pháp học chuyển giao zero-shot thường không cần ontology vì chúng được thiết kế để hoạt động trong các miền chưa nhìn thấy.

Hiểu Đọc Máy (Machine Reading Comprehension - MRC) & Học Trong Ngữ Cảnh (In-Context Learning)

Dựa Trên Hiểu Đọc Máy (Machine Reading Comprehension - MRC)

- Xử lý DST dưới dạng bài toán đọc hiểu, trong đó lịch sử hội thoại được mã hóa như đoạn văn và các slot được biểu diễn như câu hỏi.

- “**Method_From Machine Reading Comprehension to Dialogue State Tracking_Bridging the Gap.pdf**”: Bài báo này **chuyển đổi bài toán DST thành bài toán MRC** bằng cách thiết kế câu hỏi cho mỗi slot trong trạng thái hội thoại. Lịch sử hội thoại được coi như đoạn văn và các slot được biểu diễn dưới dạng câu hỏi.
 - Bài báo phân loại slot thành 2 loại: **categorical** và **extractive**.
 - * Slot categorical được xử lý bằng mô hình MRC dạng câu hỏi trắc nghiệm, chọn đáp án từ tập giới hạn.
 - * Slot extractive được xử lý bằng mô hình MRC span-based, tìm kiếm đáp án là một đoạn văn bản trong lịch sử hội thoại.
- “**Method_End-to-End Multi-Domain Task-Oriented Dialogue Systems with Multi-level Neural Belief Tracker .pdf**”: Bài báo đề cập đến mô hình **DST Reader**, coi DST như một bài toán đọc hiểu. Mô hình này **dự đoán mỗi slot như một khoảng văn bản trong lịch sử hội thoại** và sử dụng mạng nơ-ron dựa trên attention với các mô-đun bổ sung để dự đoán loại slot và xác suất chuyển tiếp slot.
- “**Method_Non-Autoregressive Dialog State Tracking.pdf**”: Bài báo cũng đề cập đến mô hình **DST Reader** và mô tả nó như một phương pháp **biểu diễn DST thành bài toán đọc hiểu**.

Học Trong Ngữ Cảnh (In-Context Learning)

- Sử dụng các mô hình ngôn ngữ lớn (như GPT) để xử lý DST trong bối cảnh zero-shot hoặc few-shot, cho phép giải quyết các miền chưa từng thấy trước đó.
- “**Method_In-Context Learning for Few-Shot Dialogue State Tracking.pdf**”: Bài báo này đề xuất khung **Học Trong Ngữ Cảnh (ICL)** cho DST zero-shot và few-shot.
 - Mô hình ngôn ngữ lớn được huấn luyện trước (LM) sẽ nhận đầu vào là một instance kiểm tra và một số ví dụ, sau đó giải mã trực tiếp trạng thái hội thoại mà không cần cập nhật tham số.
 - Phương pháp này **không cần ontology** vì LM có thể học cách dự đoán giá trị slot từ ngữ cảnh.
 - Bài báo giới thiệu **mô hình IC-DST** kết hợp các cải tiến sau:
 - * Chuyển đổi DST thành bài toán text-to-SQL, tích hợp mô tả dạng bảng của ontology vào prompt.
 - * Sử dụng trạng thái hội thoại thay vì toàn bộ lịch sử hội thoại để biểu diễn ngữ cảnh.
 - * Đề xuất phương pháp mới để học điểm tương đồng nhằm lựa chọn các ví dụ trong ngữ cảnh.

Các Kỹ Thuật Cốt Lõi

- **Transformer và Attention Networks:** Hiểu và biểu diễn thông tin hội thoại hiệu quả.
- **Cơ chế Copy:** Xử lý các giá trị chưa biết bằng cách sao chép từ lịch sử hội thoại.
- **Học chuyển giao (Transfer Learning):** Tận dụng kiến thức từ miền đã biết cho miền mới.

- **Học liên tục (Continual Learning):** Cho phép mô hình cập nhật liên tục khi dữ liệu thay đổi.
- **Học tập ít mẫu (Few-Shot Learning):** Áp dụng meta-learning và prompt learning để cải thiện hiệu suất trong điều kiện thiếu dữ liệu.

Xu Hướng Nghiên Cứu Mới

1. **Học Tập Ít Mẫu:** Giảm phụ thuộc vào dữ liệu gán nhãn nhờ các kỹ thuật meta-learning và fine-tuning.
2. **Học Tập Liên Tục:** Phát triển mô hình có khả năng thích nghi với miền mới mà không cần huấn luyện lại.
3. **DST Đa Phương Thức:** Kết hợp thông tin từ nhiều nguồn dữ liệu như văn bản, hình ảnh, âm thanh.
4. **Xử Lý Nhiều Dữ Liệu:** Áp dụng các framework kháng nhiễu như **ASSIST** (Ye et al., 2021).

Đánh Giá và Thách Thức

- **Xây Dựng Bộ Dữ Liệu:** Cần phát triển các bộ dữ liệu đa dạng như **MultiWOZ**, **SGD**, **CrossWOZ** để đánh giá toàn diện hơn.
- **Phát Triển Phương Pháp Đánh Giá Mới:** Tập trung vào các yếu tố như khả năng hiểu ngữ cảnh, xử lý thông tin đa modal và khả năng thích ứng.

Method

Results

Discussion

Conclusion

Acknowledgements

References