

進捗報告

1 今週やったこと

- 資料のまとめ

2 質問応答

質問応答タスクとは、文書をもとに、入力された質問を正しく応答することを目指すタスクである。そして対象とする知識の範囲が限らない質問応答はオープンドメイン質問応答と呼ばれる。

3 灯謎 (トウメイ)

灯謎とは、中国の伝統クイズ問題である。質問者は問題を詩や熟語の形で提出し、回答者が回答する。答えは常に字や、単語の形式である。質問応答とは違い、灯謎は質問を回答するための文書や知識など必要はない、単に質問の中に答えの情報を得る、言い換えると、質問を理解すれば回答できる。故に情報量が少ない質問応答にも見える。灯謎を解くため、謎に隠された情報をもとに、質問を理解しなければならないので、灯謎の研究は人の言葉の情報抽出、分析に役に立つと思う。

4 灯謎パターン

灯謎のパターンはだいたい、謎、ヒント (時にはヒント 2 もある) と答えで構成される。謎は詩や熟語、あるいは普通のしゃべり言葉で作られ、ヒントは答えの形を説明する。答えは問題に隠された情報 (字の構成、発音、意味など) で解くことができる。

5 実験方法

灯謎の答えは不確定 (単語や字) であるので、Sequence To Sequence 問題として処理することができ、簡単に実験できるように、これからの実験は主に字謎 (回答は一文字の灯謎) で実験する、ゆえにマルチラベル問題として実験することができる。

6 データセット

現在、人工知能における灯謎の研究はほぼないので、実験用のデータはネット (交渉中) や、大学サークルのデータ (交渉中) で収集し、データセットを作るような形になる。今は 2000 くらいのデータが集まった。

7 来週目標

- 収集されたデータで実験する