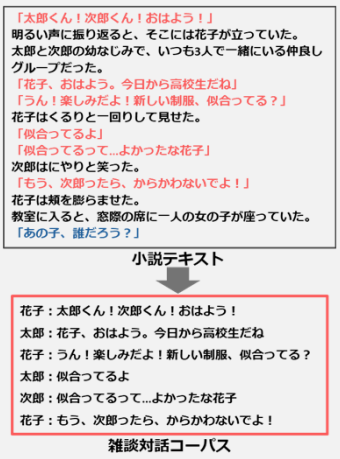
小説テキストを用いた雑談対話コーパスの自動構築とその分析に関する研究

○岩本 和真[1]　安藤 一秋[2]　（香川大学　[1]創発科学研究科，[2]創造工学部）

1. 概要

- 雑談対話コーパスの構築コストを削減するため，**小説テキスト**から会話グループを抽出し発話者を特定・付与する**雑談対話コーパス（小説対話コーパス）の自動構築手法**を提案
- 会話グループ抽出として台詞間の**発話応答関係モデル**を構築し，F1値80%の性能を確認
- 発話者特定として，ルールベース，大規模言語モデル（LLM），口調の特徴と**複数の手がかりを用いた特定手法**を提案しPrecision78%，F0.5値70%の性能を確認
- 提案手法を用いて小説対話コーパスを構築し，その特徴を分析
  - **コーパスの規模分析** ▶ 会話文脈の長い会話や複数人による会話を構築可能
  - **口調に関する分析** ▶ コーパスには多様な口調や人間関係を考慮した会話が存在
  - **対話モデルを用いた分析** ▶ 親密性が高く、口調特性を反映した会話モデルの構築が可能

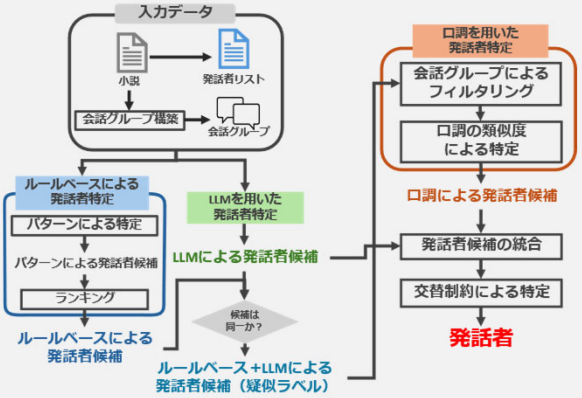


2. 会話グループ抽出

- 小説内の連続した台詞を1会話として抽出した場合，1会話に含まれる平均発話数が2.96と少ない
- **文脈を考慮するため**BERTを用いた**発話応答関係判定モデル**を構築し，その判定結果を会話グループの抽出に利用
- 提案モデルはBERTの複数の隠れ層を用いたモデル構造を採用
- 性能評価の結果、ルールベースがF1値64%の性能に対して提案モデルはF1値80%の性能であることを確認

3. 発話者特定

- ルールベース，LLM，口調ベクトルと複数の手がかりを用いることで，**発話者を正確に特定**することが目的
- 評価結果から，Precision78%とF0.5値70%の性能で特定できることを確認



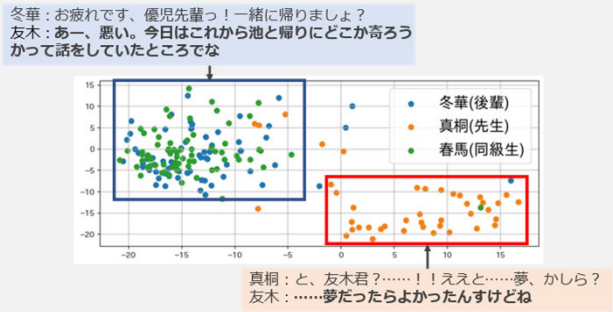
4. コーパスの規模分析

- 「小説家になろう」の恋愛ジャンル20タイトルで構築
- 7,500発話から1会話平均5発話含む会話が1,300件構築可能
- 各小説で**10発話以上含む会話**とターン数が多い会話を構築
- **3人以上**で構成されている会話など複数人による会話が存在

	発話数	会話数	1会話内の発話数	特定発話数	特定会話数
合計	149,855	26,301	-	98,550	8,317
平均	7,492.7	1315.0	5.73	4927.5	415,9

5. 口調に関する分析

- TF-IDFを用いた助詞・助動詞と人称代名詞の分析の結果，コーパス内に**11種類**の口調タイプと多様な口調の種類を確認
- 同一の発話者における話し手ごとの口調ベクトルの分布から話し相手との関係性により**口調を変える会話行為**があり，人間関係を考慮した会話も存在



6. 対話モデルを用いた分析

- **目的**：小説対話コーパスを学習させることで対話モデルを構築し，そのモデルが出力する会話の特徴を分析
- 分析対象モデル（LLMはSwallow-8Bを使用）
  - モデル1：登場人物の5つの発話例を含むプロンプトで指示
  - モデル2：プロンプトに加え，コーパスの会話文のみで学習
  - モデル3：プロンプトに加え，コーパスにおいて発話者情報が付与された，二人が交互に会話するデータで学習
- **評価指標**
  - 会話の自然性：発話の意図不明や不適切な応答がないか
  - 会話の親密性：会話を通して親しみを感じられるか
  - 口調の特性　：口調例の特徴を反映できているか
  - 口調一貫性　：口調が一貫しているか

	LLMによる評価 ⬆				人手による評価 ⬇			
	自然性	親密性	口調特性	口調一貫性	自然性	親密性	口調特性	口調一貫性
モデル1	**4.800**	3.460	1.433	**3.980**	**1**	2.3	3	**1.6**
モデル2	3.433	3.147	2.793	3.220	2.7	2.7	2	2.7
モデル3	4.155	**3.490**	**3.500**	3.690	2.3	**1.3**	**1**	**1.6**

以上より口調特性があり親密性の高い雑談対話モデルが構築可能

7. 応用活用例と今後の展望

活用例1：小説を用いたキャラクターAIの開発

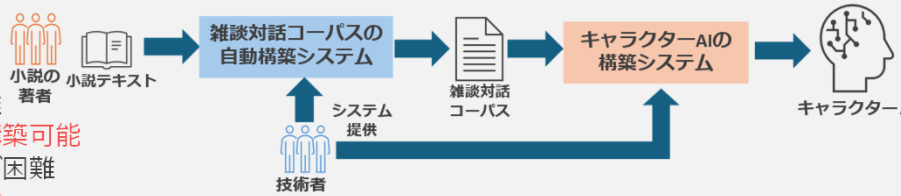
- 著者側：技術的要素、構築コストから実用化が困難
  ▶ 自動構築システムにより著者自身がAIを構築可能
- エンジニア側：構築コスト、著作権により商用化が困難
  ▶ 基盤システムの提供により、実用化を促進

活用例2：エンタメ業界における創作支援

- ゲーム作成におけるNPCの会話文生成
- キャラクターの人物像やシナリオに応じた会話文の創作支援

活用例3：教育における応用

- キャラクターAIを介したコミュニケーション練習



今後の展望

- 小説から周辺情報や人間関係、感情情報を抽出して，小説対話コーパスを拡張する手法の検討
- より柔軟にユーザと雑談できる対話モデルの検討
- 構築手法の性能向上による小説対話コーパスの質向上