

➤ Kyutechコーパス

■意思決定タスク対話データ

- ◇自由に利用できる対話要約コーパス

■収録データ・アノテーション

- ◇9つの対話データ(4種類の議論条件)
- ◇書き起こしデータ
- ◇トピックアノテーション
- ◇参照要約
- ◇**抜粋要約 (New)**

◆発話単位でのアノテーション

◆参照要約文とのトピックを分析

- ・アノテーションの妥当性
- ・より適切なアノテーションのための考察

タグ	説明
CandX	候補店1に関するトピック
CandY	候補店2に関するトピック
CandZ	候補店3に関するトピック
Cands	候補店1~3の複数に関するトピック
Closed	閉店したレストランについてのトピック
Exist1	既存店1に関するトピック
Exist2	既存店2に関するトピック
Exist3	既存店3に関するトピック
Exist4	既存店4に関するトピック
Exist5	既存店5に関するトピック
Exist6	既存店6に関するトピック
Exists	既存店1~6の複数に関するトピック
CIEx	既存店および閉店したレストランの両方に関するトピック
Mall	UBC モール全体についてのトピック
OtherMall	他の場所にあるショッピングモールに関するトピック
Location	モール内でのレストラン同士の位置関係に関するトピック
Area	地域や都市についてのトピック
People	ターゲットにする顧客についてのトピック
Price	価格に関するトピック
Menu	メニューに関するトピック
Atomos	雰囲気に関するトピック
Time	営業時間などに関するトピック
Seat	座席数や回転率などに関するトピック
Sell	売りに関するトピック
Access	交通手段などモールへのアクセスに関するトピック
Meeting	話を進めるための議事提案や最終決定部に関連するトピック
Chat	レストラン選択には直接関係ない雑談
Vague	前後の文脈からも何を言っているかわからない場合

➤ 抜粋要約の構築と分析

■対話中の発話と参照要約のリンク付け

- ◇発話毎に参照要約の要約文と関連があるかどうかの二値のアノテーション

		参照要約	
1	商業施設UBCモール内のレストラン街において、<カレーナーク>の閉店後、次に出店させる店舗を<台湾ヌードル>、<つけ麺ふうじん>、<中華料理四川>から選ぶための議論が行われた。		
2	まずはじめに、参加者の各個人の現時点でどこか良いと思うかの投票が行われ、<つけ麺ふうじん>と<中華料理四川>の二つに話が絞られた。		

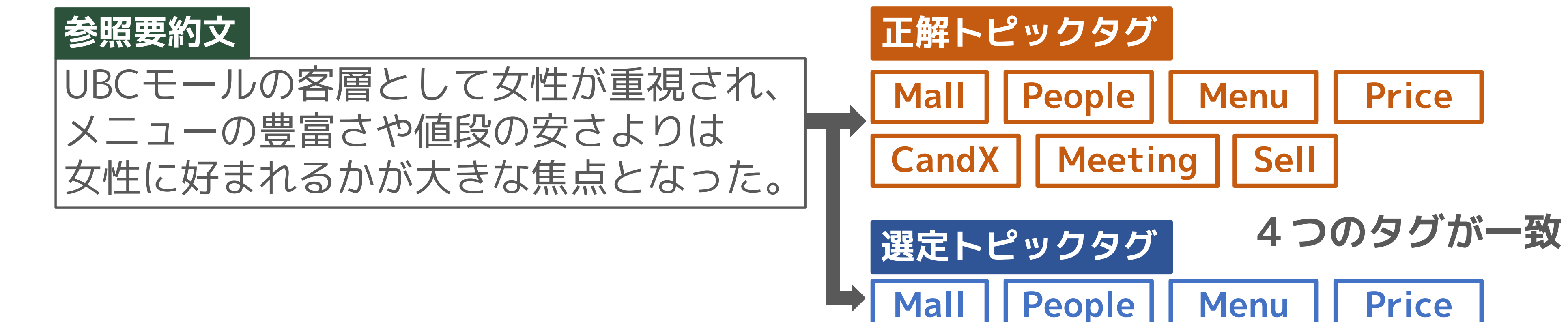
対話データ

ID	start	end	必須タグ	追加タグ1	追加タグ2	ExLink	発話
D	00:24.5	00:25.5	Meeting	Cands		2	現時点で+
D	00:26.6	00:27.6	Meeting	Cands		2	どこが良いと思う(Q)/
C	00:30.0	00:31.2	Meeting	CandY		2	私はつけ麺/
A	00:31.8	00:34.0	Meeting	CandY		2	自分も同じくつけ麺ふうじんが良いかなと/
B	00:35.0	00:38.2	Meeting	CandZ		2	僕は中華料理かなと考えてました/
D	00:38.4	00:40.7	Meeting	CandZ		2	自分も中華料理かなという/
C	01:02.6	01:06.9	Meeting	CandZ		2	なんで中華料理が良い(D 思ったか)思ったのか/
D	01:08.3	01:09.0	Closed	Price			んー/
D	01:12.7	01:14.2	Closed	Price			このカレー屋さん+
C	01:15.3	01:15.6	Closed	Price			うん/
D	01:15.8	01:17.2	Closed	Price			は(F まあ)+
D	01:18.8	01:20.2	Closed	Price			予算的にはそんなに+
D	01:20.6	01:21.8	Closed	Price			変わらなくて/
D	01:22.4	01:23.8	Closed	Sell			でもまあ売り上げあって/
C	01:25.4	01:25.8	Closed	Sell			うーん/

■参照要約と抜粋要約における分析

- ◇参照要約とリンク付けにおけるトピックタグの一致度を調査
- ◇参照要約の正解トピックタグとアノテータが選定したトピックタグを比較

- ◆**正解トピックタグ**：抜粋要約（リンク付けされた発話）のトピックタグの集合
- ◆**選定トピックタグ**：アノテータ1名が参照要約から推定したトピックタグ



タスクID	対話ID	発話数	リンク付け		トピックタグの比較			
			リンク数	割合(%)	正解タグ数	選定タグ数	一致数	F値
C1	0313	759	240	31.6	28	22	17	0.680
	0320	505	124	24.6	29	31	22	0.733
	0326	502	76	15.1	14	16	12	0.647
C2	0326	566	160	28.3	28	22	19	0.682
	0327	284	52	18.3	13	19	11	0.800
C3	0323	324	102	31.5	22	22	15	0.760
	0327	445	118	26.5	24	26	14	0.634
C4	0320	637	69	10.8	15	19	11	0.688
	0326	487	98	20.1	18	23	13	0.560

➤ 抜粋要約抽出実験

■研究の最終目標：対話要約

- ◇生成型要約手法
 - ◆重要文の抽出 & 発話の統合・補完

■実験設定

- ◇機械学習 (RandomForest, SVM)
- ◇BoW + 17の言語・非言語情報の素性
- ◇F値：0.298
- ◇議論構造の認識・対話の拡充が必要

➤ おわりに

■抜粋要約のアノテーション

- ◇各発話に対する重要度付与

リンク付けされていない参照要約文（全体の19%）

・会議では、定食屋の閉店後に出店させる店舗について議論された。
 ・UBCモールの不採算が原因で閉店することになったミスターKの代わりに新規店舗を沖縄料理あたる、台湾ヌードル、ポノパスタから選ぶ会議が行われた。
 ・UBCモールのレストラン街において閉店した定食屋に代わる新規店舗をラーメンかいづつ、つけ麺ふうじん、ポノパスタの三件の中から選ぶ会議が行われた。
 ・今回閉店するミスターKの代わりとなる新規店として沖縄料理あたる、台湾ヌードル、ポノパスタから選ぶ会議が行われた。
 ・その後、つけ麺ふうじんとポノパスタについてそれぞれの利点および欠点が議論された。

■今後の課題と展望

- ◇生成型対話要約
- ◇対話データの拡充・新たなタグ付与
 - ◆動画・音声データの公開, アノテーション
 - ◆発話意図タグの付与 (日野, 2016)

- 正解トピックタグの半数以上を選定
- 要約文の表層で単純に判断可能

抜粋要約の注釈支援システムの構築
より信頼性のあるアノテーション

