Kyutech コーパス 意思決定タスクを対象とした複数人対話コーパス (Giuseppe Carenini, Raymond Ng (UBC)

はじめに

■複数人対話コーパス

◇意思決定タスク

- ◆対話参加者: 1対話4名
- ◆架空の商業施設の新規参入店の選択
- ◆資料に基づき議論
 - ・候補店および既存店の情報
 - ・当該商業施設に関する立地や客層・性別分布など
 - ・近隣の市町村の人口分布や他の商業施設の情報
- ◆資料を10分黙読,20分の議論

◇コーパスに含まれるもの

- ◆4種類の条件&9つの対話データ
- ◆書き起こしデータ
- ◆トピックアノテーション
- ◆正解要約

この度あなた方には、商業施設「UBCモール」内のレストラン街において、「定食和屋(わや)」閉店後、 次に出店させる店舗を, UBCモールの経営者の立場で, 決定していただきたいと思います.

店名	台湾ヌードル	つけ麺 ふうじん	中華料理 四川
メニュー例	牛ヌードル:880円 ジャージャー麺:980円	つけ麺:700円 餃子:200円	麻婆豆腐:720円 中華丼:900円
予算	800円~1,200円	700円~1,000円	900円~1,500円
座席数	25	30	25
営業時間	11:00~23:00	11:00~23:00	11:00~23:00
概要	ぞ知る台湾ヌードルの老舗. 日	店. バラエティのあるメニュー	全国で300店舗以上を有する中 華料理チェーン店. 手頃な値段 で本格的な中華を味わえると有 名. 老若男女, 幅広い客層に人 気.
別店舗の口コミ	・ここでしか食べられない味がある(20代男性) ・スープのにおいがダメだった(30代男性) ・美味しい. 何度も食べに行きたい(30代女性)	ンも美味しい(30代男性)	・手頃な予算で食べられる(20 代女性) ・あともう少し安ければと思う (30代男性) ・種類も豊富で家族連れも多い (30代男性)

【UBC モールに関する情報】

UBC モールは,A 県 U 市の中心部付近にあるショッピングモール(複合商業施設)で,スーパーマーケットと 60 の専門店、ゲームセンター、映画館、7 つのレストランなどから成ります.

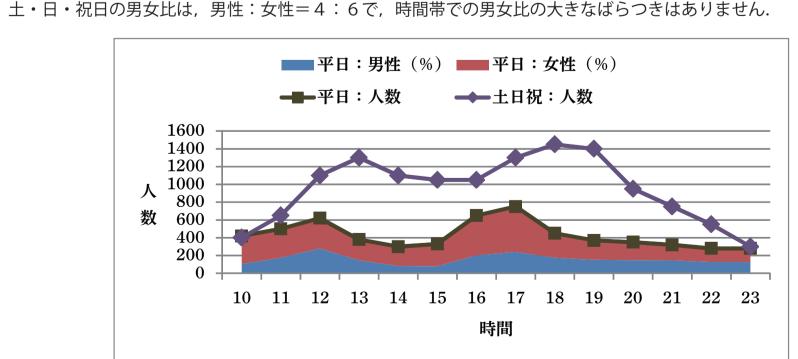
主なターゲット 所在地の U 市と隣接する X 市の住民です. 休日などにカップルや家族連れが長時間滞在する時間消費型の施設 としてのみならず, U市内に古くからある商店街の衰退に伴い, 日用品の買い物を含む多くの役割を担っています.

UBC モールには十分な無料駐車場の他に、U市内の各所から UBC モール向けのバスもあります。 郊外型のショッピングモールではないため、UBC モールの周りには規模は大きくありませんが、いくつかの

オフィスビルなども存在します. U市内には他にもスーパーマーケットやディスカウントストアなどの単体の商業施設がありますが、ショッピング

モールは UBC モールのみです. 来客者に関する情報

営業時間は店舗によって異なります.最も大きな施設であるスーパーマーケットは 10 時から 23 時まで 開いています。



対話環境





書き起こし

■転記単位

◇0. 2秒以上のポーズで区切り ◇9つの対話で4451転記単位

■タグ

タグ	説明
(F)タグ	フィラー
(D)タグ	言いよどみ・言い直し
(Q)タグ	疑問文:文末の表記やイントネーション
	から疑問文であると判断される場合
(%)タグ	上記では判断が難しい疑問文
(?)タグ	聞き取りに自信がない箇所
(L)タグ	ささやき声や独り言のような小さな声
<笑>	発話中もしくは単独の笑い

■発話単位

◇ルールと主観により発話単位を認定

Step1: まず, すべての転記単位の末尾を + とする.

Step2: 次の転記単位まで1秒以上の空白があり、以下の条件のどちらか にマッチした場合は + を / に変更する.

条件1:「動詞(もしくは助動詞)+接続詞」の形で終わる 条件2: 次の発話が「接続詞」,「フィラー」,「副詞」のいずれか

で始まる Step3: Step2で発話の終わりと認定されないが、作業者が1つの発話の 終わりであると感じる部分は + を * に変更する.

ID	Begin	End	Utterance
D	00:00:24.490	00:00:25.530	現時点で+
D	00:00:26.585	00:00:27.615	どこがいいと思う(Q)/
С	00:00:29.985	00:00:31.195	私はつけ麺/
А	00:00:31.815	00:00:33.965	自分も同じくつけ麺ふうじんがいいかな/
В	00:00:35.045	00:00:38.185	僕は中華料理かなと考えてました/
D	00:00:38.405	00:00:40.665	自分も中華料理かなっていう/
С	00:01:02.650	00:01:06.880	なんで中華料理が良いと(D 思ったか)思ったのか/
D	00:01:08.300	00:01:08.950	ル ー/

アノテーション

■研究の最終目標:対話内容の理解・要約

◇トピックセグメンテーション:転記単位に対して複数のトピックタグを付与

◇正解要約:250文字以上・500文字以下で人手により正解要約を作成

タグ	説明
CandX	候補店1に関するのトピック
CandY	候補店2に関するのトピック
CandZ	候補店3に関するのトピック
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	売り上げに関するトピック
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
	分からない場合

アノテータ 1 アノテータ2 最終タグ 発話 必須タグ 追加タグ 必須タグ 追加タグ 追加タグ 必須タグ 追加タグ D Closed Sell Closed Sell |今回の定食はあれですよね、(D ふさいけんだった)不採算だった/ Closed そうそう/ Closed Closed \ |Closed Sell |不採算ったから潰されるってことは採算取れないと+ Closed A Closed Closed D | Closed Closed |駄目みたいな空気が凄いあるからさあ/ A Closed D Closed Closed って考えたら、採算取れる店って何なんだろうっつったら、 Exist4 A Exist4 Exist4 |ファミリープレートが一番売り上げが良い訳で+ Exist4 D Exist4 Exist4 |(L んー)/ Exist4 Meeting |ファミリープレートが+ A Exist4 A |Exist4| Exist4 |なんで良いんだろうね(Q)/ D Exist4 Exist4 メニュー、色んな店とか(? かな)/ C Exist4 Exist4 People |プラス+ C Exist4 Exist4 People |家族連れが+ People |多いって書いてあったんで、多分色んな種類あるとこに(?)+ C |Exist4 Exist4 People Exist4 A Exist4 People Exist4 A Exist4 People |(?)家族があって/ People C Exist4 Exist4 |集まるんじゃないんすかね/ People D Exist4 Exist4 People 休日はなんか消費が少なく長時間滞在、消費型って書いて D People People People |ありましたね、カップルや家族連れ/

データのURL: http://www.pluto.ai.kyutech.ac.jp/~shimada/resources.html (3月末公開予定)

ピックセグメンテーション

- ■28個のトピックタグ
- ◇4名で各対話からトピックタグを事前に洗い出し
- ■2名のアノテータによるトピックタグ付与
- ◇各転記単位に2つまで
 - ◆2名アノテータの部分一致率: 0.879
- ■著者ら3名による合議
- ◇各転記単位に3つまで
 - ◆アノテータのタグを1つ以上含んでいる率: 0.965

III 解要約

- ■AMIコーパスのガイドラインに準拠
 - ◇対話内容を知らない人が呼んだ場合でも議論の内容 を理解できるような要約
- ■アノテータ2名の要約とその総意の3つを用意

既存店の中で最も売り上げが良いファミリープレートを例とし、主に3つの候補店の値段、 客層、メニューの多さについて議論がなされた。値段については既存店の中に値段が安くも 売り上げが少ない店舗があることや値段が安くても売り上げを出している店舗があるため、 優先順位は低いとされた。また客層については、ショッピングモールの主な客層が家族連れ や女性客を多いことから、それらの客層に受けの良い店舗が優先するとされた。その結果、 ラーメンかいぶつは3つの候補店のうち値段はもっとも安いが、メニューが少なくショッピ ングモールの主な客層である女性客に受けが悪いと判断され、棄却された。最終的には既存 店に女性向けの店舗がないことを考慮し、メニューの多さよりも女性客に受けが良いかとい う点を優先した。その結果、つけ麺ふうじんとボノパスタを比べたとき、つけ麺ふうじんは メニューの豊富さではボノパスタに優っているが、ボノパスタはつけ麺ふうじんよりも女性 受けが良いと判断され、ボノパスタが採択された。(433字)

- ■自由に使える意思決定タスク対話データ
- ◇書き起こしデータとトピックタグ、正解要約
- ■作業時間
- ◇アノテータ: 1対話につき2時間半程度
- ◇トピックタグの合議: 1対話につき2時間程度
- ■今後の課題と展望
- ◇生成型の対話要約
 - ◆トピックセグメンテーションについては昨日発表済み(山村・嶋田2016)
- ◇対話データの拡充、新たなタグ付与(発話意図など)、

抽出型要約の正解データなど