

1.目的

サービス分野のクチコミと評価に関する研究の一環として、航空会社のサービスを対象とした分析を行う。

2.結論

(1)トピックモデルの1つであるLDAにより、航空会社のサービスのクチコミについてのトピックを自動分類したところ、予約の変更に関するもの、予約の際の航空会社の対応に関するもの、コロナによるキャンセルに関するもの、搭乗の際の荷物に関するものに分類できた(図表2)。トピック数はCoherence Measureによって4つの場合を選択した(図表1)。

(2)航空会社のサービスについての評価(満足度)について、順序プロビットモデルを用いて統計分析した(図表3)。説明変数にはクチコミ内容を形態素解析し頻度の高いもの、評価ごとのクチコミタイトルとクチコミ内容を形態素解析し頻度の高いものに紐付きタグ数、参考になったクリック数、気持ちアイコンIDを加えて候補として、統計的に優位なものを選んだ。

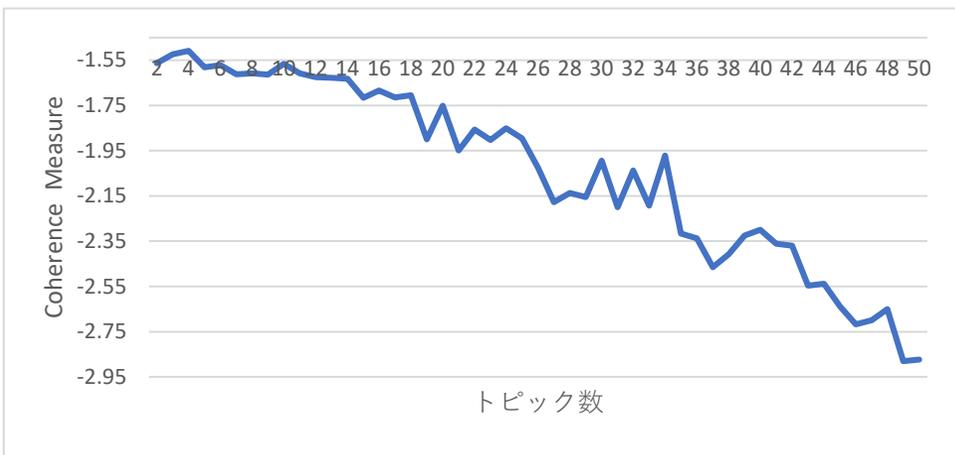
(3) 図表3で用いた説明変数を7用いて、評価(満足度)に関して、教師あり機械学習(SVM(線形)、ニューラルネットワーク、ランダムフォレスト)を試みた。データを学習に70%、テストに30%割り当てた。現データの5段階評価の場合、accuracy scoreが0.705~0.732となり満足できる結果は得られなかった(図表4以下同じ)。評価点が1に偏っていることから、評価点1と2~5の2つに分類し同様の分析をしたところ、ランダムフォレストを用いた場合accuracy scoreが0.803と向上した。

3.今後の課題

(1)トピック分類に関して、より広範囲のサービス内容をカバーしたい。

(2) 評価モデルについて、評価点2~4の文章の特徴をより捉えたい。

図表1 Coherence Measureによる最適トピック数の決定



図表2 LDA結果

	1	2	3	4	5
Topic1	予約 0.012	便 0.012	変更 0.009	前 0.009	円 0.008
Topic2	対応 0.022	利用 0.018	会社 0.014	時間 0.013	便 0.012
Topic3	キャンセル 0.019	対応 0.015	利用 0.015	コロナ 0.014	予約 0.013
Topic4	荷物 0.017	空港 0.013	対応 0.012	時間 0.012	搭乗 0.011
	6	7	8	9	10
Topic1	電話 0.008	空港 0.007	時間 0.007	対応 0.007	搭乗 0.007
Topic2	電話 0.009	飛行 0.008	予約 0.008	時 0.007	航空 0.007
Topic3	便 0.010	会社 0.009	スタッフ 0.009	返金 0.008	航空 0.008
Topic4	時 0.011	会社 0.010	利用 0.010	航空 0.010	前 0.008

図表4評価(満足度)について教師あり機械学習による分析結果

	評価満足度5段階		評価満足度2段階	
	accuracy score	f1 score	accuracy score	f1 score
SVM	0.726	0.627	0.735	0.650
ニューラルネットワーク	0.705	0.641	0.746	0.737
ランダムフォレスト	0.732	0.688	0.803	0.796

図表3 評価(満足度:5段階)について順序プロビットモデル推計結果

変数	係数	
紐付きタグ数	0.8529	***
	(0.2858)	
参考になったクリック数	-1.0965	***
	(0.3458)	
気持ちアイコンID	-0.4912	***
	(0.1186)	
利用	0.2722	***
	(0.0550)	
対応	-0.2486	***
	(0.0567)	
機内	0.3069	***
	(0.0691)	
座席	0.3116	***
	(0.0845)	
予約	-0.1822	**
	(0.0749)	
返金	-0.2456	**
	(0.1149)	
最低	-1.0795	***
	(0.2304)	
最悪	-1.0699	***
	(0.1247)	
度	-0.9514	***
	(0.1241)	
搭乗	-0.3084	*
	(0.1657)	
欠航	-0.3823	**
	(0.1816)	
遅延	-0.4201	**
	(0.1888)	
キャンセル	-1.4697	***
	(0.4205)	
電話	-0.3559	***
	(0.0894)	
変更	-0.1785	*
	(0.0934)	
安い	0.3418	*
	(0.2025)	
良かった	1.4605	***
	(0.3136)	
LCC	0.1878	**
	(0.0798)	
良い	0.3369	***
	(0.0799)	
最高	1.7418	***
	(0.3030)	
素晴らしい	1.9085	***
	(0.3617)	
快適	1.3602	***
	(0.2615)	
感謝	2.3938	***
	(0.4394)	
JAL	0.2246	***
	(0.0748)	
観測データ数	2,583	
LR chi2(19)	700.09	
有意確率	0.0000	
Pseudo R2	0.1358	

***: p < 0.01, **: p < 0.05, *: p < 0.1