

コールセンターの応対音声を対象とした声の印象の自動評価

今井 貴史[†] 村木 友子[‡] 市川 治[§]
 滋賀大学[†] ビーウィズ[‡] 滋賀大学[§]

1 序論

顧客満足度を高められる高品質なコールセンターは、顧客の離反を防ぐ上で重要な存在である。それゆえ、多くのコールセンターでは、応対品質の向上を目指し、応対品質について評価・採点を実施している。応対品質は様々な要素に分解できるが、中でもオペレータの声の印象という要素は顧客満足度に与える影響が大きい。したがって、オペレータの声の印象を評価・採点することは、実務的観点から特に重要である。しかし、現状では声の印象についての評価・採点は人の手によって行うしかなく、その作業には多大なコストが必要となる。そこで本研究では、機械学習によって声の印象評価を自動化することを試みる。

声の印象の自動評価を実用水準まで引き上げるためには、様々な要件を満たす必要がある。例えば、評価器は話者についての汎化能力をもつ必要がある。したがって、評価器が獲得する判断基準は話者に依存するようものであってはならない。また、評価の間の順序関係も考慮する必要がある。たとえ完全な正解を返すことはできなかったとしても、せめて正解に近い評価を返すことが必要である。正解と正反対の評価を下すようなことはあってはならない。

これらの要件を満たすため、本研究では、敵対的ドメイン適応 [1] およびソフト順序ベクト

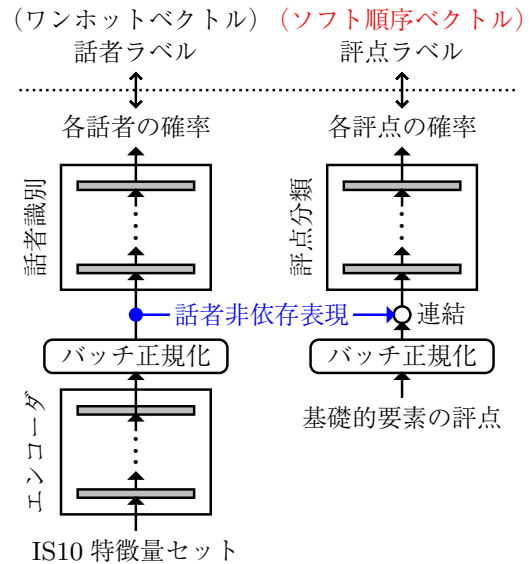


図1 評価器の構成の概略。

ル [2] を評価器に導入する。

2 評価器の構成

本研究ではニューラルネットワークを用いて評価器を構成する。構成の概略を図1に示す。評価器への入力は、オペレータの発話音声から抽出されたIS10特徴量セット [3] である。入力された特徴量はまずエンコーダネットワークで処理され、次に二つに分岐して、一方は話者識別ネットワークへ、もう一方は声の印象以外のより基礎的な要素（声の大きさや滑舌など）についての評点と連結されたうえで評点分類ネットワークへと入力される。

話者依存性をなくすため、話者識別ネットワークとエンコーダネットワーク・評点分類ネットワークとは敵対的に訓練される。すなわち、話者識別の精度を高めるように話者識別

Automated grading of call center agents' tones of voice

[†] Takashi Imai, Shiga University

[‡] Tomoko Muraki, Bewith

[§] Osamu Ichikawa, Shiga University

ネットワークは訓練され、それと同時に、評点分類の精度を高めつつ各話者の予測確率が一樣となるようにエンコーダネットワーク・評点分類ネットワークが訓練される [4].

声の印象の善し悪しは数値として表現されており、それら評点の間には順序関係が存在する。したがって、評点の推定は一種の順序回帰問題である。このような順序回帰問題に対して有効なアプローチの一つはソフト順序ベクトルの方法 [2] であり、すなわち正解ラベルをソフトラベルとして表現することでラベル同士の距離に差をもたせることである。今回はこの方法を採用し、評点をソフト順序ベクトルとして表現する。

3 実験

前節の評価器を、89 人のオペレータから取得された 2312 件の発話データに対して適用する。これらの発話データには人の手によって 1 から 3 の評点がつけられており、1 から 3 へと順に評価が高くなる。評価の客観性を高めるため、採点は複数人の協議により行われている。発話データのうち声の印象の評点が 1, 2, 3 のものは、それぞれ 400, 1350, 562 件である。また、順伝播の途中で潜在表現に連結される基礎的要素の評点について、実用化のためにはこれらの評点も自動で推定されるべきだが、今回は人の手でつけられた評点を使用する。

モデル評価のため、各話者に対し、その話者に属するデータをテスト事例、それ以外の話者に属するデータを訓練事例として検証を実施した。正解率および均衡正解率を全話者にわたって平均した値を表 1 に示す。いずれの指標においても、敵対的ドメイン適応の導入により成績が向上している。また、全話者に対する結果を合計した混同行列を、ソフト順序ベクトルの有無別に表 2 に示す。ソフト順序ベクトルの導入により、正反対の評価である 1 と 3 の間の間違いが減少している。

表 1 (a) 敵対的ドメイン適応および (b) ソフト順序ベクトル有無別の分類成績。

(a)	(b)	正解率	均衡正解率
無	無	0.606	0.530
有	無	0.646	0.564
無	有	0.627	0.553
有	有	0.664	0.590

表 2 混同行列。各行が実際の評点に対応し、各列が予測された評点に対応する。敵対的ドメイン適応は (a), (b) いずれにおいても使用されている。

(a) ソフト順序ベクトルなし (b) ソフト順序ベクトルあり

	1	2	3		1	2	3
1	141	251	8	1	191	204	5
2	148	1005	197	2	151	992	207
3	6	273	283	3	2	244	316

4 結論

声の印象の自動評価において敵対的ドメイン適応およびソフト順序ベクトルが有効であることを確認した。敵対的ドメイン適応を用いることで、予測の話者依存性を排除することができ、また、訓練ラベルをソフトラベル化して評点の順序関係を考慮することで、正解と正反対の評価を下すような間違いを減らすことができる。これらの技術は、声の印象の自動評価を実用水準まで引き上げるために特に重要である。

参考文献

- [1] Ganin, Y. et al.: *Journal of Machine Learning Research*, Vol. 17, No. 59, pp. 1–35 (2016).
- [2] Díaz, R. and Marathe, A.: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4738–4747 (2019).
- [3] Schuller, B. et al.: *Proceedings of INTER-SPEECH 2010*, pp. 2794–2797 (2010).
- [4] Li, H. et al.: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7144–7148 (2020).