

Politeness分析を用いた OSSコミュニティのコミュニケーションの理解

宮崎 智己

背景

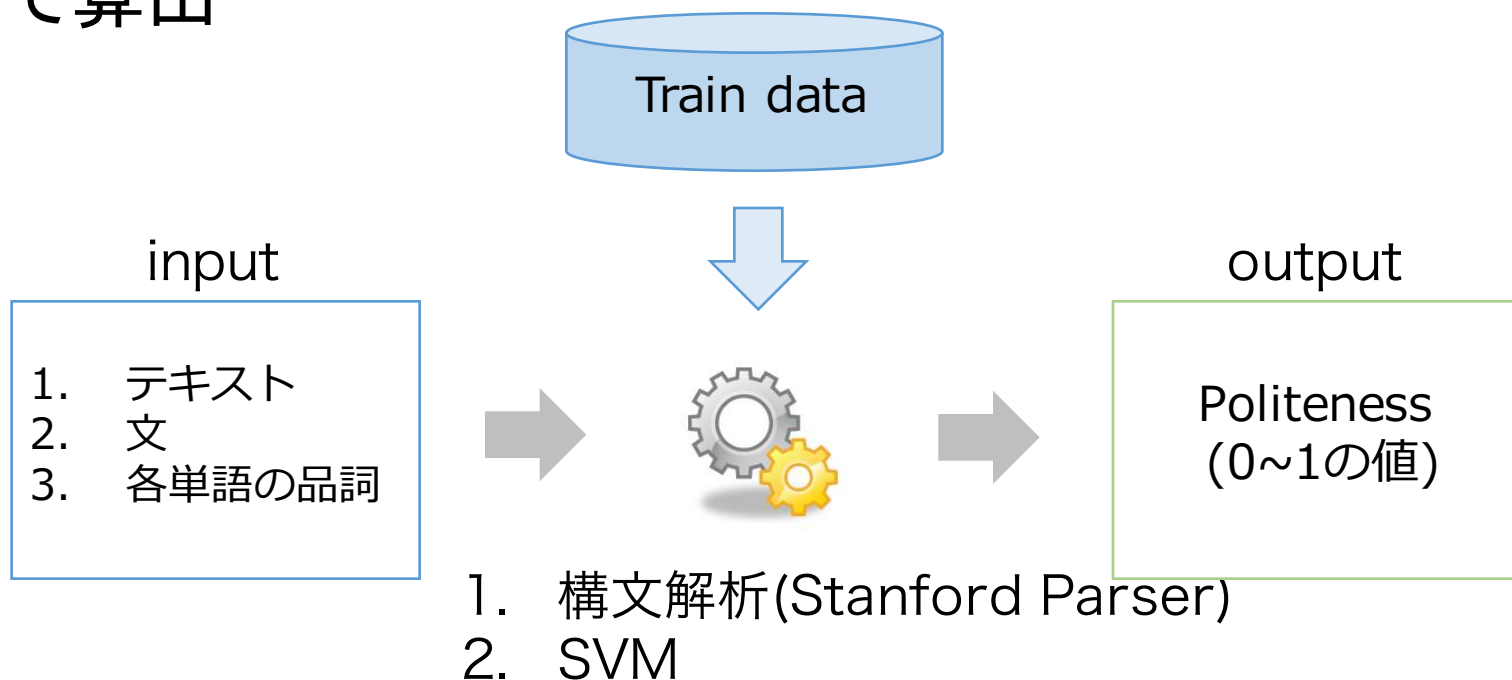
- OSSの開発には多くの開発者が開発に携わっており、開発者間のコミュニケーションは必要不可欠なもの
- 開発者間のコミュニケーションの質は開発者にどのような影響を与えるのかを知りたい

Politenessとは

- 会話の参加者がお互いのフェイスを侵さないために行う言語的配慮
- 会話の「誠実さ」や「丁寧さ」
- 1987年にBrownとLevinsonによって確立

Politeness解析ツール

- テキストデータの前処理を行ったあと, SVMを用いて算出



学習データ

- テキストデータの取得
 - Wikipedia:編集者の会話ページ
 - StackExchanged:Yahoo知恵袋の海外版
- Amazon Mechanical Turkを用いラベル付け
 - 異なる5人が, 各リクエストに対しPoliteness値をつける
 - 5つの値を正規化したものをPolitenessとする

domain	#requests	#annotated	#annotators
Wiki	35,661	4,353	219
SE	373,519	6,604	212

Table 1: Summary of the request data and its politeness annotations.

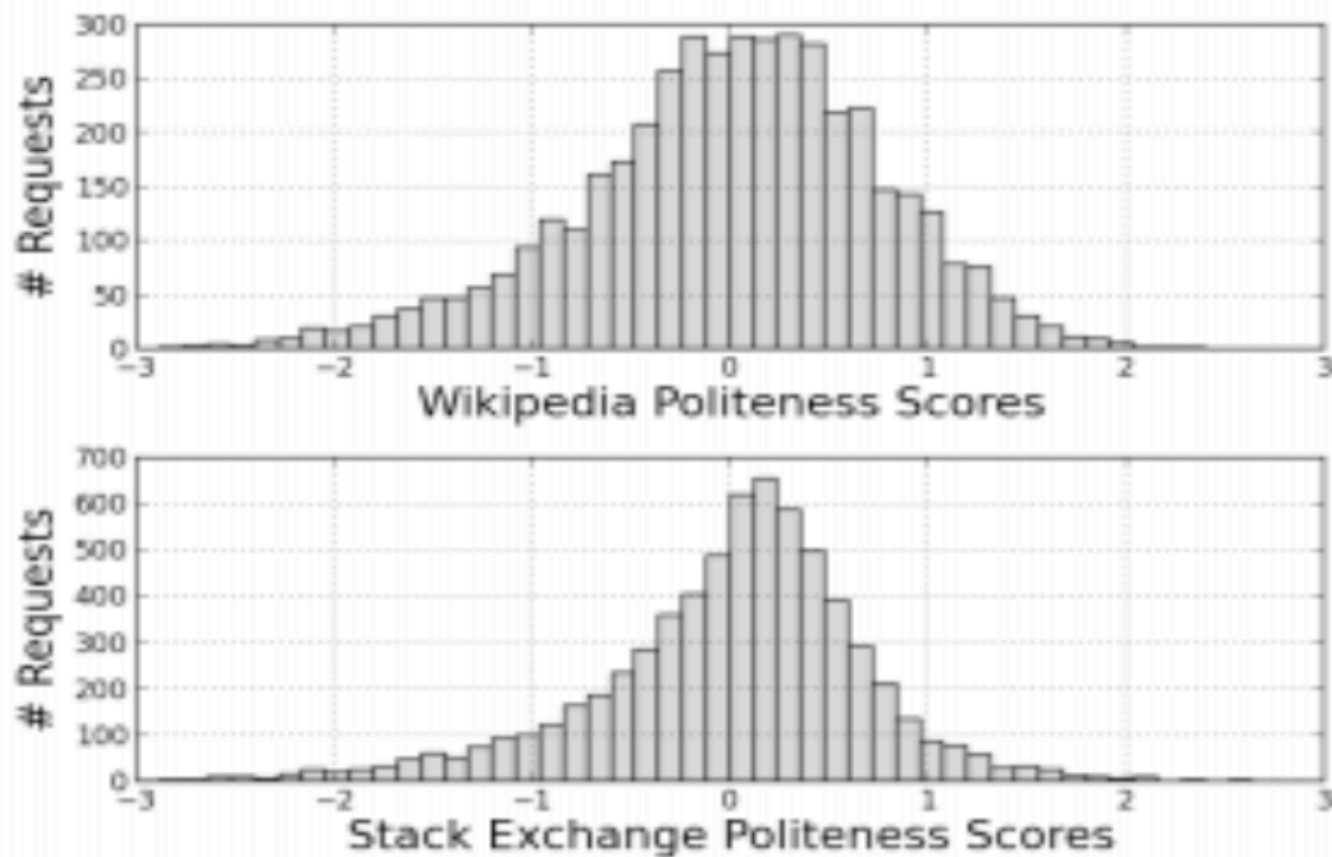


Figure 1: Distribution of politeness scores. Positive scores indicate requests perceived as polite.

Strategy	Politeness	In top quartile	Example
1. Gratitude	0.87 ^{***}	78% ^{***}	I really appreciate that you've done them.
2. Deference	0.78 ^{***}	70% ^{***}	Nice work so far on your rewrite.
3. Greeting	0.43 ^{***}	45% ^{***}	Hey , I just tried to ...
4. Positive lexicon	0.12 ^{***}	32% ^{***}	Wow! / This is a great way to deal...
5. Negative lexicon	-0.13 ^{***}	22% ^{**}	If you're going to accuse me ...
6. Apologizing	0.36 ^{***}	53% ^{***}	Sorry to bother you ...
7. Please	0.49 ^{***}	57% ^{***}	Could you please say more...
8. Please start	-0.30 [*]	22%	Please do not remove warnings ...
9. Indirect (btw)	0.63 ^{***}	58% ^{**}	By the way , where did you find ...
10. Direct question	-0.27 ^{***}	15% ^{***}	What is your native language?
11. Direct start	-0.43 ^{***}	9% ^{***}	So can you retrieve it or not?
12. Counterfactual modal	0.47 ^{***}	52% ^{***}	Could/Would you ...
13. Indicative modal	0.09	27%	Can/Will you ...
14. 1st person start	0.12 ^{***}	29% ^{**}	I have just put the article ...
15. 1st person pl.	0.08 [*]	27%	Could we find a less complex name ...
16. 1st person	0.08 ^{***}	28% ^{***}	It is my view that ...
17. 2nd person	0.05 ^{***}	30% ^{***}	But what's the good source you have in mind?
18. 2nd person start	-0.30 ^{***}	17% ^{**}	You've reverted yourself ...
19. Hedges	0.14 ^{***}	28%	I suggest we start with ...
20. Factuality	-0.38 ^{***}	13% ^{***}	In fact you did link, ...

Positive
politeness

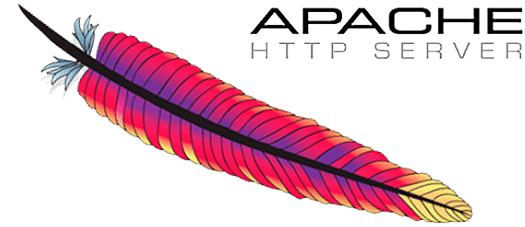
Negative
politeness

解析した例

例文	Politeness 値
OK thanks for your suggestion. I will change that. Regards XXX.	0.997
Oh dear. My question is already slipping down the list into obscurity... I guess I must be the only person using eclipse 3.4 who would like to be able to do this.	0.507
So, what was the solution for this problem then??	0.122

アプローチ

- 対象プロジェクト
 - Apache httpd



- 研究方法
 - 開発者のメーリングリストから、多くの開発者に返信している開発者を絞り込み、その開発者のPoliteness値に変動があるかを見る

Politeness解析までの手順

1. 開発者メーリングリストのmboxファイルからメールの内容をデータベースに格納
2. 対象の開発者の絞り込み(メトリクスは考え中)
3. 引用など邪魔なテキストの消去
4. 解析