

Sentiment Analysis in Tickets for IT Support



Cássio Castaldi Araujo Blaz
Instituto de Informática
Universidade Federal do Rio
Grande do Sul Porto Alegre, Brazil



Karin Becker
Instituto de Informática
Universidade Federal do Rio
Grande do Sul Porto Alegre, Brazil

[MSR '16](#) Proceedings of the 13th International Conference on Mining
Software Repositories Pages 235-246

背景

- Sentiment Analysisはソフトウェア工学の分野で認められてきている
- 開発者のSentimentとソフトウェア開発の間に関係があるといわれている
- しかし、多くの研究は一般的な分析を行っているため異なった結果が出ている
 - ソフトウェア工学ならではの分析を行っていないからではないか？

目的とアプローチ

- 目的
 - チケットに含まれているSentimentの評価手法の提案
- アプローチ
 - Domain Dictionaryの自動生成
 - 3つの手法を提案
 - 既存のツール/提案手法の比較

チケット

- ユーザ要求をドキュメントにしたもの

Table 2: Types of Tickets and Examples

Type	Subtype	Example
Issue	Infrastructure	<i>"I'm having trouble sending external emails, I urgently need help, because I need to send commercial proposals."</i>
	Software	<i>"Good morning, kindly check why when I open the certificate it is not stating the covered contract"</i>
Request	Data Extraction	<i>"Please send us the list of cancelled contributions."</i>
	System Maintenance	<i>"We need to create a report of payment for a given period. Today we looked one by one to generate it manually."</i>
	Infrastructure	<i>"Install CRM on my computer."</i>
	Administrative	<i>"Register John's work hours immediately."</i>
Feedback	IT Area	<i>"Excellent support of the team involved with the environment maintenance."</i>
	Software	<i>"What about automate importing data from the old client system?"</i>
	Infrastructure	<i>"Enabling the submission of more jobs without compromising the cluster can help people not delay in the pool."</i>

チケットの構造

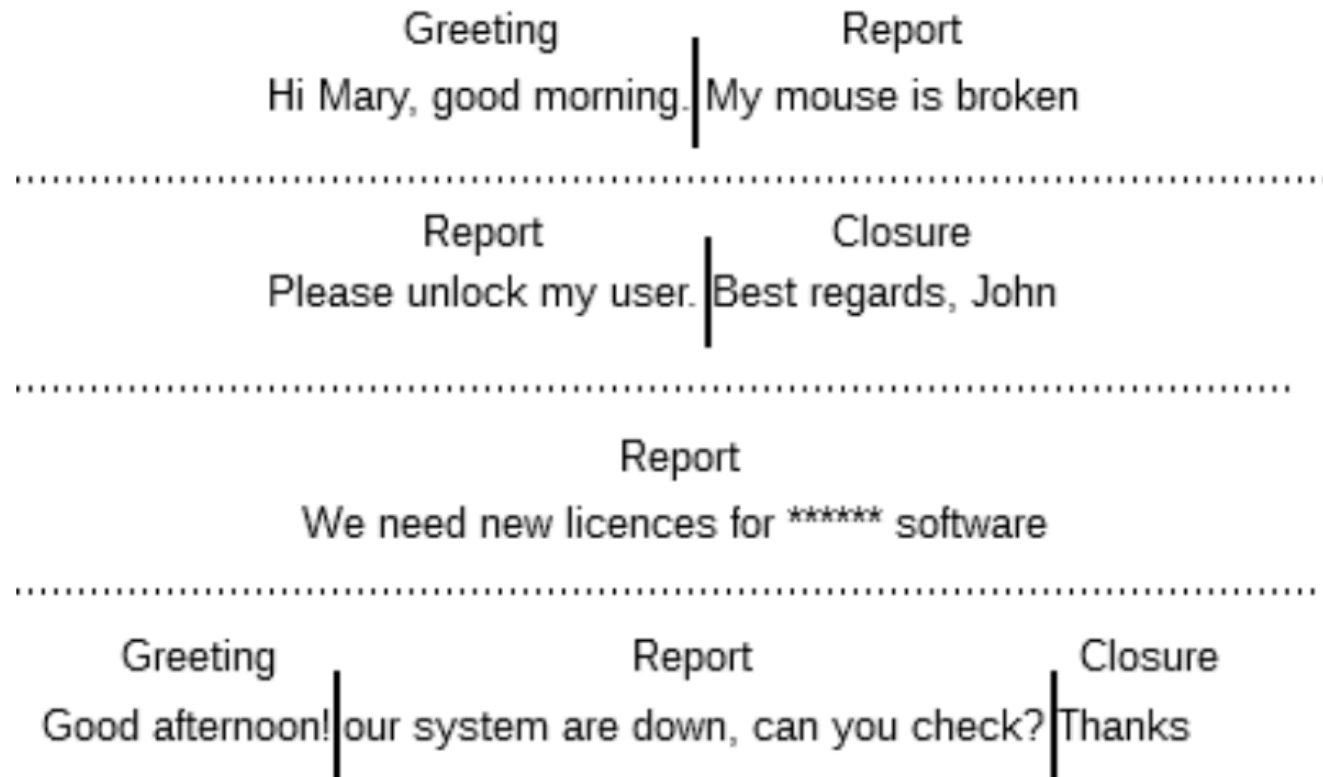


Figure 1: Examples of tickets and their structure

Domain Dictionary

- Sentimentの性質を持った単語の辞書
- 類義語, N-gramにも対応
- 生成方法
 1. チケットからSentimentの性質を持った単語を抽出
 - SentiWordNetを適用
 2. 単語の性質の判別 (Positive/Negative/Neutral)
 3. 単語の位置の判定 (Greeting/Report/Closure)
 4. 単語の品詞を判別

提案手法

- Dictionary method (DM)
 - 各単語の意味からpositive/negative/neutralを判別
 - SentiWordNetを用い単語ごとの値の総和を算出
 - SentiWordNetにない単語はdefault scoreが割り当てられる
- Template method (TM)
 - 各単語の位置と性質 (positive/negative/neutral) から6つに分類分け
 - Category score (分類ごとの値) の総和を算出
- Hybrid method (HM)
 - DMとTMの合わせ技

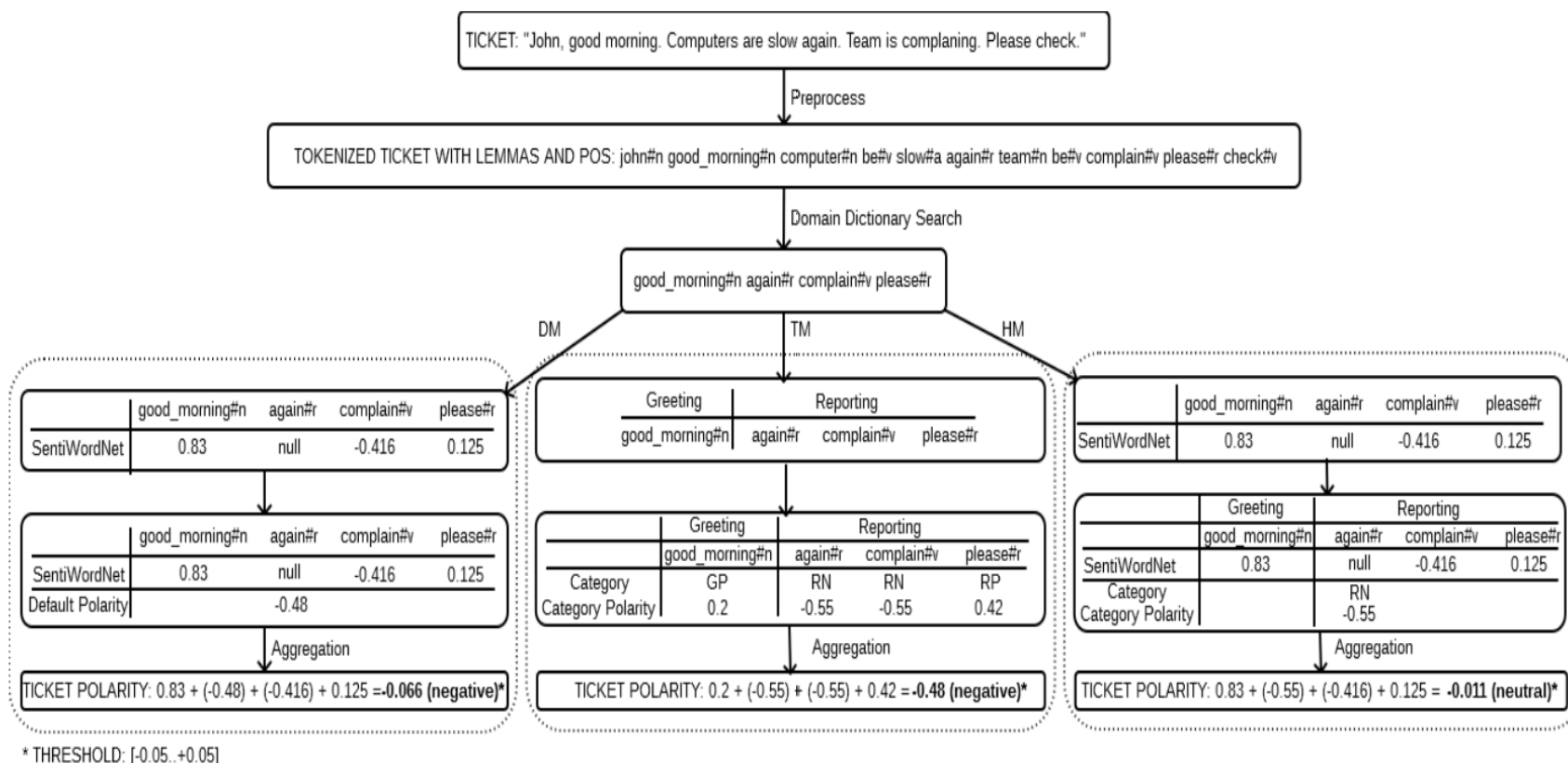


Figure 4: Sentiment Analysis Illustration according to DM, TM and HM Methods

データセット

- 5つの企業から計34,895件のチケットを取得
- 34,895件からランダムに2,333件のチケットを選択
- 選択されたチケットに対し, 3人の専門家がSentimentを判別

Table 5: Gold Standard Inter-Annotator Agreement

Annotator	Positive	Negative	Neutral	Disagreement
A	273	154	1942	B: 12%
				C: 10%
				BC: 7%
B	358	172	1803	A: 12%
				C: 21%
				AC: 7%
C	187	67	2079	A: 10%
				B: 21%
				AB: 7%
Total	233	105	1995	

評価実験

- SentiWordNetとSentiStrengthの比較
 - Dictionary methodを適用した場合で比較
- 提案した3つの手法をデータセットに適用し
precisionとrecallを算出

結果

SentiWordNetとSentiStrengthの比較

	Baseline	
	Senti WordNet	Senti Strength
Accuracy	25%	71%
Positive Precision	17%	53%
Negative Precision	07%	08%
Neutral Precision	100%	94%
Positive Recall	91%	78%
Negative Recall	50%	37%
Neutral Recall	16%	71%
Positive F-measure	29%	63%
Negative F-measure	12%	14%
Neutral F-measure	27%	81%

SentiwordNetはNeutral Precisionが100%だったが、Positive Precision はSentiStrengthと比べて非常に精度が低い

結果

提案手法の比較

Table 7: Comparison of DM, TM, and HM Methods

Metric	DM	TM	HM
Positive Precision	83.00(0.82)	82.50(1.73)	83.75(0.50)
Negative Precision	66.75(2.06)	66.25(2.87)	65.00(3.56)
Neutral Precision	98.75(0.50)	98.50(0.58)	98.50(0.58)
Positive Recall	95.75(1.50)	94.00(2.45)	92.75(1.89)
Negative Recall	73.50(4.12)	72.00(4.08)	75.00(4.69)
Neutral Recall	96.00(0.00)	96.00(0.00)	96.00(0.00)
Positive F-measure	89.00(0.00)	87.50(0.58)*	88.00(0.82)
Negative F-measure	70.00(2.16)	69.00(0.82)	69.75(0.96)
Neutral F-measure	97.00(0.00)	97.00(0.00)	97.00(0.00)
Average Precision	82.83(13.70)	82.42(13.87)	82.42(14.44)
Average Recall	88.42(11.25)	87.33(11.63)	87.92(10.00)
Average F-measure	85.33(11.88)	84.50(12.15)*	84.92(11.86)

他の二つと比べてTemplate methodの精度がわずかに低かったが、どの手法も精度が高かった

考察

- Negativeチケットへの精度が低いのは、チケットの数が著しく少ないからではないか
 - Negativeは全体の4.5%
 - Neutralは全体の85.5%を占めており、精度もかなり高い
- 精度の低下には、Domain Dictionaryに含まれる単語が少ないことが考えられる
 - 実際に含まれていない単語があった

まとめと所感

- まとめ

- Domain Dictionaryの自動生成はITチケットに対して効果的である
- DM (Dictionary method) と HM (Hybrid method) は TM (Template method) よりもわずかに精度が高かった

- 所感

- Template method で用いられていた Category score はさすがにざっくりすぎ
- 結果はそうなるだろうなあと思った

Dictionary method (辞書ベース)

1. それぞれの単語がDomain Dictionaryに含まれているかを判定
 - 含まれていれば, positive/negativeに判別
 - 含まれていなければneutral
2. SentiWordNetを用いて値を算出
 - SentiWordNetに含まれていなければ, Default scoreを割り当て
3. 出た値の総和を計算する

Template method (構造ベース)

- トークンの位置と極性 (positive/negative) で分類分け
 - 分類
 - Greeting Positive
 - Greeting Negative
 - Report Positive
 - Report Negative
 - Closure Positive
 - Closure Negative
 - 分類ごとに値を設定

Template method (構造ベース)

1. 単語の位置を判別し分類分け
 - チケットの最初のトークンをgreeting
 - 最後のトークンをclosure
 - その間に含まれているトークンをreport
2. それぞれの単語がpositiveかnegativeかを判別し再度分類分け
3. 分類ごとの値を算出
4. 出た値の総和を計算