

単語の類似性を考慮した セキュリティ要件分類手法の構築

オープンソースソフトウェア工学研究室

宮崎智己

背景

- 近年の情報システム開発は委託開発が主流
- ベンダーは委託者と要件定義を行い，システムが満たすべき機能要件と非機能要件を要件定義書にまとめる
- 開発者は要件定義書を目視して機能要件や非機能要件を特定し，設計や実装，テストを実施する
 - 多くの場合，各工程を担当する開発者は異なる
 - 各開発者がそれぞれ目視する必要がある

非機能要件の特定の問題点

- 非機能要件の多くは機能要件に含まれる[1]
- 機能要件と比べて曖昧になりやすい[2]
- 同じ内容の要件であっても開発者によって表現方法が異なる[3]

時間的コストがかかる
見落としをしてしまう

[1]Eckhardt, J., Vogelsang, A. and Fern´andez, D. M.: Are "Non-functional" Requirements Really Non-functional?: An Investigation of Non-functional Requirements in Practice, Proc. of the 38th International Conference on Software Engineering (ICSE '16), pp. 832-842 (2016).

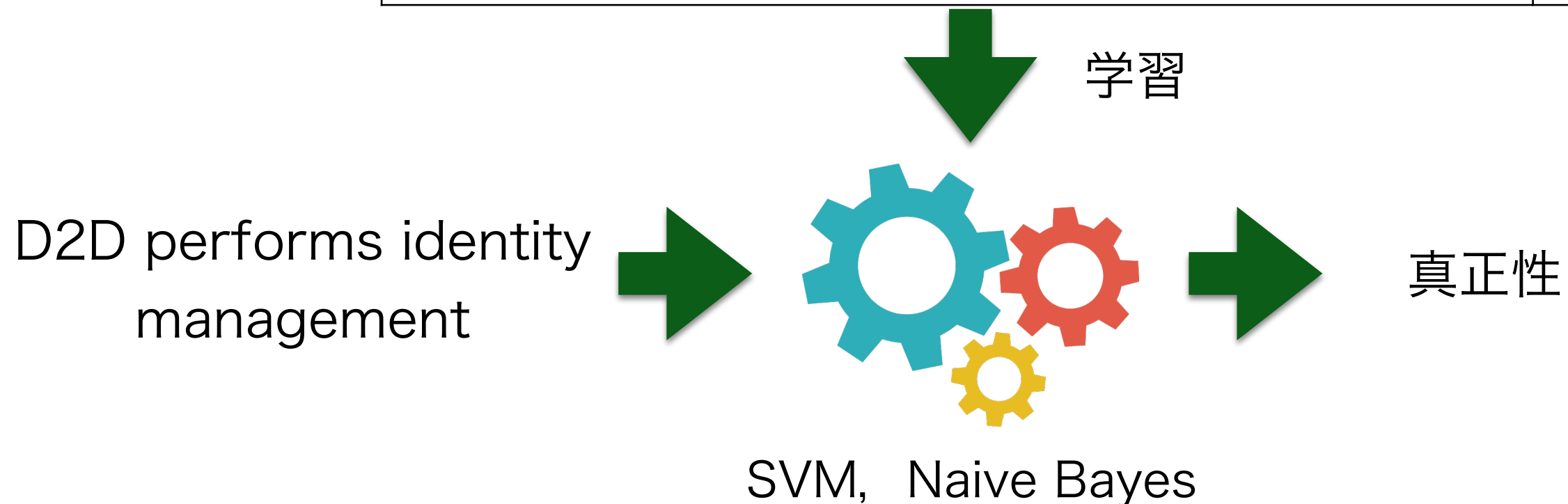
[2]Borg, A., Yong, A., Carlshamre, P. and Sandahl, K.: The Bad Conscience of Requirements Engineering: An In- vestigation in Real-World Treatment of Non-Functional Requirements, Proc. of the 3rd Conference on Software Engineering Research and Practice in Sweden (SERPS '03), pp. 1-8 (2003).

[3]Abad, Z. S. H., Shymka, A., Pant, S., Currie, A. and Ruhe, G.: What are Practitioners Asking about Require- ments Engineering? An Exploratory Analysis of Social Q A Sites, Proc. of IEEE 24th International Require- ments Engineering Conference Workshops (REW '16), pp. 334-343 (2016).

非機能要件分類手法[4,5]

- 教師あり学習を用いた非機能要件分類手法が提案されている

Allows role-based access to EHR system	機密性
Data used during user authentication	真正性
⋮	⋮



[4]Wen Zhang, Ye Yang, Qing Wang, Fengdi Shu, An Empirical Study on Classification of Non-Functional Requirements, Proceedings of the 23rd International Conference on Software Engineering & Knowledge Engineering (SEKE'2011)

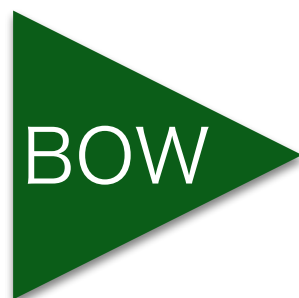
[5]John Slankas and Laurie Williams, Automated Extraction of Non-functional Requirements in Available Documentation, 1st International Workshop on Natural Language Analysis in Software Engineering (NaturaLiSE), 2013

既存手法の問題点

- 類義語に対応することができない
 - 意味的には同じ単語を全く違う単語として扱っている

(1) **login** as oscar or mac2002 or 1117.

(2) **logon** as oscar or mac2002 or 1117.



	login	logon	as	• •	mac2002
(1)	1	0	0	• •	1
(2)	0	1	0	• •	1

同じ意味を持つフレーズだが、
異なるフレーズとして判断している

既存手法の問題点

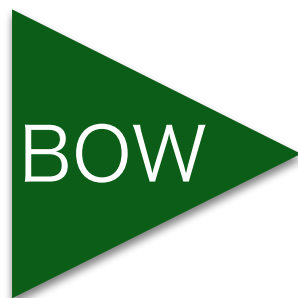
- 単語間の順序関係を考慮しておらず、イディオムなどの複合語を考慮できていない

(1)log in as oscar or mac2002 or 1117.

(2)login as oscar or mac2002 or 1117.

(3)log into as oscar or mac2002 or 1117.

(4)All POS systems connected to the EHRi must record in an audit log.



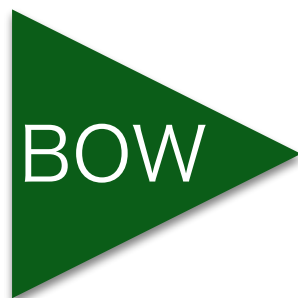
	in	log	into	• •	login
(1)	1	1	0	• •	0
(2)	0	0	0	• •	1
(3)	0	1	1	• •	0
(4)	1	1	0	• •	0

既存手法の問題点

- 単語間の順序関係を考慮しておらず、イディオムなどの複合語を考慮できていない

(1) **log in** as oscar or mac2002 or 1117.

(4) All POS systems connected to the EHRi must record **in** an audit **log**.



	in	log	into	• •	login
(1)	1	1	0	• •	0
(2)	0	0	0	• •	1
(3)	0	1	1	• •	0
(4)	1	1	0	• •	0

同じものと判断

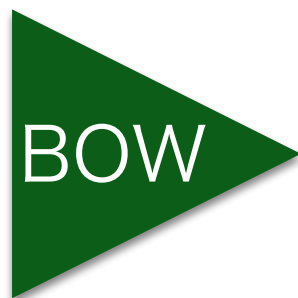
既存手法の問題点

- 単語間の順序関係を考慮しておらず、イディオムなどの複合語を考慮できていない

(1) **log in** as oscar or mac2002 or 1117.

(2) **login** as oscar or mac2002 or 1117.

(3) **log into** as oscar or mac2002 or 1117.



	in	log	into	• •	login
(1)	1	1	0	• •	0
(2)	0	0	0	• •	1
(3)	0	1	1	• •	0
(4)	1	1	0	• •	0

表記揺れや類義イディオムを考慮できていない

本研究の目的・アプローチ

- 目的

- 非機能要件分類手法の精度向上
 - 本研究では、セキュリティ要件の詳細な分類を対象

- アプローチ

- Nグラムを用いたBOWの作成
- 分散表現を用いた類義語・類義フレーズへの対応

Nグラムを用いたBOWの作成

- Nグラムとは文章を任意のN個の単語に分割する手法
 - 既存手法では1グラムを用いている

(1) log in as oscar or mac2002 or 1117.

1gram

“log” “in” “as” “oscar” “or” “mac2002” “or” “1117”

2gram

“log in” “in as” “as oscar” “oscar or” “or mac2002”
“mac2002 or” “or 1117”

3gram

“log in as” “in as oscar” “as oscar or” “oscar or
mac2002” “or mac2002 or” “mac2002 or 1117”

Nグラムを用いたBOWの作成

- Nグラムとは文章を任意のN個の単語に分割する手法
 - 既存手法では1グラムを用いている

(1) log in as oscar or mac2002 or 1117.

1gram

	log	in	as	• •	1117
(1)	1	1	1	• •	1

2gram

	log in	in as	as oscar	•	or 1117
(1)	1	1	1	•	1

3gram

	log in as	in as oscar	• •	mac2002 or 1117
(1)	1	1	• •	1

Ngramを用いた手法の問題点

- あまり連続して出ないような単語列も抽出してしまう
 - 計算コストがかかる
- あらかじめ定義したN個の連続分しか単語列を抽出することができない
 - 例えば $N > 1$ の時，単語一つが判断基準となっている場合に，その単語一つを特徴量として扱うことができない

Multiword Expression Distance(MED)[6]

.....

- コルモゴロフ複雑性に基づいた単語間結合度を測定するアルゴリズム

$$MED(g) = \log \frac{df(w_1, w_2, \dots, w_n)}{df(g)}$$

$MED(g)$: Nグラムgの値

$df(w_1, w_2, \dots, w_n)$: Nグラムgを構成する単語を全て含む文書数

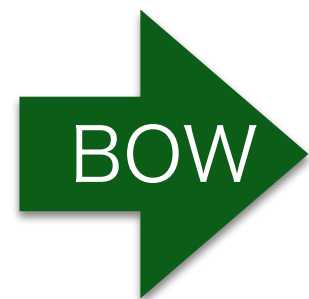
$df(g)$: Nグラムgを含む文書数

- MEDの値が大きいNグラムは, 不自然なNグラムであるとみなせる

MEDを適用した場合

- 1 gramだけでなく, 2gramや3gramの情報も利用する

(1) log in as oscar or mac2002 or 1117.



	log	in	• •	log in	log in as
(1)	1	1	• •	1	1

評価実験

- 単語の順序関係が精度向上につながるかを明らかにする
 - 既存研究で用いられていたナイーブベイズとSVMを学習モデルに用いる
 - 1gram（既存手法），2gram，3gram，N-gram（MED）の場合の精度を比較
 - N-gramは $N \leq 3$ ，MED $\leq \log(10/1)$ のNgramを追加
- 10交差検証を実施
 - Precision，Recall，F値で評価

データセット

.....

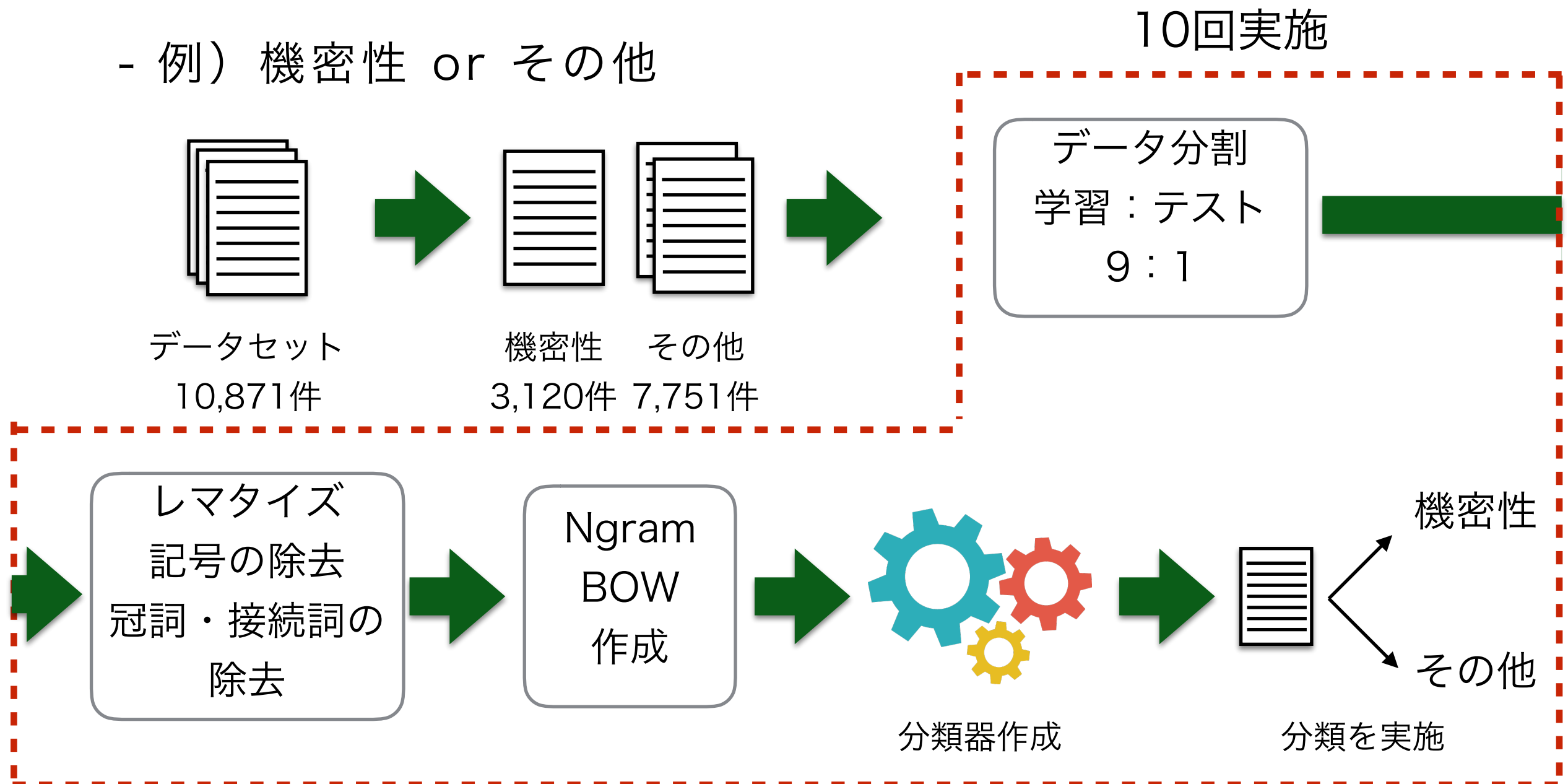
- 5つの医療系システムの要件定義書・機能要求
 - 1つの文に複数のセキュリティ要件が含まれる場合がある

セキュリティ要件	件数
機密性 (CONFIDENTIALITY)	3,120
インテグリティ (INTEGRITY)	3,386
可用性 (AVAILABILITY)	174
責任追跡性 (ACCOUNTABILITY)	3,882
真正性 (Identification & Authentication)	311
プライバシー (PRIVACY)	246
Not Security (その他)	6,617
total	10,871

実験手順

- 各セキュリティ要件の2値分類器を作成し10交差検証をそれぞれ実施

- 例) 機密性 or その他



ナイーブベイズの実験結果

- ・ 機密性の場合

N-gram	Precision	Recall	F-score
1gram	0.71	0.76	0.73
2gram	0.74	0.80	0.77
3gram	0.79	0.73	0.76
Ngram	0.74	0.76	0.75

ナイーブベイズの実験結果

- 真正性の場合

N-gram	Precision	Recall	F-score
1gram	0.42	0.34	0.37
2gram	0.14	0.58	0.22
3gram	0.08	0.77	0.14
Ngram	0.83	0.38	0.51

SVMの実験結果

- インテグリティの場合

N-gram	Precision	Recall	F-score
1gram	0.79	0.77	0.78
2gram	0.86	0.71	0.78
3gram	0.91	0.54	0.68
Ngram	0.81	0.78	0.80

SVMの実験結果

- 真正性の場合

N-gram	Precision	Recall	F-score
1gram	0.70	0.58	0.63
2gram	0.86	0.38	0.52
3gram	0.97	0.32	0.47
Ngram	0.76	0.53	0.62

実験結果まとめ

.....

- Ngram (MED) を用いた時
 - ナイーブベイズの場合：F1 値が+0.01～+0.14の誤差
 - SVMの場合：F1 値が-0.01～+0.02の誤差
 - SVMの場合，特徴量が増えすぎてきれいに分類できる識別面を作るのが困難になるため？
- 1gramと比べて2,3gramと増やすにつれてPrecisionがプラスにRecallはマイナスになる
 - 2gramや3gramの時は，できる特徴量が限定的になるためハズレは少なくなるが見落としが増える
 - Ngram (MED) の時がマシなのは，1グラムの情報も扱っているため

今後の予定

- 用いるN-gramの設定方法の改善
 - 今回はMEDの閾値を設定して閾値以下のNgramを適用した
 - Ngram IDFや出現回数による方法などはまだ試せていない

N-gram IDF

$$\begin{aligned} IDF_{N-gram}(g) &= IDF(w_1, \dots, w_N) - MED(g) \\ &= \log \frac{|D| \cdot df(g)}{df(w_1, \dots, w_N)^2} \end{aligned}$$

- 頻繁に出現するNgramの重みを低くすることができる
 - 特に頻出単語のみで構成されるようなNgramは分母が大きくなる

例) Ngram = “to in”

df(to,in)=812, df(to in)=3, D=10000の時

IDF(to, in)=log(10000*3/(812*812))=-3.09

目標

- 類義語, 類義フレーズを同一として扱う処理を追加する
 - 以下の赤字を同じ意味として扱いたい
 - (1) **log in** as oscar or mac2002 or 1117.
 - (2) **login** as oscar or mac2002 or 1117.
 - (3) **log into** as oscar or mac2002 or 1117.
- Ngram ($N > 1$) を上手くマージして分散表現手法を適用したい

Ngramのマージの方法

- スペースをなくした時に1 gramの単語と一致する場合はマージ

○ “log in” → “login”

✕ “audio log” → “auditlog”

- Ngramの構成単語が共通する場合はMEDもしくはNgram-IDFの小さい方をマージする

log into as oscar or mac2002 or 1117.

- “log into”と”into as”があった場合、値を比較

“loginto” as oscar or mac2002 or 1117.

補足

ナイーブベイズの実験結果

.....

- インテグリティの場合

N-gram	Precision	Recall	F-score
1gram	0.71	0.79	0.75
2gram	0.75	0.83	0.79
3gram	0.76	0.77	0.77
Ngram	0.74	0.79	0.76

ナイーブベイズの実験結果

- 可用性の場合

N-gram	Precision	Recall	F-score
1gram	0.10	0.07	0.08
2gram	0.06	0.37	0.10
3gram	0.04	0.66	0.08
Ngram	0.54	0.06	0.10

ナイーブベイズの実験結果

- 責任追跡性の場合

N-gram	Precision	Recall	F-score
1gram	0.76	0.85	0.80
2gram	0.83	0.84	0.84
3gram	0.90	0.70	0.79
Ngram	0.80	0.83	0.81

ナイーブベイズの実験結果

- プライバシーの場合

N-gram	Precision	Recall	F-score
1gram	0.25	0.30	0.27
2gram	0.09	0.51	0.16
3gram	0.06	0.78	0.12
Ngram	0.39	0.27	0.32

SVMの実験結果

.....

- 機密性の場合

N-gram	Precision	Recall	F-score
1gram	0.78	0.75	0.76
2gram	0.85	0.70	0.77
3gram	0.91	0.51	0.66
Ngram	0.81	0.77	0.79

SVMの実験結果

- 可用性の場合

N-gram	Precision	Recall	F-score
1gram	0.44	0.34	0.38
2gram	0.74	0.21	0.33
3gram	0.68	0.15	0.23
Ngram	0.56	0.29	0.38

SVMの実験結果

.....

- 責任追跡性の場合

N-gram	Precision	Recall	F-score
1gram	0.83	0.82	0.82
2gram	0.90	0.75	0.82
3gram	0.96	0.57	0.71
Ngram	0.85	0.83	0.84

SVMの実験結果

- プライバシーの場合

N-gram	Precision	Recall	F-score
1gram	0.43	0.32	0.36
2gram	0.66	0.22	0.33
3gram	0.49	0.07	0.12
Ngram	0.52	0.31	0.38