特徴重み付けを用いた
低・ゼロ頻度N-gramに対する尤度比の推定法

菊地 真人, 吉田 光男, 梅村 恭司
（豊橋技術科学大学）

※ 本資料は下記のサイトからもご覧いただけます。

DEIM 2020 F8-4(day2 p55)

http://www.ss.cs.tut.ac.jp/~kikuchi14/

尤度比とは？

- 尤度比
  - 確率分布の比で表され、二値分類や統計検定でよく用いられる

$$r(x) = \frac{p_{\mathrm{nu}}(x)}{p_{\mathrm{de}}(x)}$$

x は単語列や文字列などの離散要素

  - 尤度比 $r(x)$ が大きいほど x は確率分布 $p_{\mathrm{nu}}(x)$ から生成されやすく、小さいほど $p_{\mathrm{de}}(x)$ から生成されやすい

- 尤度比の実用には推定が必要
  - アプリケーションに影響を及ぼす重要なプロセス
  - **本研究では、x の観測頻度から尤度比を推定する**

尤度比の素朴な推定法

- 確率分布を相対頻度で求め、その比を取ること

$$r_{\mathrm{MLE}}(x) = \frac{\hat{p}_{\mathrm{nu}}(x)}{\hat{p}_{\mathrm{de}}(x)} \qquad \hat{p}_*(x) = \frac{f_*(x)}{\sum_x f_*(x)}, \quad *\in \{\mathrm{de}, \mathrm{nu}\}$$

  - $f_*(x)$ は確率分布 $p_*(x)$ から得た x の頻度

- $\sum_x f_{\mathrm{de}}(x_{\mathrm{A}}) = 10^7$, $f_{\mathrm{de}}(x_{\mathrm{A}}) = 5000$, $\sum_x f_{\mathrm{nu}}(x_{\mathrm{A}}) = 10^4$, $f_{\mathrm{nu}}(x_{\mathrm{A}}) = 100$

$$\hat{p}_{\mathrm{de}}(x_{\mathrm{A}}) = \frac{f_{\mathrm{de}}(x_{\mathrm{A}})}{\sum_x f_{\mathrm{de}}(x_{\mathrm{A}})} = \frac{5000}{10^7} \qquad \hat{p}_{\mathrm{nu}}(x_{\mathrm{A}}) = \frac{f_{\mathrm{nu}}(x_{\mathrm{A}})}{\sum_x f_{\mathrm{nu}}(x_{\mathrm{A}})} = \frac{100}{10^4} \qquad r_{\mathrm{MLE}}(x_{\mathrm{A}}) = \frac{10^{-2}}{5 \times 10^{-4}} = 20$$

素朴な推定法の問題点①

- 要素 x_{B} から x_{D} についても、推定量 $r_{\mathrm{MLE}}(x)$ を計算

x	頻度				$r_{\mathrm{MLE}}(x)$
	$\sum_x f_{\mathrm{de}}(x)$	$f_{\mathrm{de}}(x)$	$\sum_x f_{\mathrm{nu}}(x)$	$f_{\mathrm{nu}}(x)$	
x_{A}	10^7	5000	10^4	100	20
x_{B}	10^7	50	10^4	1	20
x_{C}	10^7	50	10^4	2	40
x_{D}	10^7	14	10^4	0	0

大差

偶然かも

- **① 低頻度に対する推定値は変動が大きく不安定**
  - $f_*(x_{\mathrm{A}})$ と $f_*(x_{\mathrm{B}})$ は大差あるが、要素 x_{A} と x_{B} の推定値は20と等しい

    ⇒ 頻度差を推定値に反映したい

素朴な推定法の問題点①

- 要素 x_{B} から x_{D} についても、推定量 $r_{\mathrm{MLE}}(x)$ を計算

x	頻度				$r_{\mathrm{MLE}}(x)$
	$\sum_x f_{\mathrm{de}}(x)$	$f_{\mathrm{de}}(x)$	$\sum_x f_{\mathrm{nu}}(x)$	$f_{\mathrm{nu}}(x)$	
x_{A}	10^7	5000	10^4	100	20
x_{B}	10^7	50	10^4	1	20
x_{C}	10^7	50	10^4	2	40
x_{D}	10^7	14	10^4	0	0

差1

大きく変動

- **① 低頻度に対する推定値は変動が大きく不安定**
  - $f_{\mathrm{nu}}(x_{\mathrm{B}})$ と $f_{\mathrm{nu}}(x_{\mathrm{C}})$ の差は 1 しかないが、推定値は20と40で大きく変動

    ⇒ <u>低頻度に対する推定値をロバストにしたい</u>

素朴な推定法の問題点②

- 要素 x_{B} から x_{D} についても、推定量 $r_{\mathrm{MLE}}(x)$ を計算

x	頻度				$r_{\mathrm{MLE}}(x)$
	$\sum_x f_{\mathrm{de}}(x)$	$f_{\mathrm{de}}(x)$	$\sum_x f_{\mathrm{nu}}(x)$	$f_{\mathrm{nu}}(x)$	
x_{A}	10^7	5000	10^4	100	20
x_{B}	10^7	50	10^4	1	20
x_{C}	10^7	50	10^4	2	40
x_{D}	10^7	14	10^4	0	0

- **② ゼロ頻度に対して有効な推定値を算出できない**
  - $f_{\mathrm{nu}}(x_{\mathrm{D}}) = 0$ のとき、他の頻度に依らず $r_{\mathrm{MLE}}(x_{\mathrm{D}}) = 0$

    ⇒ データにない要素に対しても、有効な推定値を与えたい

先行研究：正則化を導入した手法 [菊地:19]

- 確率分布の推定を経由せず、最適化の枠組みで尤度比を直接推定
  - 正則化パラメータ λ が推定値を低めに見積もる
  - λ がゼロのときは $r_{\mathrm{MLE}}(x)$ と等価

$$\hat{r}(x) = \frac{\hat{p}_{\mathrm{nu}}(x)}{\hat{p}_{\mathrm{de}}(x) + \lambda} \qquad \hat{p}_*(x) = \frac{f_*(x)}{\sum_x f_*(x)}, \quad *\in \{\mathrm{de}, \mathrm{nu}\}$$

x	頻度				$r_{\mathrm{MLE}}(x)$	$\hat{r}(x)$
	$\sum_x f_{\mathrm{de}}(x)$	$f_{\mathrm{de}}(x)$	$\sum_x f_{\mathrm{nu}}(x)$	$f_{\mathrm{nu}}(x)$		
x_{A}	10^7	5000	10^4	100	20	19.6
x_{B}	10^7	50	10^4	1	20	6.7
x_{C}	10^7	50	10^4	2	40	13.3
x_{D}	10^7	14	10^4	0	0	0

※ $\hat{r}(x)$ の正則化パラメータ $\lambda = 10^{-5}$

先行研究 [菊地:19] の利点（1/2）

- **正則化パラメータ λ が低頻度に起因する問題を緩和**
  - 頻度の差を推定値に反映できる
  - 低頻度に対する推定値をロバストにできる

$$\hat{r}(x) = \frac{\hat{p}_{\mathrm{nu}}(x)}{\hat{p}_{\mathrm{de}}(x) + \lambda} \qquad \hat{p}_*(x) = \frac{f_*(x)}{\sum_x f_*(x)}, \quad *\in \{\mathrm{de}, \mathrm{nu}\}$$

x	頻度				$r_{\mathrm{MLE}}(x)$	$\hat{r}(x)$
	$\sum_x f_{\mathrm{de}}(x)$	$f_{\mathrm{de}}(x)$	$\sum_x f_{\mathrm{nu}}(x)$	$f_{\mathrm{nu}}(x)$		
x_{A}	10^7	5000	10^4	100	20	19.6
x_{B}	10^7	50	10^4	1	20	6.7
x_{C}	10^7	50	10^4	2	40	13.3
x_{D}	10^7	14	10^4	0	0	0

※ $\hat{r}(x)$ の正則化パラメータ $\lambda = 10^{-5}$

先行研究 [菊地:19] の利点（2/2）

- **正則化パラメータ λ が低頻度に起因する問題を緩和**
  - 頻度の差を推定値に反映できる
  - 低頻度に対する推定値をロバストにできる

$$\hat{r}(x) = \frac{\hat{p}_{\mathrm{nu}}(x)}{\hat{p}_{\mathrm{de}}(x) + \lambda} \qquad \hat{p}_*(x) = \frac{f_*(x)}{\sum_x f_*(x)}, \quad *\in \{\mathrm{de}, \mathrm{nu}\}$$

x	頻度				$r_{\mathrm{MLE}}(x)$	$\hat{r}(x)$
	$\sum_x f_{\mathrm{de}}(x)$	$f_{\mathrm{de}}(x)$	$\sum_x f_{\mathrm{nu}}(x)$	$f_{\mathrm{nu}}(x)$		
x_{A}	10^7	5000	10^4	100	20	19.6
x_{B}	10^7	50	10^4	1	20	6.7
x_{C}	10^7	50	10^4	2	40	13.3
x_{D}	10^7	14	10^4	0	0	0

※ $\hat{r}(x)$ の正則化パラメータ $\lambda = 10^{-5}$

先行研究 [菊地:19] の欠点

- $r_{MLE}(x)$ **と同様、ゼロ頻度の問題に対処できない**
  - $f_{\mathrm{nu}}(x_{\mathrm{D}}) = 0$ のとき、$\hat{p}_{\mathrm{nu}}(x) = 0$ なので下式もゼロ

$$\hat{r}(x) = \frac{\hat{p}_{\mathrm{nu}}(x)}{\hat{p}_{\mathrm{de}}(x) + \lambda} \qquad \hat{p}_*(x) = \frac{f_*(x)}{\sum_x f_*(x)}, \quad *\in \{\mathrm{de}, \mathrm{nu}\}$$

x	頻度				$r_{\mathrm{MLE}}(x)$	$\hat{r}(x)$
	$\sum_x f_{\mathrm{de}}(x)$	$f_{\mathrm{de}}(x)$	$\sum_x f_{\mathrm{nu}}(x)$	$f_{\mathrm{nu}}(x)$		
x_{A}	10^7	5000	10^4	100	20	19.6
x_{B}	10^7	50	10^4	1	20	6.7
x_{C}	10^7	50	10^4	2	40	13.3
x_{D}	10^7	14	10^4	0	0	0

※ $\hat{r}(x)$ の正則化パラメータ $\lambda = 10^{-5}$

本研究のモチベーション

- 離散要素 x を自然言語処理で多用される **N-gram** とする
  - N個の文字や単語が連続した系列
  - A_k $(1 \leq k \leq N)$ を特徴変数とし、 A_k が取る値（単語や文字）を a_k とする

例）単語4-gram　$x = \langle \mathrm{This}, \mathrm{is}, \mathrm{a}, \mathrm{sentence} \rangle$

$a_1 = \text{"This"}$　$a_2 = \text{"is"}$　$a_3 = \text{"a"}$　$a_4 = \text{"sentence"}$

- x はN個の特徴を持つインスタンスとも考えられる
  - a_k の頻度が疎な点で一般のデータセットとは異なる

- 研究目的：**低・ゼロ頻度を同時に対処する尤度比推定の実現**
  - 低頻度が多い状況では、出現しない要素も多いことが想定される
  - 未観測の要素にも何らかの対処を要する　（例：未知語の処理）

ゼロ頻度の問題への簡単な対策

- a_k $(1 \leq k \leq N)$ の出現に独立を仮定し、 a_k に対する尤度比の積をとる

$$r_R(x) = \prod_{k=1}^{N} \hat{r}(a_k) = \hat{r}(a_1) \times \hat{r}(a_2) \cdots \times \hat{r}(a_N)$$

$$\hat{r}(a_k) = \frac{\hat{p}_{\mathrm{nu}}(a_k)}{\hat{p}_{\mathrm{de}}(a_k) + \lambda}$$

正則化パラメータの導入によって
低頻度にも安定した推定値を付与

  - 要素 x がなくても、a_k が出現していれば計算できる

- **しかし a_k の統計的独立は、実際はほとんど成立しない**
  - 推定性能を低下させる原因
  - **特徴重み付け法を応用する**（本研究のポイント！）

提案手法

- **a_k の独立仮定を軽減したい**
  - **ポイント１　A_k に対する重み W_k を導入**
  - ポイント２　a_k の頻度を補正（本発表では省略）

$$r_{\mathrm{ours}}(x) = \prod_{k=1}^{N} \tilde{r}(a_k)^{W_k}$$

$\tilde{r}(a_k)$：頻度を補正した正則化付きの推定量

- **特徴重み付け法 [Jiang:18]（既存手法）を応用**
  - 重み W_K を計算するアルゴリズム
  - 推定への影響小な A_k には小さい重み、影響大な A_k には大きい重み

本研究で用いる重み付け法 CFW

- **Correlation-based Feature Weighting (CFW) filter [Jiang:18]**
  - ナイーブベイズ分類器のための手法でシンプル・高速・効果的

前提：分類に強く寄与する特徴 A_k は

① クラス $c \in C$ と強い相関を持つ（maximum mutual relevance）

② 他の特徴 $A_{k'}$ と相関を持たない（minimum mutual redundancy）

$$D_k = \underbrace{NI(A_k; C)}_{① \, relevance} - \overbrace{\frac{1}{N-1} \sum_{k'=1 \wedge k' \neq k}^{N} NI(A_k; A_{k'})}^{② \, average \, redundancy}$$

$NI(\cdot)$：正規化した相互情報量

- D_k をシグモイド変換し、$(0, 1)$ の範囲に収めたものを W_k とする

提案手法へ CFW を応用

- CFW は特徴とクラスが定義されていれば使用可能
  - 尤度比は二値分類でよく用いられる

$p_{\mathrm{de}}(x)$ から得た要素 x はクラス $c_{\mathrm{de}} \in C$ に属する
$p_{\mathrm{nu}}(x)$ から得た要素 x はクラス $c_{\mathrm{nu}} \in C$ に属する
と考えて CFW を適用

$$r_{\mathrm{ours}}(x) = \prod_{k=1}^{N} \tilde{r}(a_k)^{W_k} \qquad W_k \in (0, 1)$$

$\tilde{r}(a_k)$：頻度を補正した正則化付きの推定量

- 重みの効果
  - W_k が 0 に近ければ、$\tilde{r}(a_k)^{W_k}$ は 1 に近づいて $r_{\mathrm{ours}}(x)$ への影響小
  - W_k が 1 に近ければ、$\tilde{r}(a_k)^{W_k}$ は $\tilde{r}(a_k)$ に近づいて $r_{\mathrm{ours}}(x)$ への影響大

実験の概要

- 目的：低・ゼロ頻度の両方に対する提案手法の有効性を検証
  - 正則化パラメータ λ と重み W_k を組み合わせたときの有効性
- 実験内容
  - 英文コーパスで固有表現（地名・人名）の左にある単語 N-gram を予測
  - N-gram が固有表現の左文脈 or Not を当てる

地名：直前に前置詞が現れやすいが、前置詞は他文脈にも現れやすい
人名：直前に敬称・人を表す単語（Mr., President）などが現れやすい

- なぜこの実験か？
  - <u>地名と人名は扱う文脈の性質が異なり</u>、提案手法のふるまい分析に便利
  - 正解が一意に定まるため、性能を定量的に評価できる
  - N-gram の次数 N が大きいと低・ゼロ頻度を多く扱う

実験データと実験条件

- データは “ウォール・ストリート・ジャーナルコーパス” から作成
  - Stanford Named Entity Recognizer でコーパスに固有表現タグを付ける
  - コーパスから記事をランダム抽出し、訓練・開発・評価データに分配
    （開発・評価データは1000記事に固定）

- 実験条件：3条件から一つずつ選び、その組み合わせで実験
  - N-gram の次数 N　：　2、4、6、8、10
  - 訓練データのサイズ：　2500、5000、7500、10000記事
  - 固有表現　　：　地名、人名

N=10 のとき、
開発・評価データに含まれる99%以上の N-gram が訓練データに現れない

尤度比推定までの流れ

1．頻度の学習

- 訓練データを N-gram $\langle a_1, a_2, \cdots, a_N \rangle$ に分けて a_k の頻度を数え上げる
  - 訓練データ全体
  - 条件で指定した固有表現（地名 or 人名）の左
- 訓練データから重み W_k を計算

2．手法のパラメータ決定（後述）

3．尤度比の推定

- 評価データにある N-gram x について、学習した頻度で尤度比を推定

$$r(x) = \frac{p_{\mathrm{ne}}(x)}{p_{\mathrm{tr}}(x)}$$

$p_{\mathrm{ne}}(x)$：固有表現の左側での出現確率

$p_{\mathrm{tr}}(x)$：訓練データ全体での出現確率

手法の評価手順（1/3）

(1) 推定値の降順に N-gram を並び替えて順位づけ

(2) N-gram の正誤判定*

(3) 上位8000件についてランクー再現率曲線を描く

（評価の例）

ランク	N-gram	推定値	正誤
1	x_1	35.8	
2	x_2	18.3	
3	x_3	14.9	
4	x_4	13.2	
5	x_5	9.5	
6	x_6	3.6	
7	x_7	0.3	

手法の評価手順（2/3）

(1) 推定値の降順に N-gram を並び替えて順位づけ

(2) N-gram の正誤判定*

(3) 上位8000件についてランクー再現率曲線を描く

* N-gramが評価データで固有表現の左に現れれば正解

ランク	N-gram	推定値	正誤
1	x_1	35.8	正
2	x_2	18.3	正
3	x_3	14.9	正
4	x_4	13.2	誤
5	x_5	9.5	誤
6	x_6	3.6	正
7	x_7	0.3	誤

手法の評価手順（3/3）

(1) 推定値の降順に N-gram を並び替えて順位づけ

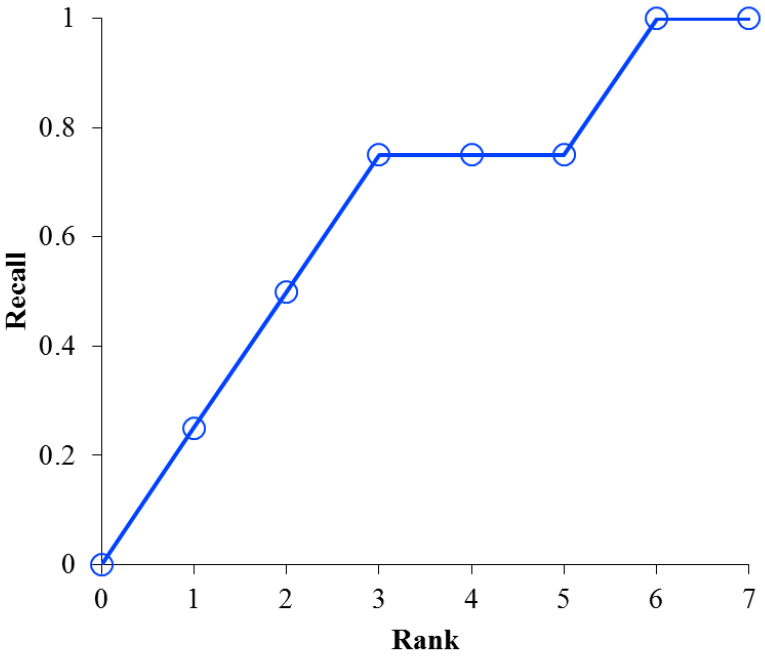
(2) N-gram の正誤判定*

(3) 上位8000件についてランク－再現率曲線を描く（下の例では7件）

	ランク						
	1	2	3	4	5	6	7
正誤	正	正	正	誤	誤	正	誤

$$\text{再現率} = \frac{\text{あるランクまでの正解数}}{\text{評価データに含まれる正解数}}$$

$$\text{適合率} = \frac{\text{そのランクまでの正解数}}{\text{あるランク}}$$



※ グラフ上のある点と原点を結んだ直線の傾きが、その点での適合率に比例

比較手法

- 正則化＋重み付け（提案手法）

$$r_{\mathrm{ours}}(x) = \prod_{k=1}^{N} \tilde{r}(a_k)^{W_k}$$

$W_k \in (0, 1)$

$\tilde{r}(a_k)$：頻度を補正した正則化付きの推定量

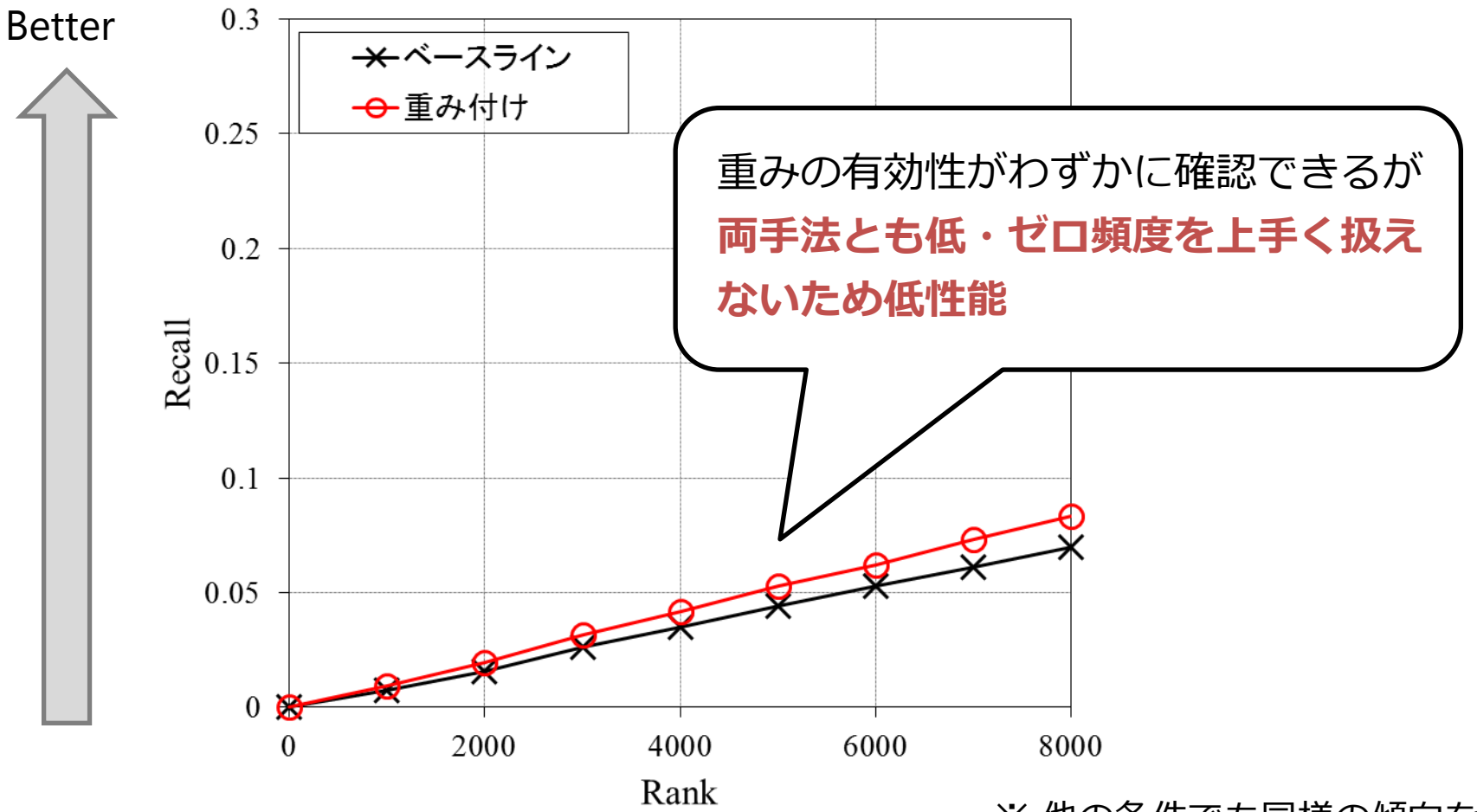
- 比較手法*
  - ベースライン： 確率分布を相対頻度で求めて比を取る（$\lambda = 0, W_k = 1$）
  - 重み付け　　： 重み付けのみを導入（$\lambda = 0$）
  - 正則化　　： 正則化のみを導入　（$W_k = 1$）

- 正則化パラメータ λ は開発データを用いて決定
  - 訓練データが7500、10000記事のとき $\lambda = 10^{-5}$、それ以外は $\lambda = 10^{-4}$

* 頻度は提案手法と同様、補正したものを使用する

ベースライン・重み付けのみの手法は低性能

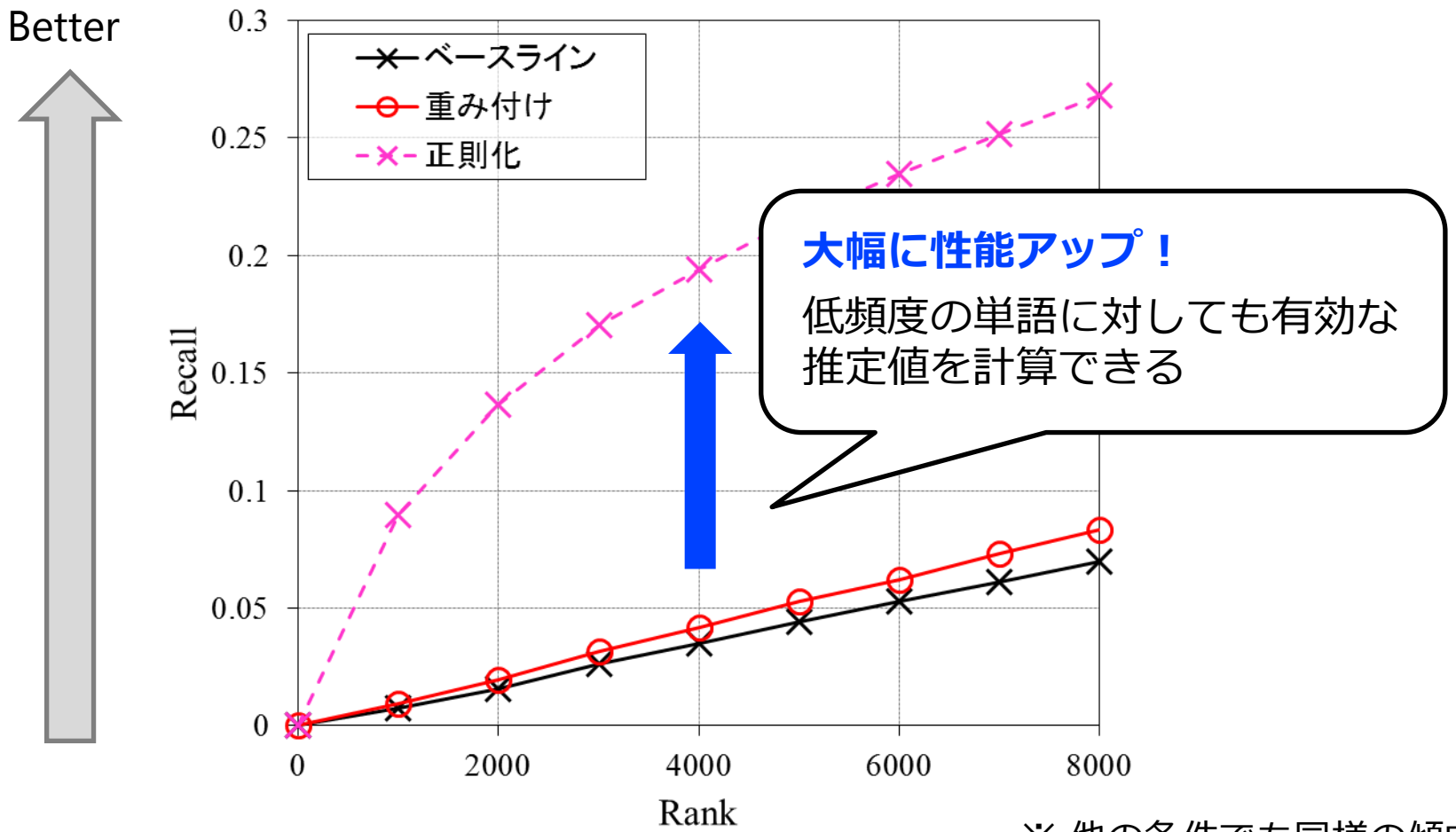
- ランクー再現率曲線（10-gram、訓練データ10000記事、地名）



※ 他の条件でも同様の傾向を確認

単語 a_k 間の出現に独立を仮定しても正則化は有効に働く

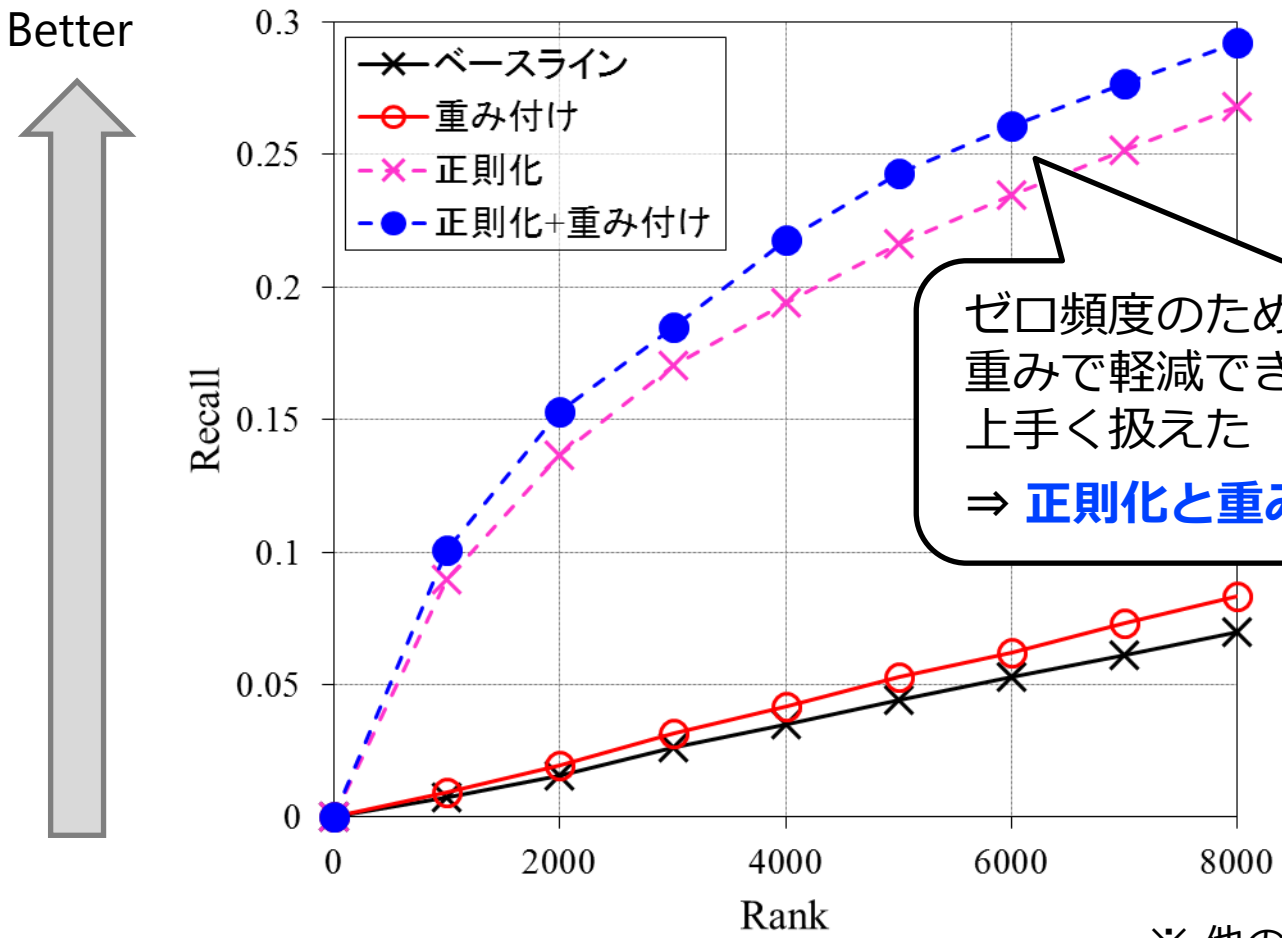
- ランクー再現率曲線（10-gram、訓練データ10000記事、地名）



※ 他の条件でも同様の傾向を確認

正則化＋重み付けが最良性能

- ランクー再現率曲線（10-gram、訓練データ10000記事、地名）



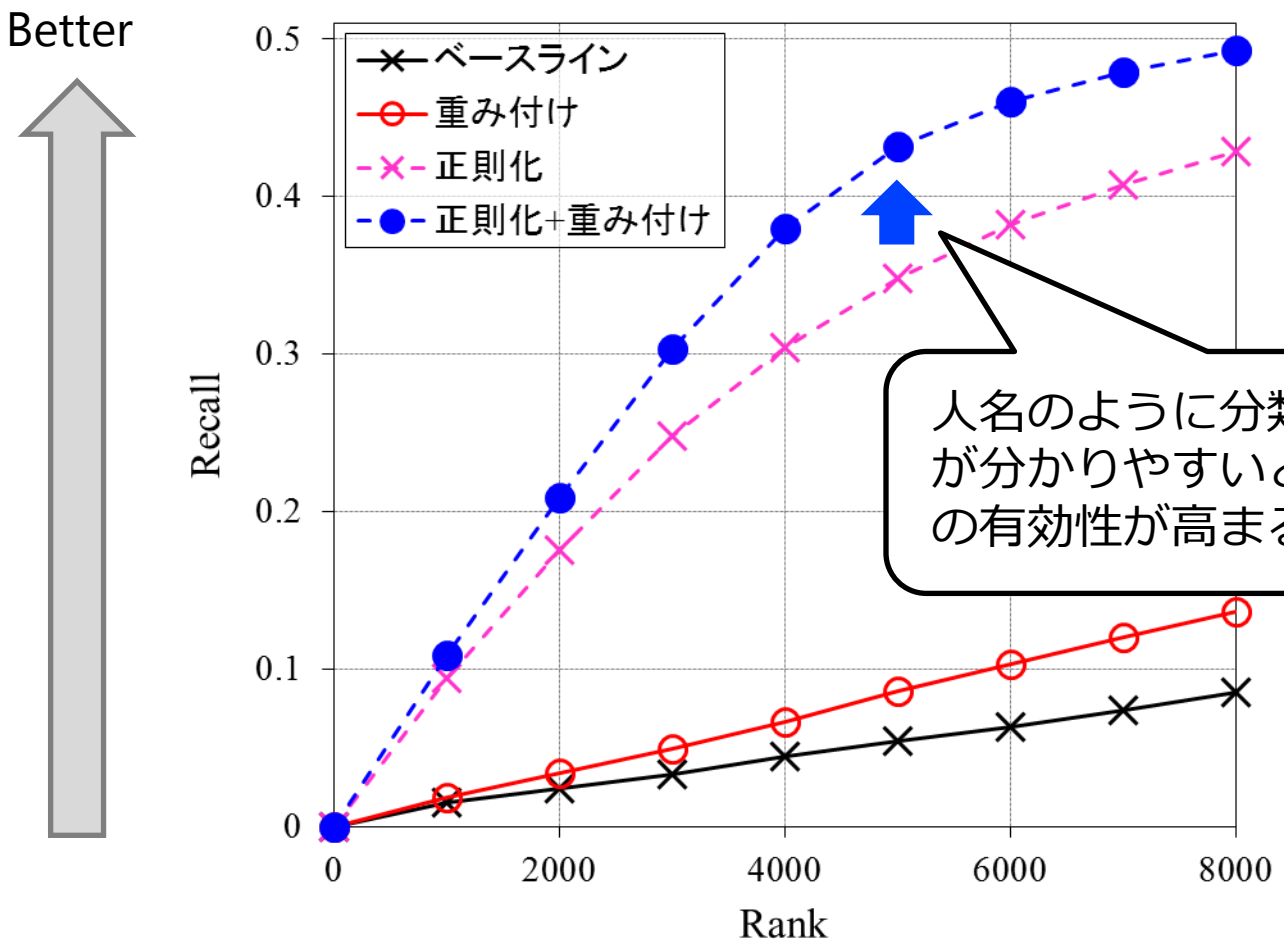
ゼロ頻度のために仮定した独立性を重みで軽減でき、低頻度も正則化で上手く扱えた

⇒ **正則化と重み付けは互いに効果的**

※ 他の条件でも同様の傾向を確認

人名では提案手法の有効性が顕著

- ランクー再現率曲線（10-gram、訓練データ10000記事、人名）

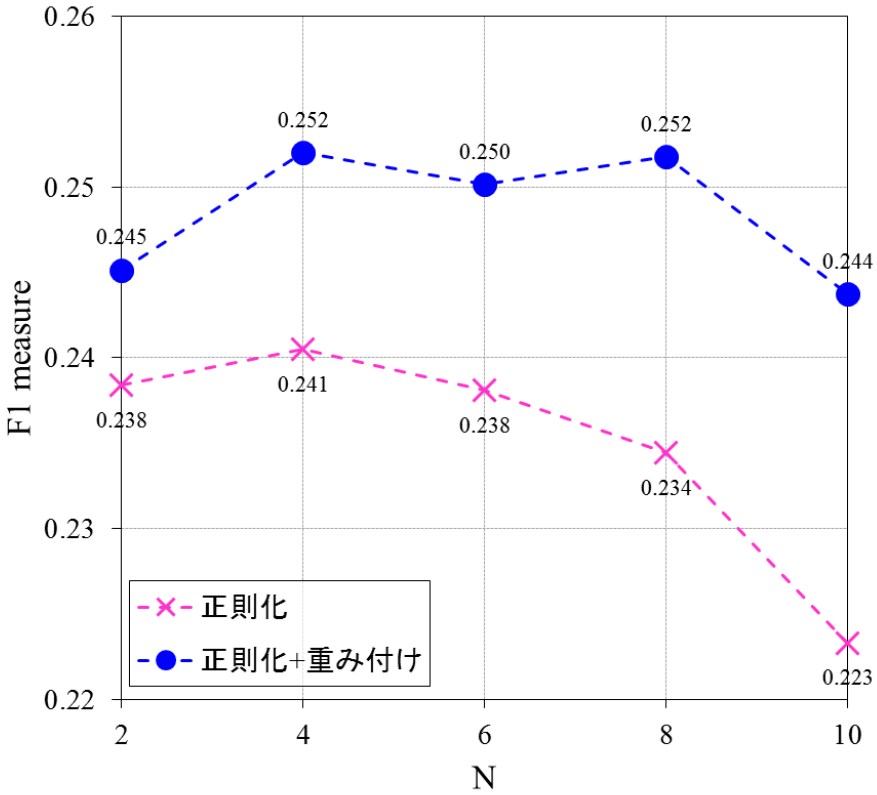


頻度が疎な状況でも正解の約50%を発見

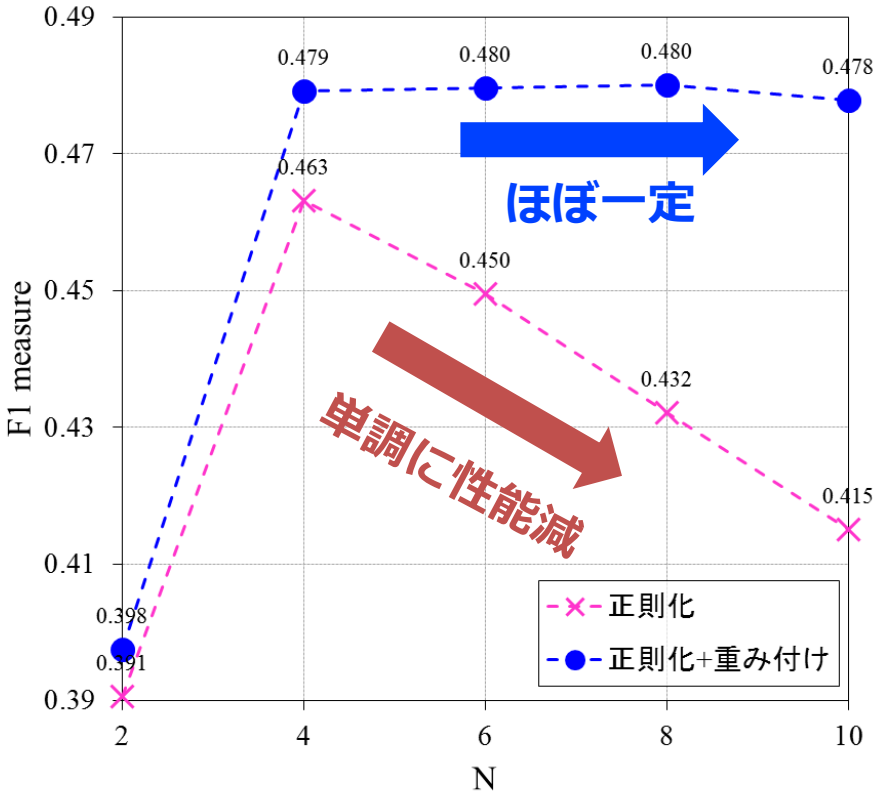
人名のように分類に役立つ特徴が分かりやすいとき、重み付けの有効性が高まる

重みはNの変化に対して提案手法の性能をロバストにする

- Nを変化させた場合のF1-尺度（適合率と再現率の調和平均）
  - 訓練データは10000記事に固定



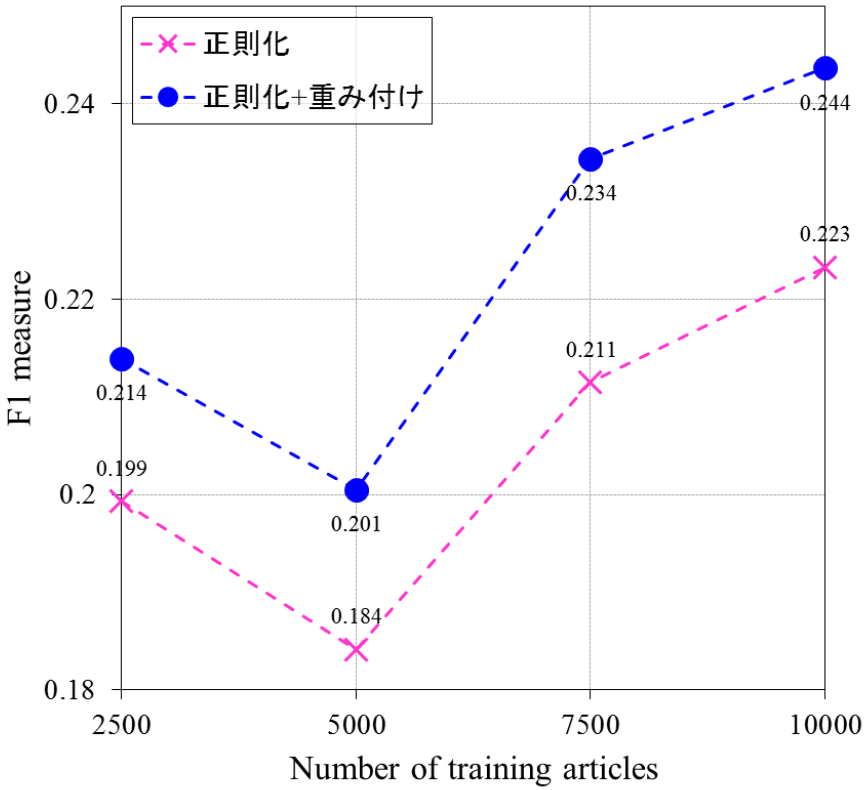
（地名）



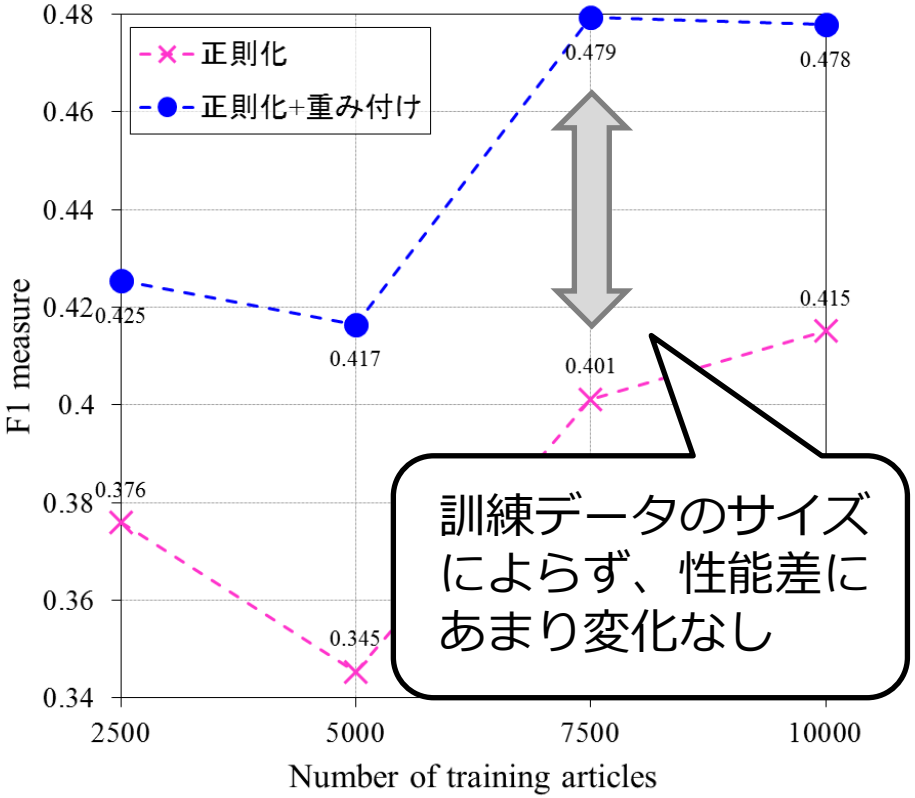
（人名）

Nが十分に大きいとき、訓練データが小さくても重み付けは有効

- 訓練データのサイズを変化させた場合のF1-尺度
  - Nは10に固定



（地名）



（人名）

まとめ

- 研究目的：
  - 低・ゼロ頻度の N-gram に対する尤度比推定の実現
- 提案手法：
  - a_k の出現に独立を仮定し、 a_k の尤度比の積をとる
  - 独立性仮定を緩和するために**特徴重み付け法**を応用

$$r_{\mathrm{ours}}(x) = \prod_{k=1}^{N} \tilde{r}(a_k)^{W_k}$$

- 評価実験：
  - 単語N-gram が固有表現の文脈 or Not を当てる問題
    - 正則化＋重み付けが、疎性の高い文脈予測に最も有効
    - Nが大きくても性能低下を軽減でき、訓練データが小さくても効果的

今後の課題

- 課題１：特徴値 a_k 単位の重み付けによる性能向上
  - 今回、使用したのは特徴変数 A_k への重み付け
  - より細かい重み付けで、微細な推定値のコントロールを可能にしたい
- 課題２：N の値にロバストな推定法の実現を目指す
  - N が大きい（長い）ものに対しても、安心して利用できる手法の開発
- 課題３：提案手法が有効に働く実用的タスクを探す
  - 尤度比は“二値分類の順位づけ関数”として理論的に最適 [中西:15]
  - 提案手法は疎性の高い特徴を持つデータセットで特に有効

中西健太郎, 田中利幸, 上田修功. 尤度比に基づく順位づけ関数による受信者操作特性曲線下面積の漸近的性質. 電子情報通信学会技術研究報告, 2015. IBISML2014-92.

参考文献・連絡先

- スライドで引用した重要文献
  - 菊地真人, 川上賢十, 吉田光男, 梅村恭司. 観測頻度に基づくゆう度比の保守的な直接推定. 電子情報通信学会論文誌D, Vol. J102-D, No. 4, pp. 289-301, 2019.
  - Liangxiao Jiang, Lungan Zhang, Chaoqun Li, and Jia Wu. A correlation-based feature weighting filter for naive Bayes. IEEE Transactions on Knowledge and Data Engineering, Vol. 31, No. 2, pp. 201-213, 2018.

- 研究に関するご質問など
  - 菊地 真人

    Email: m143313 (at) edu.tut.ac.jp

    Homepage: http://www.ss.cs.tut.ac.jp/~kikuchi14/index_ja.html

就活中です！！

重み W_k の具体的な値

- 重み W_k
  - 値が大きいほど、特徴 A_k が尤度比の推定に重要
  - Nが大きいときに人名直前の特徴 A_N が尤度比推定に強く寄与

CFW で計算された重み W_k.

固有表現	N	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9	W_{10}
地名	2	0.453	0.547	—	—	—	—	—	—	—	—
	6	0.432	0.422	0.438	0.468	0.545	0.687	—	—	—	—
	10	0.449	0.441	0.436	0.437	0.442	0.450	0.474	0.511	0.597	0.743
人名	2	0.414	0.586	—	—	—	—	—	—	—	—
	6	0.398	0.386	0.386	0.413	0.544	0.820	—	—	—	—
	10	0.424	0.414	0.410	0.410	0.410	0.416	0.422	0.459	0.617	0.890

※ 訓練データは10000記事

a_k の頻度を補正する必要性

- a_k の独立仮定を軽減したい
  - ポイント１　A_k に対する重み W_k を導入
  - **ポイント２　a_k の頻度を補正**

$$r_{\mathrm{ours}}(x) = \prod_{k=1}^{N} \tilde{r}(a_k)^{W_k}$$

$\tilde{r}(a_k)$：頻度を補正した正則化付きの推定量

- **問題：$f_{\mathrm{nu}}(a_k) = 0$ となる a_k があれば、$r(x)$ の推定値がゼロになる**
  - 【解決策】a_k の頻度に微小な正の値を加算

補正法 - 微小な正の値を加算

- $f_{\mathrm{nu}}(a_2) = 0$ のとき、
  - 補正なし：

$$\hat{r}(a_k) = \frac{\hat{p}_{\mathrm{nu}}(a_k)}{\hat{p}_{\mathrm{de}}(a_k) + \lambda} \qquad \hat{p}_*(a_k) = \frac{f_*(a_k)}{\sum_x f_*(a_k)}$$

$$\prod_{k=1}^{N} \hat{r}(a_k)^{W_k} = \hat{r}(a_1)^{W_1} \times \underbrace{\hat{r}(a_2)^{W_2}}_{0} \cdots \times \hat{r}(a_N)^{W_N} = 0$$

  - 補正あり：

$$\tilde{r}(a_k) = \frac{\tilde{p}_{\mathrm{nu}}(a_k)}{\tilde{p}_{\mathrm{de}}(a_k) + \lambda} \qquad \tilde{p}_*(a_k) = \frac{f_*(a_k) + 1}{\sum_x f_*(a_k) + 2}$$

$$\prod_{k=1}^{N} \tilde{r}(a_k)^{W_k} = \tilde{r}(a_1)^{W_1} \times \underbrace{\tilde{r}(a_2)^{W_2}}_{\text{Not 0 !}} \cdots \times \tilde{r}(a_N)^{W_N} \neq 0$$

比較手法

- 正則化＋重み付け（提案手法）

$$r_{\mathrm{ours}}(x) = \prod_{k=1}^{N} \tilde{r}(a_k)^{W_k}$$

$W_k \in (0, 1)$

$\tilde{r}(a_k)$：頻度を補正した正則化付きの推定量

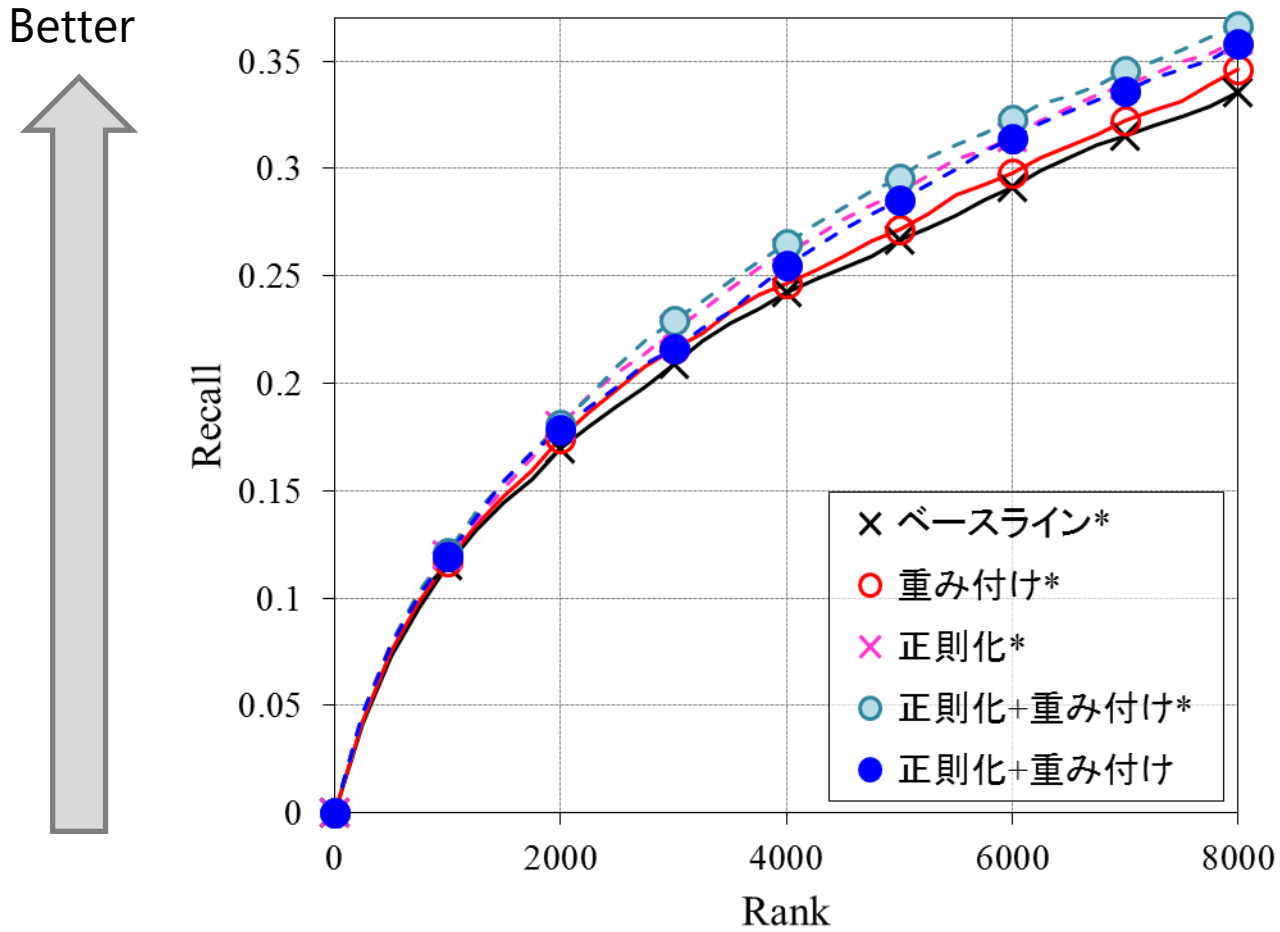
- 比較手法
  - ベースライン*
  - 重み付け*
  - 正則化*
  - 正則化＋重み付け*

補正しない頻度をそのまま使用

- 正則化パラメータ λ は開発データを用いて決定

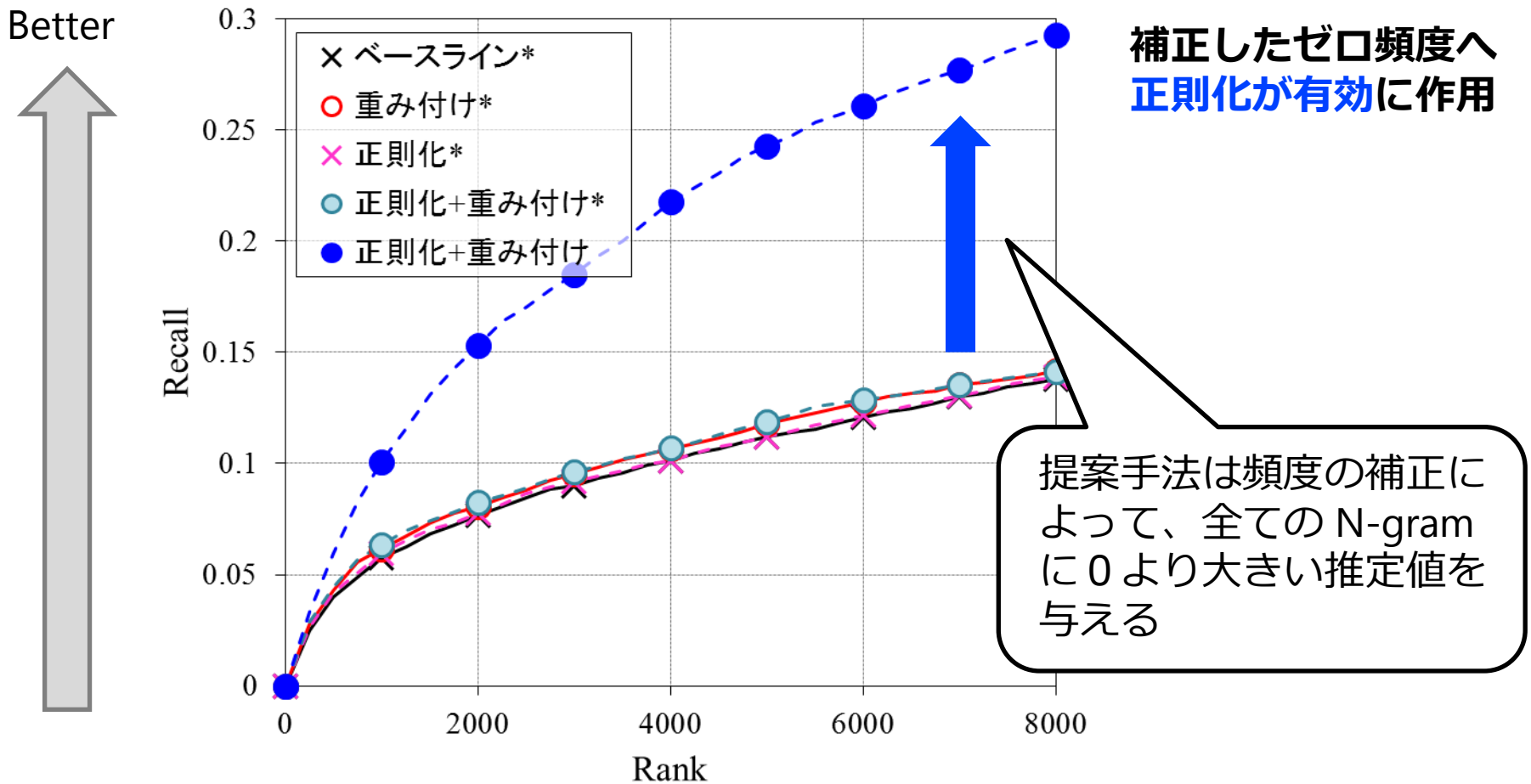
Nが小さいとき、地名ではあまり差が出ない

- ランクー再現率曲線（**2-gram**、訓練データ10000記事、地名）



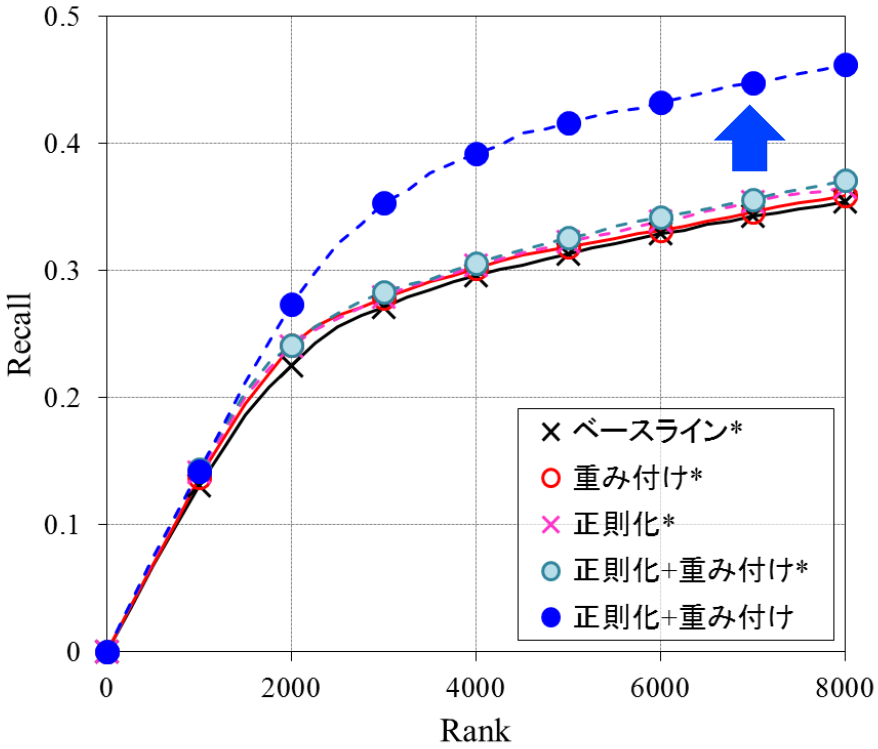
Nが大きくなると補正が効果的

- ランクー再現率曲線（**10-gram**、訓練データ10000記事、地名）

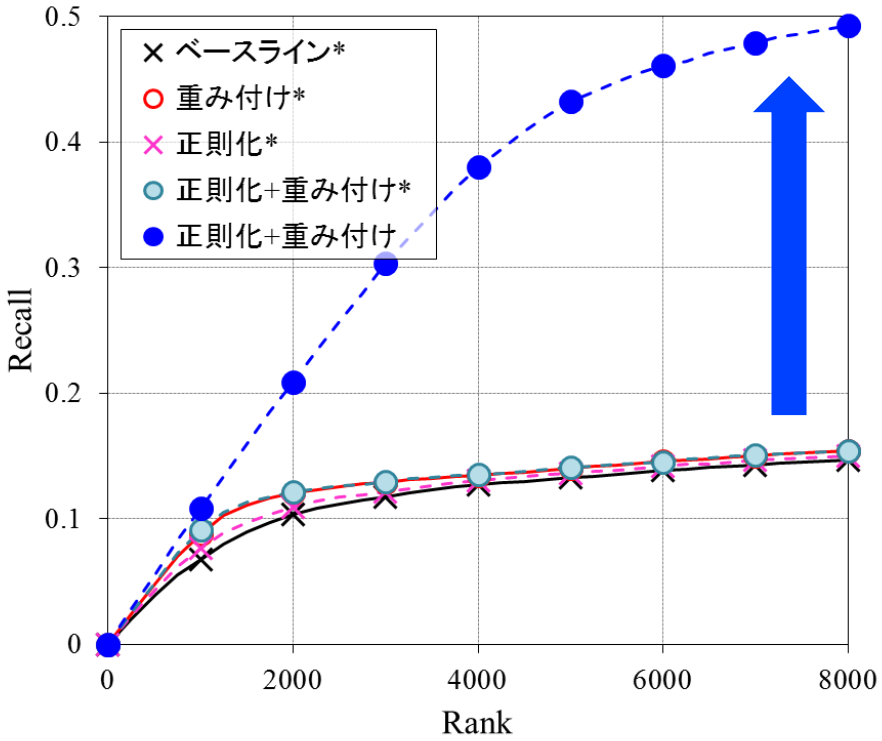


人名のケースではNに依らず有効

- ランクー再現率曲線（訓練データ10000記事、人名）



（N=2）



（N=10）