2Q-9

共起距離を考慮した有害情報フィルタリングの検討

武田　健志†　　中村　健二‡　　小柳　滋†
†立命館大学情報理工学部　　‡大阪経済大学情報社会学部

1　はじめに

インターネットには，青少年の健全な育成に不適切な有害情報が存在している．これらの情報を機械的に判定する様々な有害情報フィルタリングの研究が行われている．その中でも，単語間の共起に基づき抽出した特徴を用いて有害情報を判定する手法が注目されている．これらの研究において，単語の共起は，抽出対象文章に含まれる単語の組み合わせを用いている．しかし，抽出対象によっては，距離が離れている単語の組み合わせは，誤判定を招く可能性があり，精度低下に繋がると考えられる．そこで本研究では，単語間の距離に応じて，共起関係の有無を判定し，有害度を算出する手法を提案する．

2　単語間の共起情報に基づくフィルタリング

語の共起情報に基づくフィルタリング手法 [1] では，学習部と判定処理部がある．学習部では，教師データを解析し，有害辞書の構築をする．判定処理部では，判定データを有害辞書を用いて，有害・無害判定をする．なお，本文における文書とは，Web 上に存在するページを指し，トークンとは文書を構成する最小単位であり，本稿においては単語をトークンとして用いる．以下に判定手順を示す．手順 1〜3 は事前処理部，手順 4 は判定処理部とする．

1. 教師データ文書からトークン w_i を取得する．
2. トークン w_i の有害確率を求め，トークンの有害度辞書を構築する．
3. トークン w_i と他のトークンの共起有害確率を求め，共起有害度辞書を構築する．
4. 辞書の情報に基づき，判定文章の有害確率を求める．

2.1　学習部

学習部では，ベイジアンフィルタの Robinson-Fisher 方式 [3] で，トークン w_i の有害確率を求める．次にトークン w_i とトークン w_j が有害文書で共起する確率 $p(w_i . w_j)$ を式 (1) で求め，その値を用いてトークンの共起有害確率 $f(w_i, w_j)$ を式 (2) で求める．$cogood_{ij}$ は w_i と w_j が非有害サイトで共起した回数であり，$cobad_{ij}$ は，有害サイトで共起した回数，n_{w_i} はトークン w_i が出現した回数である．

$$p(w_i, w_j) = \frac{\frac{cobad_{ij}}{b_i}}{\frac{cogood_{ij}}{g_i} + \frac{cobad_{ij}}{b_i}} \tag{1}$$

$$f(w_i, w_j) = \frac{s \cdot x + n_{w_i} \cdot p(w_i, w_j)}{s + n_{w_i}} \tag{2}$$

2.2　判定処理部

教師データの文章から抽出したトークンごとに有害文書に出現する確率 $p(w_i)$ を求める．

判定データの文章 D から抽出したトークン w_i との共起有害確率 $f(w_i, w_j)$ と 0.5 との差の絶対値が大きい順にトークン w_j を n 個抽出し，式 (3)，式 (4) でトークン w_i の有害性 $S(w_i)$ と非有害性 $H(w_i)$ を求める．次に式 (5) で，トークンの有害確率 $F(w_i)$ を計算し，この値をベイジアンフィルタの Robinson-Fisher 方式 [3] で用いるトークンの有害確率とする．$C(x, df)$ は自由度 df の x^2 分布における x の片側確率である．

$$S(w_i) = C(-2ln\{f(w_i)\prod_j f(w_i, w_j)\}, 2n) \tag{3}$$

$$H(w_i) = C(-2ln\{(1-f(w_i))\prod_j (1-f(w_i, w_j))\}, 2n) \tag{4}$$

$$F(w_i) = \frac{1 - H(w_i + S(w_i))}{2} \tag{5}$$

3　提案手法

図 1 に処理の流れを示す．提案手法には，学習部と判定処理部があり，学習部は既存研究 [1] と同様である．判定処理の判定手順を以下に示す．

1. 判定文書からトークン w_i を取得する．
2. トークン w_i とトークン w_j が共起有害辞書に存在するかを確認する．
3. トークン w_i とトークン w_j の共起距離を算出する．

Study of Filtering Harmful Information Considering Distance Co-occurrence
†Kenji TAKEDA　‡Kenji NAKAMURA　†Shigeru OYANAGI
†College of Informaition Science and Engennering，Ritsumeikan University
‡Faculty of Information Technology and Social Science，Osaka University of Economics

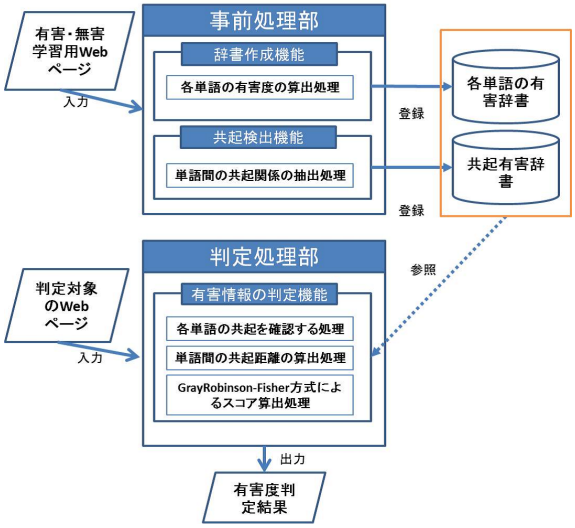

図 1: 処理の流れ

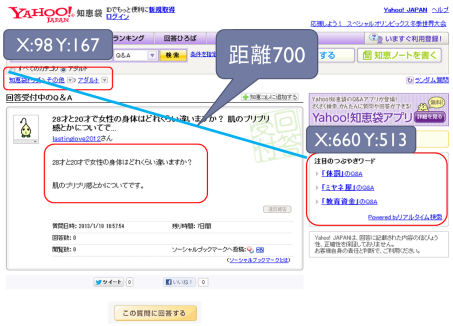

図 2: 単語間の距離のイメージ

4. 共起距離が閾値以下の単語の組み合わせを抽出し，トークンの有害確率 $F(w_i)$ を算出する．
5. Robinson-Fisher 方式 [3] によるスコア算出で，文章の有害度を算出する．

Web ページ上には，様々なトークンが存在し，その全てが共起関係という考えは有用ではない．例えば，有害サイト，無害サイトに頻出する「ログイン」と Web ページ内のヘッダー部に頻出する「ガイドライン」との共起は，誤判定に繋がる可能性があると考えられる．そのため本研究では，判定処理部において，各単語が共起有害辞書に存在するかを確認し，共起有害辞書に存在した場合，単語間の距離を求める．単語間の距離とは Web ページの見た目における距離であり図 2 のようなイメージである．事前に閾値を設定し，単語間の距離が閾値を下回っていた場合，共起有害度辞書の値を単語の有害度とし，既存研究と同様，Robinson-Fisher 方式によりスコアを算出し，判定処理を行う．

4 評価実験

4.1 実験内容

本研究では，判定時における共起距離と判定精度の関係を明らかにするため，単語のみを利用する方式，共起を利用する方式，判定時に単語間の共起距離を考慮する方式の 3 方式に比較実験を行う．実験データは事前に人手で収集した教師有害データ 400 件，教師無害データ 400 件，有害判定データ 1,000 件，無害判定データ 1,000 件を使用する．

4.2 実験結果と考察

共起を考慮しない場合は，2000 件中 84 件の誤判定が見られた．共起を考慮することで 2000 件中誤判定が 2 件まで減り，精度が向上した．判定時における共起距離を考慮することでは，判定が変わるようなデータは見られなかった．しかし共起距離を考慮しない場合と，考慮する場合で，文章の有害度の平均をとったところ，共起距離の閾値が 5000 の時，無害判定データでは変化が見られなかったが，有害判定データの平均が 0. 938 から 0. 967 となり，より有害と判定していることがわかった．

5 おわりに

本研究では，単語間の距離に応じて，共起関係の有無を判定し，有害度を算出する手法を提案した．有害判定データの有害度の平均が向上していることから，本提案手法の有用性を確認した．本研究では判定処理にのみ，共起距離の手法を適用したが，学習部においても共起距離を考慮することで，精度が向上すると考えられる．今後は学習部にも適用し，その効果を検証する予定である．

参考文献

[1] 菊池琢弥， 内海彰: 語の共起情報に基づく有害サイトフィルタリング手法，第 9 回情報科学技術フォーラム講演論文集，情報処理学会・電子情報通信学会，Vol. FIT2010，No2，pp. 1-6(2010).

[2] 中村健二， 田中成典， 山本雄平， 安彦智史: 共起関係の抽出範囲を考慮した有害情報フィルタリング手法， 情報処理学会論文誌，情報処理学会，Vol. 54. No2(2013 発行予定).

[3] G. Robinson:Spam Detection，(2002) <http://radio-weblogs.com/0101454/stories/2002/09/16/spamDetection.html>.