

知識検索サイトにおける不適切な投稿の分類

小林 大祐¹⁾ 松村 真宏²⁾ 木戸 冬子³⁾ 石塚 満¹⁾

1) 東京大学 大学院情報理工学系研究科

2) 大阪大学 大学院経済学研究科

3) ヤフー株式会社

1. はじめに

近年、Web の急速な普及によって膨大な量の知識が Web 上に蓄積されるようになってきた。このような知識を利用したサイトの一つとして、知識検索サイトがあり、ユーザは質問や回答を投稿することができる。しかし、全てのユーザが想定された適切な行動をとるとは限らないため、不適切な投稿の排除は欠かせない。

そこで、本研究では現在人手で行っている不適切な投稿の削除を機械学習による不適切投稿の発見を半自動化するために、教師つき負例と未知データからなる半教師つき学習コーパスからの SVM (Support Vector Machine) 学習器の作成に取り組んだ。

2 章では Web 上の不適切な行動の自動推定に関する関連研究を、3 章では知識検索サイトについて述べる。4 章で本研究の手法について述べ、5 章で実験とその考察を行い、6 章でまとめる。

2. 関連研究

Web 上の不適切な投稿を自動的に分類する研究は、昔からさまざまな分野で行われてきた。有名なものとしては e-mail のスパムフィルタが挙げられる[1]。また、有害な Web サイトをブロックするための研究も行われてきた[2]。また、blog のエントリを対象として、spam かどうかを判定する研究にも取り組まれている [3]。しかし、本研究で対象としている不特定多数のユーザが文書を書き込むサイトに対してはあまり研究が行われていないのが現状である。

3. 知識検索サイト

知識検索サイトとは、あるユーザが投稿した質問に他のユーザが回答を寄せるサイトであり、お互いに知恵や知識を教え合うことを目的としている。過去に投稿された質問や回答は記録されており、後から検索することによって直接その質問と回答には関係ないユーザも情報を得ることもできる。このような

サイトとして、Yahoo! 知恵袋[4]や教えて! goo[5]などがある。

本研究では、Yahoo! 知恵袋の一月分の投稿データと、削除された投稿データの協力をいただいた。人手により削除された投稿データは、Yahoo! 知恵袋のスタッフがチェックして不適切であると判定されたものである。今回は削除された投稿の約 85% を占めていた質問文書を分析対象とする。一月分の投稿データにおける適切な質問文書 (215, 288 投稿) と削除された質問文書 (33, 852 投稿) をコーパスとした。なお、1 つの質問の長さは 60~70 文字であった。

4. 手法

4.1 実験用コーパスの作成

本研究で用いるコーパスについて考察する。削除された質問文書は不適切であることがわかるため、これを負例として用いる。正例は、削除されていない質問文書を用いる。

本研究では、不適切であることが判明している文書と、適切・不適切が混ざった未知の文書があったときに、コーパスを精錬することによって如何にして精度の高いコーパスを得るか、という問題に取り組む。そのため、コーパスに含まれる正例と負例を混ぜて、教師つき負例と未知データからなる半教師つき学習コーパスを作成する。

4.2 コーパスの精錬

教師つき負例データセットを D_0 、未知データセット 2 組をそれぞれ P_0^A と P_0^B とする。

次に、 D_0 と P_0^A から文書を 1 : 1 にランダムに抽出して SVM 学習器を作成し、 P_0^B を分類する。分類して不適切な文書だと判定されたものを $\bar{P}_1^B$ とし、取り除く。残りは P_1^B とする。また、 $D_1 = D_0$ 、 $P_1^A = P_0^A$ とする。

次に、 D_1 と P_1^B より SVM 学習器を作成し、 P_1^A を分類する。不適切と判定された文の集合 $\bar{P}_2^A$ を取り除き、そうでない文の集合 P_2^A はそのままにする。これを交互に繰り返していき、どの程度コー

Automatic Estimation of Inappropriate Posts on Knowledge Exchange Website

Daisuke Kobayashi 1), Naohiro Matsumura 2), Fuyuko Kido 3) and Mitsuru Ishizuka 1)

1) Graduate School of Information Science and Technology, The University of Tokyo

2) Graduate School of Economics, Osaka University

3) Yahoo Japan Corporation

パスがせいせいされる程度を調べる。取り除く文は、SVM の出力がある一定値を越える、スコアの大きい M 件を取り除く、などの方法によって決めることができるが、今回はスコアの大きい M 件を除くこととした。また、取り除かれた文の集合を $\bar{P} = \bar{P}_1^B + \bar{P}_2^A + \bar{P}_3^B + \bar{P}_4^A + \dots$ と定義する。

5. 実験

図 1 は、 D_0 を不適切な文 5000、 P_0^A と P_0^B をそれぞれ適切文 8000、不適切文 2000 で構成させ、 $M=100$ としてコーパスの精錬を繰り返したときの Precision と Recall、そして F 値の変化である。ここで、 $\bar{P}$ のうち実際に不適切な文の割合を Precision、不適切な文章のうち $\bar{P}$ へ取り除かれた割合を Recall とする。精度は 5 回実験した結果の平均値である。図 1 より、Recall はコーパスの精錬を繰り返す度に上がっていくが、Precision は下がっていることが分かる。F 値はおよそ操作を 40 回繰り返したところで最大となっている。

図 2 は P_0^A と P_0^B をそれぞれ適切文 5000、不適切文 5000 で構成させ、 $M=200$ としてコーパスの精製操作を繰り返したときの結果である。未知データにおける不適切な事例の割合が変化すると、コーパスの精錬にも影響があることがわかる。

また、図 2 の実験のときに、適切・不適切それぞれ 500 件ずつの教師つきテストデータを用い、コーパスの精錬が分類精度にどの程度影響を与えるか調べた。del-Precision は SVM 分類器が出力したもののうち実際に不適切な文であった割合、del-Recall は実際に不適切な文のうち SVM 分類器で不適切だと分類できた割合を表す。このとき、全くコーパスを精錬しなかったときは del-Recall が 0.8 を下回っていたが、最も精度がよくなった時には 0.84 となった。

6. まとめ

今回の実験により、教師つき負例と未知データからなる半教師つきコーパスを精錬し、分類精度を高めるコーパスを作成することに成功した。今後はよりよい分類精度を求めて、コーパス精錬手順を改良していく予定である。

謝辞

本研究では、ヤフー株式会社様より Yahoo! 知恵袋のデータの提供をいただき、そのデータを分析に利用させていただき、また、実験についてのご協力もいただきました。ご協力いただきましたヤフー株式会社様には記してお礼を申し上げます。

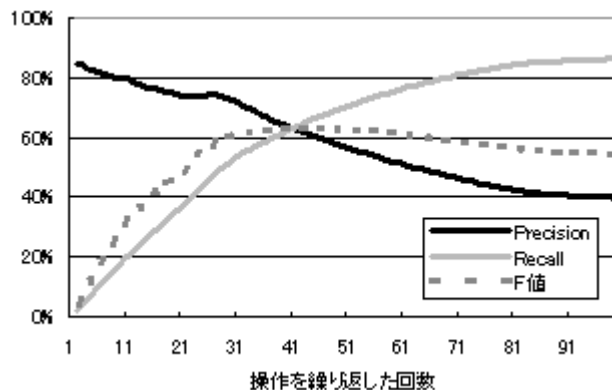


図 1 適切文 8000、不適切文 2000 の時コーパスからの不適切文の排除の様子

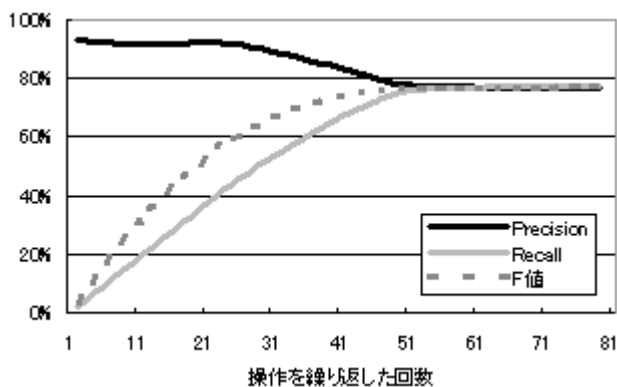


図 2 適切文 5000、不適切文 5000 の時コーパスからの不適切文の排除の様子

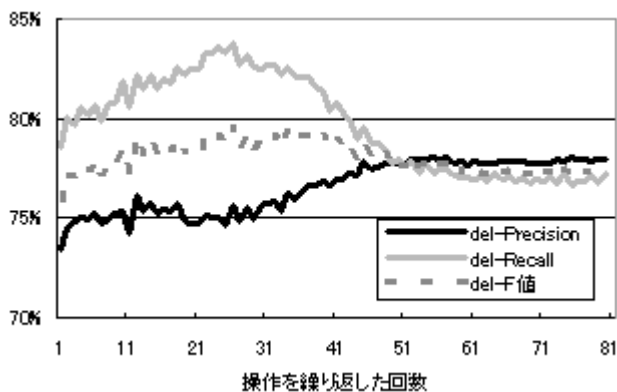


図 3 コーパスの精製による分類精度の変化

参考文献

- [1] H. Drucker, C. Wu and V. Vapnik, Support Vector Machines for Spam Categorization. IEEE Trans. On Neural Networks, vol. 10, number 5, pp.1048-1054, 1999.
- [2] B. Grailheres, S. Brunessaux, P. Leray, Combining Classifiers for harmful document filtering, RIAO' 2004, Coupling Approaches, Coupling Media and Coupling Languages for Information Retrieval, 2004.
- [3] P. Kolari, A. Java and T. Finin, Characterizing the Splogosphere. Proceedings of the 3rd Annual Workshop on Weblogging Ecosystem: Aggregation, Analysis and Dynamics, 15th World Wide Web Conference, 2006.
- [4] Yahoo! 知恵袋 <http://chiebukuro.yahoo.co.jp/>
- [5] 教えて! goo <http://oshiete.goo.ne.jp/>