教師のためのレポート評価支援
～『コピペ』レポートの検出～

峯脇　さやか*

Report Evaluation Support System
— Detecting Copy and Paste from Students' Reports —

Sayaka Minewaki *

Abstract

For teachers, reports evaluations are hard work. And nearly all of students' reports are copy and paste. Students are copying and pasting texts, from web pages and others reports. Author is developing report evaluation support system. In this paper, a method for detecting copy and paste from students' reports is described. Using a text similarity method, copy and paste texts are detected. The method for detecting copy and paste texts and the result of experiment are shown.

1　はじめに

大学生や高専生のレポートはコピペが多い。マウスを使うだけで，Web 上のテキストや他人の文書から簡単にコピー&ペーストすることができ，自分で努力してレポートを作成するよりも，コピペするほうがはるかに楽な作業であるからである。学生は,他人が作成した文章を盗用することについて，意識が低く，また，教育の現場では，盗用が著作権侵害であることを指導しているところは，あまり見かけられない。さらに，安易にコピペすることにより,学生は，考える力や文章作成能力が向上しないという問題が生じる。

教員にとって，レポートチェックの評価や指導は，1 つ 1 つ手作業で行うため，非常に手間のかかる作業である。コピペされたレポートをチェックするのは，時間と手間の無駄である。

本研究では，教師がレポートを評価する際の作業を支援するため，学生レポートのコピペ部分を検出する。本研究の立場は, 教師を支援する立場であり，本研究によって，レポートの評価を行うという立場ではない。これは，評価は，教師がすべきものであって，支援ソフトなどの使い方は，その教師次第であるという考えに基づく。

近年行われているレポートのコピペを判定する研究では，文書間類似度を用いて，レポートのオリジナリティを評価するものが多い。文書間類似度は，文書分類などで用いられるもので，同じテーマのレポートを文書間類似度で計算しても，コピペかどうか判断するのは難しい。そこで，本研究では，文書を対象とするのではなく，レポート内の文を対象として，1 文ごとに類似度を計算する。テキスト間類似度計算法として，BLEU[5]を用いる。

コピペレポートには，(1)他学生のレポートからコピペしたもの, (2)Web 上のテキストからのコピペの 2 つの傾向がみられる。2 つの場合について，それぞれ，テキスト間類似度計算を行う。(1)の場合，2 つの文書中のテキストの全ての組み合わせについて，BLEU による類似度計算を行う。(2)の場合，まず，対象テキストの先頭 4 文節を含むテキストを Web 上から検索する。先頭 4 文節と一致した Web テキストについて，類似度計算を行い，類似度が最も高かったものを，コピペの参照先とする。

以下，2 では，関連研究について述べる。3 では，BLEU について述べる。4 では，コピペレポートの検出方法について述べる。5 では，具体例を示し，最後に 6 でまとめる。

2　関連研究

近年，レポートのコピペを判定する研究が盛んに行われており，文書間類似度を用いたものが多い。文献[1]では，文書の類似度をもとにコピーレポートを判定する「模倣レポート判定支援システム」を開発している。文献[2]では，客観的なレポート間の関係を，コサイン尺度による類似度で求めることで，レポートの独自性を評価している。さらに，評価値を視覚化することで，レポート評価者を支援するシ

*情報工学科科
平成 22 年 9 月 30 日受理

ステムを開発している。文献[3]では，1 対比較法と TF･IDF 法でコピーレポート判定している。さらに，理解度チェック単語数で考察の評価を行っている。これらの値からニューラルネットワークを用いて，レポート評価するシステムを開発している。

さらに，金沢工業大学の杉光氏が考案し，株式会社アンクが開発したコピペ判定支援ソフト『コピペルナー』[4]が 2009 年 12 月より発売されている。

3　BLEU[5]

BLEU は機械翻訳の自動評価基準である。機械翻訳文（candidate）と人手による正解訳（reference）の類似度を計算し，類似度が高ければ，機械翻訳文の精度が高いと評価している。

BLEU は次式によって定義される。

$$\log \mathrm{BLEU} = \mathrm{BP} + \sum_{n=1}^{N} w_n \log p_n$$

ここで, BP は n-gram 数の短さによるペナルティである。w_n は重みであり，$w_n = 1/N$ である。p_n は，n-gram 適合率で，次式によって定義される。

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{ngram \in C} Count_{clip}(ngram)}{\sum_{C \in \{Candidates\}} \sum_{ngram \in C} Count(ngram)}$$

ここで，$Count(ngram)$ は，機械翻訳文の n-gram 数であり，$Count_{clip}(ngram)$ は，機械翻訳文と人手による正解訳において，共通する n-gram 数である。

BP は，次式によって定義される。

$$\mathrm{BP} = \begin{cases} 0 & c > r \\ 1 - \dfrac{r}{c} & c \le r \end{cases}$$

ここで，c は，機械翻訳文（candidate）の n-gram 数である。また，r は，人手による正解訳（reference）の n-gram 数である。

以上より，BLEU は，一致する n-gram の割合（適合率）と，n-gram 数の短さによるペナルティによって計算される。なお，文献[5]では，$N = 4$ としている。

4　コピペレポートの検出

コピペレポートには，(1)他学生のレポートからコピペしたもの，(2)Web 上のテキストからのコピペの 2 つの傾向がみられる。2 つの場合に分けて，コピペレポートの検出を行う。

4.1　前処理

本研究の対象となるレポートは，テキスト形式および，Word で作成された文書である。コピペ検出の前処理として，次の作業を行う。

- Word で作成された文書からテキストデータを抽出する。
- 全てのレポート抽出されたテキストデータについて，1 文ごとに分割する。
- 各文にインデックスを付与する。
- 各レポートをインデックスで表現する。

4.2　他学生のレポートからのコピペ検出

評価するレポートをレポート A，コピペ参照先の候補のレポートをレポート B とする。前処理により，レポート A，B は図 1 のように，文インデックスで表現されている。レポート A の各文について，レポート B と一致するもの，および，類似度が高いものを検出すると図 2 のようになる。一致するものは，「match full」と表示し，類似度が高いものは，参照先の文インデックスと BLEU の計算結果を表示する。

4.3　Web 上のテキストからのコピペ検出

レポート中の各文について Web 検索を行う。一般に，レポートで記述される 1 文の長さについて，20 字以下という文はほとんど見られず，また，文節数も 5 文節以下というものも見かけられない。ここで，レポートで記述されるような長文をダブルクオーテーションで囲んでフレーズ検索しても，一致する Web テキストを発見することは困難である。また，学生によっては，あまりに長い文章を適度に削ってコピペする場合もある。

そこで，文全体を検索するのではなく，先頭 N 文節と一致する Web テキストを検索する。そして，検索対象の文と先頭 N 文節と一致する Web テキストを BLEU で類似度計算を行い，類似度か高いものをコピペの参照先とみなす。なお，本研究では，$N = 4$ としている。また，Web 検索では，検索結果の上位 20 位までの Web ページを対象としている。検索結果を図 3 に示す。図 3 では，文インデックス，BLEU の計算結果，参照先の URL を表示している。

5 おわりに

本稿では, 教師のレポート評価を支援するために, 学生レポートのコピペ部分を検出する方法について述べた。コピペレポートの傾向から，他学生のレポートからのコピペ検出方法と，Web 上のテキストからのコピペ検出方法について述べた。本研究では, テキスト間類似度計算手法として BLEU を用いた。他学生のレポートからのコピペ検出では，レポート中の 1 文ごとに，類似度計算を行う。Web 上のテキストからのコピペ検出方法では，検索対象の文の先頭 4 文節と一致する Web テキストを検索し, 類似度計算を行う。

今後は，コピペ検索結果を教師に見やすく表示するため, 結果を可視化することが課題である。また, 表記のゆれや同義語に対応したコピペ検出への改良も必要と考えられる。

本研究の立場は，教師を支援する立場であり，レポートの評価を行うという立場ではない。本稿では, コピペ部分の検出について述べたが，レポートチェックに要する教師の労力を軽減するため，指導箇所の検出に取り組みたい。

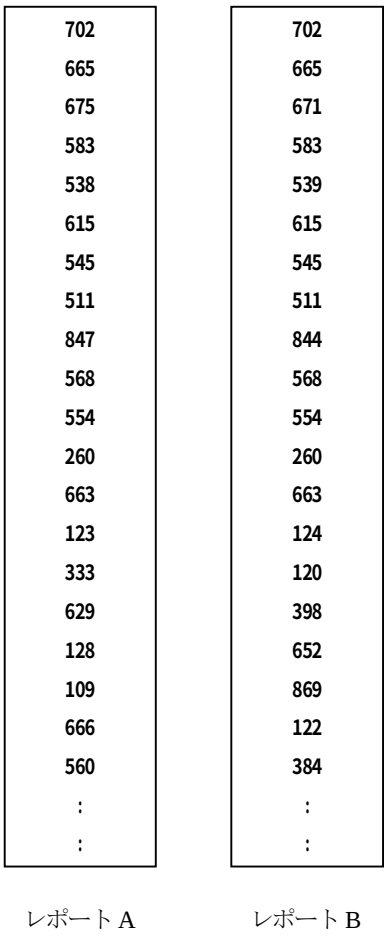

図 1　文インデックスで表現されたレポート

参考文献

[1] 太田，増山：模倣レポート判定支援システムの開発，言語処理学会第 11 回年次大会，pp。293-296，2005

[2] 川口，砂山：内容の独自性を視覚化するレポート評価支援システム, 第 21 回人工知能学会全国大会 2H4-2，2007

[3] 渡邊：ニューラルネットワークを用いた実習レポート評価支援システムの開発，電子情報通信学会技術研究報告 ET, 教育工学 108(146)，pp。7-12，2008

[4] http://www。ank。co。jp/works/products/copypelna/

[5] K。Papineni, et al。BLEU: a Method for Automatic Evaluation of Machine Translation。 *In Proceedings of ACL 2002*, pp。 311-318, 2002

```
702     match full
665     match full
675     671     -0。833051127543801
583     match full
538     539     -0。519860385419959
615     match full
545     match full
511     match full
847     844     -0。835589220409805
568     match full
554     match full
260     match full
663     match full
123
333
629
128
109
666     match full
560
 :
 :
```

図 2　他学生からのコピペ検出結果

```
353     -1
        http://www。naxnet。or。jp/ -rider/cho/cho。htm
330     0
        http://niwango 。 jp/mobile/search/niwango_wiki 。
        php?wid=161&amp;f=hbv99b
395     0
        http://niwango 。 jp/mobile/search/niwango_wiki 。
        php?wid=161&amp;f=hbv99b
372     0
        http://niwango 。 jp/mobile/search/niwango_wiki 。
        php?wid=161&amp;f=hbv99b
390
471     0
        http://www。geocities。jp/chachadess/page014。html
397     0
        http://www 。 euro8 。
        net/wiki/?word=%E8%91%97%E4%BD%9C%E6%A8%A9
472     0
        http://yomi 。
        mobi/wgate/%E8%91%97%E4%BD%9C%E6%A8%A9/
440
409     0
        http://www 。 wiki-movie 。
        net/wiki/E89197E4BD9CE6A8A9E6B395。html
148     0
        http://www 。 wiki-movie 。
        net/wiki/E89197E4BD9CE6A8A9E6B395。html
115     -0。184699610360374
        http://yomi 。
        mobi/wgate/%E8%91%97%E4%BD%9C%E6%A8%A9%E6%B3%95
        /
334     0
        http://www。birdlandgolden。com/copyright。html
315
103     0
        http://wikipedia 。 atpedia 。
        jp/m/wiki/%E8%91%97%E4%BD%9C%E6%A8%A9%E6%B3%95
138     0
        http://www。greenmarketstl。com/017。html
280
82
84
279
21
452
```

図 3　Web からのコピペ検出結果