レポート自動採点支援システムによる Cプログラミング演習の評価

Evaluation of C Programming Exercises by the Automatic Reports Grading Support System

渡 邊 博 之^{※1}

Hiroyuki WATANABE

In this research, the degree of similarity between reports is calculated by the one pair comparison method and the TF·IDF method, the z-scores are calculated by all words and some keywords in a report. The training data for the neural network is made by these values, and the automatic grading support system, which evaluates reports in 5 steps, has been developed. The reports of the C programming exercises submitted by e-mail could predict the score within an error of 10 % by using this system, compared to manual operation. Furthermore, the office hours of 3 TAs could be reduced to 0, and contributed to improving the quality of writing reports for the students.

Keywords : Reports, Neural Network, Automatic Grading Support System, Evaluation in 5 Steps

キーワード：レポート，ニューラルネットワーク，自動採点支援システム，5段階評価

1. はじめに

PCの普及とペーパーレス化に伴って，演習・実験のレポートをファイルやメールなどの電子媒体で提出する機会が増えている．電子レポートは手書きのレポートのように自分で読んで書かなくても，コピーすることが容易である．また，最近の学力低下に伴って，学習内容を理解してない考察不足のレポートも目立っている．教員はこれらの電子レポートを一つ一つ開いて読み，さまざまな項目を考慮して総合的に採点している．レポートに不備がある場合は，その項目をフィードバックによって指摘し，再提出させることもある．学生数が100人以上となると，1人で公平な採点を行なうには作業負荷が増大し，作業時間も授業時間外におよぶ．授業時間内に採点しようとするために，レポートを総単語数のみで採点せざるを得ない場合もある．採点作業を項目ごとに教員が分担すると総合的な採点が困難となる．コピーレポートを防止するため，提出されるレポートの課題や内容を毎回変更すると採点は更に困難となる．これらのレポートの採点支援をシステムによって短時間に処理できれば，教員は余った時間で他の作業が行える．

従来の研究では，MoodleなどのようなLMS（学習支援システム）を用いたレポート管理システムが活用されている¹⁾．教員と学生との双方向コミュニケーションによってレポートの書き方を向上させるには有

効であるが，自動採点でないために多くの労力と作業時間が必要である．Webによってレポートが提出された場合を対象に，TF·IDF法を用いて文章のコピーを検索する方法^{2), 3)}がある．TF（Term Frequency）は1通のレポート内での単語 t の出現回数を示し，IDF（Inverse Document Frequency）は単語 t の出現する総レポート内でのレポート数を示している．これら2つの積がTF·IDF法である．文章中の特徴的な単語を抽出するアルゴリズムであるため，コピーレポートの検索に適用することができる．しかし，レポートの採点では更に点数化しなければならない問題がある．SVM（Support Vector Machine）を用いてレポートをグループに分類⁴⁾し，グループ毎に点数化する方法も考えられる．この方法は数通のレポートのパラメータ（例えば，類似度，総単語数，z得点）を基に総レポートを2分類（例えば，よい考察と考察不足に分類）するような場合に，2分類間の距離が最大となるような分離面を求める方法で，2段階評価にはよい方法である．しかし，3段階以上の評価には適用の限界がある．z得点を用いて3段階評価する方法⁵⁾はレポートの採点に有効であるが，5段階評価に拡張する場合は閾値の設定が問題となる．プログラミング演習に限定された環境においてはいくつかの支援システムが開発されている^{6), 7)}．しかし，コンパイルに関する支援が主体であって，考察などを含むレポートの評価は行われていない．自由記述式レポートの内容を自動採点するシステム⁸⁾はレポートの順序立てや文体などの書式をチェックし，指導するには有効である．

平成20年8月29日受付

※1 日本大学工学部

しかし、実際の運用では評価の重みが固定されている問題がある。また、採点結果の精度については検討されていない。論文の課題と採点項目の重みを設定すると、小論文の回答を自動採点するシステム⁹⁾は、論文構成と評価を行うことができる。しかし、重みの設定を教員が設定しなければならないため、異なる課題が混在する場合には問題であり、コピーなどを考慮した総合的な評価は行われていない。

本研究では、1対比較法とTF・IDF法で類似度を求め、考察に含まれる合計単語数と理解度チェック単語を基に z 得点を求めている。これらを基に教師データを作成し、ニューラルネットワークによって点数を予測すると共に¹⁰⁾、学生にメールで返信する実用的なシステムを開発している¹¹⁾。本システムを用いた5段階評価は実際の手作業による採点結果と比較して、誤差10%以内で点数を予測することができた。ニューラルネットワークによる予測は教員による採点時の重みが考慮でき、年度別推移による学力の評価も可能であることを示した。また、従来は毎日届くメールのレポートに対して、教員1名の採点者とTA3名の補助者を必要としていたが、本システムによって教員1名で済み、評価に要する時間が従来の1/4にすることができた。更に、コピーレポートが減少し、質の高いレポートの書き方に貢献できた。

2. レポート評価のアルゴリズム

図1はレポート自動採点支援システムの処理手順を示している。以下、各処理項目について述べる。

2.1 提出条件のチェック

テキスト形式のメールで提出されたレポートを1通ずつ分離し、件名を基に課題に対応するディレクトリに分類する。ディレクトリの中から1通のレポートが選択される。

図2は課題とレポートの提出書式を示している。学生はプログラムを空欄に穴埋め方式で完成させ、実行結果を含めてプログラム部分とする。次に、「考察」とだけ書かれた1行を区切記号として続け、考察と問題回答を考察部分として書く。これをレポートとしてメールで送信すると、本文の前にヘッダが付加される。自動採点支援システムではこれらを1行ずつ解読し、書式が整っていなければ、再提出の理由を付加した返信メッセージを作成する。

2.2 単語の抽出

英語のように分かち書きされていない日本語は単語に分解する必要がある。このため、元の文章(図3(a))を提出条件用XMLファイルに書かれた排除する語句(図3(b))を基に、単語に分解する(図3(c))。XML(eXtensible Markup Language)は図3(b)に示すように〈タグ〉要素〈/タグ〉で構成される言語で、タグに対して文章の処理を行うことがで

図1 レポート自動採点支援システムの処理手順

【課題】2つの実数を入力し、その平均を求めるプログラムを作れ。

```
From u12345@cc.ce.nihon-u.ac.jp July 12 12:00 2000
#include <stdio.h>
int main() {
    float a, b, c;
    printf("2つの実数を入力してください?");
    scanf("%f, %f", &a, &b);
    c = (a + b) / 2;
    printf("2つの実数の平均は %5.2f です。¥n", c);
    return 0;
}
```

考察

私のプログラムの特徴は平均値を書式付フォーマットで表示することです。データとして、1と2を入力すると、結果は1.50と表示する。

図2 課題とレポートの提出書式

1と2を入力すると、結果は1.50と表示する。

(a) 元の文章

```
<keywords> <key>する</key> <key>は</key>
<key>が</key> <key>の</key> <key>を</key>
<key>と</key> </keywords>
```

(b) 排除する語句

1 2 入力 結果 1.50 表示

(c) 文章分解後の単語

図3 文章分解の例

き、色やレイアウトを定義するスタイルシートによってWebブラウザ上に整形して表示できる。文章から排除する語句を除くと、多くは名詞、形容詞などが単語として抽出される。

2.3 類似度の計算

レポートdで単語tが抽出されると、レポートd内には単語tがいくつか出現する。総レポート数をn通としたとき、レポートdがコピーかオリジナルかを検索するためには、総レポート数からレポートdを取り除いた残りの(n-1)通のレポートと比較して類似度を求める必要がある。以下では(n-1)通のレポートを比較レポートと呼ぶことにする。単語tと比較レポートの単語tを1通ずつ比較して一致する単語数を求めるのが1対比較法であり、単語tのレポートdでの出現回数と、単語tを含むレポートがn通の総レポート内で何通あるかを求めるのがTF・IDF法である。コピーレポートの検索漏れが少なく(再現率が高く)、オリジナルレポートが誤ってコピーレポートと判定されない(適合率が高い)ようにするため、類似度は1対比較法とTF・IDF法の2種類によって求める。また、2種類の類似度をニューラルネットワークの入力データとして用いる。

(1) 1対比較法

レポート(1通)と比較レポート(1通)のそれぞれの単語数を N_i と N_j とし、両方のレポートに共通に含まれる単語数を N_{ij} としたとき、次式によって類似度sim1を求める方法が1対比較法である。

$$\text{sim1}(N_i, N_j) = \frac{N_{ij}}{(N_i + N_j)/2} \quad (1)$$

式(1)からsim1(N_i, N_j)とsim1(N_j, N_i)は同じ値になる。部分コピーを含むコピーレポートの閾値はsim1(N_i, N_j) $\geq 50\%$ に設定する³⁾。

(2) TF・IDF法

レポートd内で出現回数の多い単語tはそのレポートを特徴付けると考えられる尺度がtf(t, d)である。しかし、tf(t, d)の値が大きいだけではレポートを特徴付けることはできない。比較レポートにも多く含まれるような単語は特徴があると言えないためである。そこで、単語tがn通の総レポート内で少数のレポートにしか出現しない尺度としてdf(t)を考える。その逆数がidfであり、nをdf(t)で割って対数をとった値で求められる。従って、tf(t, d)の値が大きくidfの値が大きい単語tはレポートdを特徴付けることになる。レポート(1通)と比較レポート(総レポート数-1)との単語の出現回数の重みをw(t, d)とすると式(2)で示される。

$$w(t, d) = \frac{1}{m} \left\{ \text{tf}(t, d) \cdot (\log_2 \frac{n}{\text{df}(t)} + 1) \right\} \quad (2)$$

但し、mはレポートdにおける総単語数である。2つのレポートで同一の重みが数多くある場合、内容が類似していると考えられる。そこで、単語に対応する重みのベクトルを作成し、式(3)で定義される2つのレポート間の余弦を用いて類似度sim2(d_j, d_k)を求める。

$$\text{sim2}(d_j, d_k) = \frac{\sum_{i=1}^m w(t_i, d_j) w(t_i, d_k)}{\sqrt{\sum_{i=1}^m w(t_i, d_j)^2} \sqrt{\sum_{i=1}^m w(t_i, d_k)^2}} \quad (3)$$

但し、 $0 \leq \text{sim2}(d_j, d_k) \leq 1.0$ である。閾値以上は部分コピーを含むコピーレポートであり、閾値はsim2(d_j, d_k)=1.3×sim1(d_j, d_k)=65%に設定する³⁾。式(3)の値が大きいほど d_j と d_k とは類似している。また、式(3)はsim2(d_j, d_k)とsim2(d_k, d_j)では値が異なる。多くの場合、コピー先のレポートはコピー元に比べて総単語数が少なく、かつコピー元の単語数を含むため、sim2の大きいレポートがコピー先と特定できる。例えば、レポート d_j とレポート d_k の2通において、前者がsim1(d_j, d_k)=60%、sim2(d_j, d_k)=65%で、後者がsim1(d_k, d_j)=60%、sim2(d_k, d_j)=78%の場合、sim1は両者同じ値でかつ50%以上のため、コピーレポートとなる。また、後者のsim2は前者より大きいため、後者がコピー先(前者はコピー元)と特定できる。コピー先をより確実にするには、更にレポートの提出日の遅い(早い)を調べる必要があるが、sim2の値が大きいレポートは提出日の遅いレポートと多くの場合一致する。

2.4 理解度の計算

一般には、単語数が多いほど豊かな表現力で、学習内容を多く含んでいると考えられる。しかし、感想文や、結果を文章で説明したにすぎない考察不足のレポートがあるため、考察の合計単語数だけでは理解度が評価できない。また、考察の合計単語数を直接用いると、出現回数の多い単語の影響が強く現われるため、出現回数と同じでも1種類と多種類では閾値の設定が問題となる。課題に関して授業中に教員が説明した文法、用語、キーワードなど重要なN種類の単語を理解度チェック単語と呼び、あらかじめ提出条件用XMLファイルに登録して、単語の種類を考慮して理解度を求める。考察の合計単語数と理解度チェック単語との相対値では文章が短い場合、少ない単語数で高い値となる問題が生じる。そこで、理解度は比較レポートとの相対値で3段階に分類し、総合的な採点は総単語数(絶対値)を加えてニューラルネットワークを用いて5段階で評価する。この評価法は、一般に数通のレポートを大まかに見て全体の相対的基準を決め、その後各レポートを絶対的基準で総合的に採点する方法に相当する。以下では、理解度チェック単語数から求めたz得点の合計を z_N 、考察の合計単語数から求めたz得

点を z_T として、両者の併用によって理解度を評価する。また、 z_N と z_T をニューラルネットワークの入力データとして用いる。

(1) 理解度チェック単語による z 得点

z 得点(z_N)は、レポートから抽出された単語 t_i の期待値 E_i と標準偏差 σ_i から、式(4)で表される偏りの強さを表す統計量である。

$$z_N = \sum_{i=1}^N z_i = \sum_{i=1}^N \frac{tf(t_i, d) - E_i}{\sigma_i} \quad (4)$$

但し、 N は理解度チェック単語の種類である。式(4)の正負と大きさによって理解度の閾値を設定することができる。

(2) 考察の合計単語数による z 得点

考察の合計単語数による z 得点(z_T)は、式(4)において $N=1$ として表される。 $\sigma=\pm 1.0$ 内には68%のレポートが含まれるので、閾値を $-1.0 < z_T < 1.0$ で「普通」とし、 $z_N > 0$ で $z_T \geq 1.0$ は「よい考察」、 $z_N < 0$ で $z_T \leq -1.0$ は「考察不足」とする返信メッセージを作成する。

2.5 点数の予測

ニューラルネットワークは図4に示すように複数のユニットを階層型に並べ、各階層のユニットは1つ前の層からの入力を受け、次の層へ出力される構成で示される。結合荷重 w_{ij} の値を求めることが目的であるが、これには入力(レポートの類似度や理解度など)とそれに対応する出力(点数)とを教師データとして準備する必要がある。出力層の値(予想値)と教師データ(目標値)との誤差が最小となるように w_{ij} を変化させ、出力層からその前の層に逆向きに順々に求めていくバックプロパゲーション(誤差逆伝播法)によって w_{ij} が求められる¹²⁾。 w_{ij} が求められると、その後は入力層の値によって絶対的基準で点数が出力層から出力される。このとき、予測精度は階層型ネットワークの構成、教師データの数と種類などに関係するため、次の条件で行う。

図4 階層型ネットワークの構成

(1) 階層型ネットワークの構成

一般に階層型ネットワークは入力層の全ユニットを中間層の各ユニットに全て結合し、中間層の各ユニットを出力層の各ユニットに全て結合する構成が多い。この場合、点数分布は一様分布となる問題点がある。

本研究では、図4に示すように点数のマイナス項目(例えば類似度)とプラス項目(例えばレポートの総単語数や z 得点)とに分けて中間層の各ユニットで結合し、出力層で合成している。この場合、点数分布は任意分布となる。

(2) 教師データの数と種類

入力ユニットには採点で考慮する項目を全て入力する必要がある。しかし、入力ユニット数が多いと解の収束が悪くなるため、「1対比較法による類似度(sim1)、TF・IDF法による類似度(sim2)、レポートの総単語数、考察の合計単語数による z 得点(z_T)、理解度チェック単語数による z 得点(z_N)」とする。

予測値は新しい学習パターンでは発火率が低くなるので、学習パターンの種類はなるべく多いほうがよい。教師データの個数を同一にした場合、特定の点数の入力データが多すぎると予測値はその点数に特化するため、汎化能力は低下する。従って、1点から5点までさまざまなパターンの教師データでバックプロパゲーションによる学習を行う。

(3) バックプロパゲーションによる学習

s 番目の教師データを入力したときの第 k 層の i 番目ユニットの出力 v_{si}^k は入力 x_{sj}^{k-1} と結合荷重 w_{ij}^{k-1} との積の総和から閾値 t_i^k を引いた次式で表される。

$$v_{si}^k = \sum_j w_{ij}^{k-1} x_{sj}^{k-1} - t_i^k \quad (5)$$

式(5)で求めた v_{si}^k を入力として出力を求める閾値関数として、一般にステップ関数(v_{si}^k の値が正で1、負で0)や、シグモイド関数(入出力がS字状曲線の関係)が用いられる。本研究では、閾値関数をシグモイド関数とした次式を用いる。

$$y_{si}^k = \frac{1}{1 + \exp(-v_{si}^k)} \quad (6)$$

また、出力を y_{si}^w 、教師データを y_{si}^d 、教師データの数を L とすると、誤差関数 $E(w)$ は次式で表される。

$$E(w) = \frac{1}{2} \sum_{i=1}^L (y_{si}^w - y_{si}^d)^2 \quad (7)$$

ニューラルネットワークの学習は $E(w)$ を小さくする方向に結合荷重を修正するよう行われるので、式(5)の初期値の w_{ij} をランダムに決め、その値を繰り返し更新することによって最適な w_{ij} を求める手法である最急降下法で求める。中間層のユニット i と出力層のユニット j とを結合する結合荷重の修正量 $\Delta w_{ij}^k(t)$ は

次式で決定できる。

$$\Delta w_{ij}^k(t) = -\varepsilon \frac{\partial E(w)}{\partial w_{ij}} - \alpha \Delta w_{ij}^k(t-1) \quad (8)$$

但し、 t は学習の反復回数、 ε は学習係数 ($\varepsilon > 0$) で1回あたりの結合荷重の変化を表し、小さいと学習回数が膨大となる。 α は慣性係数 ($0 < \alpha < 1$) で前回の結合荷重の変化の影響を表している。中間層と出力層との結合荷重が求められると、入力層と中間層との結合荷重の修正量が求められ、ネットワーク全体をバックプロパゲーションによって学習できる。

従って、結合荷重の新しい修正量は、 $(t-1)$ 回の修正量を $w_{ij}(t-1)$ 、今回(t)の修正量を $\Delta w_{ij}(t)$ としたとき、次式で求められる。

$$w_{ij}(t) = w_{ij}(t-1) + \Delta w_{ij}(t) \quad (9)$$

図4の()内にバックプロパゲーションで求めた結合荷重の一例を示している。 w_{12} は類似度が高いと点数は低く、 w_{13} は総単語数が多いと点数は高いため、 w_{12} と w_{13} の結合荷重の符号は反対となっている。

2.6 採点画面

図5は採点画面の一例である。本システムによる処理結果はタグを付けた採点用XMLファイルで生成される。タグに対応させたスタイルシートを適用することによってWebブラウザ上に統一された書式で表示できるため、公平に採点できる。プログラムと考察が表示され、それぞれに対する類似度が表示されている。採点で考慮すべき注意事項は「赤」で表示される。理解度チェック単語を基に計算した理解度に対応するメッセージとして「よい考察です」「考察不足です」などが表示される。教員はWebブラウザ上に表示されたレポートとニューラルネットワークで予測された

図5 採点画面

5段階の点数を見ながら採点を行う。採点結果を次のニューラルネットワークに反映させたい場合は予測値にチェックマークを入れると教師データに追加される。採点結果はCGIによってデータベースに保存される。また、学生にメールで返信される。採点後は、採点用XMLファイルの表示を「未」から「済」に変更することにより、未入力 of 採点を防止している。また、「決定」ボタンを無効にすることで、二重採点を防止している。

3. 結果と考察

テキスト形式のメールで提出された学部2年次生のCプログラミング演習の合計90名(2008年度)を対象にニューラルネットワークで点数を予測した。教師データ数は15名(=各段階3名×5段階)としている¹¹⁾。

3.1 予測精度

図6は点数分布と誤差を示している。「目標値」は実際に手作業による点数を示している。「誤差」は、目標値に対して予測値が異なる場合を1、目標値と同じ場合を0として総和を求め、合計人数で割っている。「予想値」が1点のレポートは、「目標値」を全て含むために「誤差」はない。「目標値」の2点が「予想値」で1点のための誤差である。点数差が1点の原因は、「目標値」が2点のレポートは理解度チェック単語よりも総単語数で評価(逆に、3点や4点のレポートは理解度チェック単語で評価)する傾向が強いことによる。このことは、図7から総単語数で採点すると2点を中心、 z 得点で採点すると3点を中心の分布となることから説明できる。ニューラルネットワークでは教師データから結合荷重を決めるため、教師データ数が少ないと誤差となる。点数差は最大で2点である。「目標値」の3点が「予想値」で5点となった中には考察が重複しているレポートがある。本システムでは文章の重複が見分けられないため、実際には予測値を参考にレポートを採点しなければならないが、その労力は少ない。レポート全体の誤差は10%である。誤差

図6 点数分布と誤差

図7 採点法の違いと点数分布

の原因は考察部分で考察と問題回答との区切記号が無いレポートに見られる。実際に手作業による採点では、3点か4点か迷う場合があり、これらを考慮すると、全体の誤差10%は許容範囲と考えられる。

3.2 採点法の違い

ニューラルネットワークの構造と教師データの入力データを変化させず、出力データ(点数)だけを変化させることによって、採点で考慮した項目が予測値に正しく反映されるかを検討する。

図7の○は図6と同じ「予測値」である。また、□は出力データを「総単語数」のみ考慮した点数で予測した結果である。この場合は全体に点数が低く採点されている。×は出力データを「類似度」のみ考慮した点数による予測値を示している。 $100\% \geq \text{sim1} > 80\%$ を1点、 $80\% \geq \text{sim1} > 60\%$ を2点、 $60\% \geq \text{sim1} > 40\%$ を3点、 $40\% \geq \text{sim1} > 20\%$ を4点、 $20\% \geq \text{sim1} > 0\%$ を5点とした。2点以下において、合計人数の5%が部分コピーを含むコピーレポートであることがわかる。教員が採点時に考慮した項目の重みによって点数分布が異なるが、ニューラルネットワークではそれを反映した結果が得られている。△の「z得点」は閾値を($z_T < -0.7$, $z_N < 0$)で1点、($|z_T| < 0.4$, $|z_N| < 0.6$)で3点、($z_T > 0.7$, $z_N > 0$)で5点、($z_T < 0$, 点数≠1, 3)で2点、($z_T > 0$, 点数≠3, 5)で4点とした教師データによる予測値である。z得点は点数が正規分布の場合には有効である。しかし、点数分布に偏りがある場合にはニューラルネットワークの「予測値」が図6の目標値(○)と良く一致している。ニューラルネットワークを用いてレポートの点数を予測することは採点に有効であると考えられる。

3.3 年度別推移

図8は点数分布の年度別推移を示している。学習内容が同一であれば教師データは再利用できるため、異なる年度でも同一基準で採点できる。2005, 2006, 2007, 2008の各年度の平均点はそれぞれ1.7, 2.4, 2.6, 2.8点である。2005年度は類似度50%以上のレポート

図8 点数分布の年度別推移

が10%程度あり、点数が4点以上は少ない。2008年度は「類似度、総単語数、理解度チェック単語」を学生にフィードバックすることによって、コピーレポートが5%以下と減少している。また、学習内容も理解度チェック単語に気をつけて書くようになっており、質の高いレポートが多く見られる。

4. あとがき

テキスト形式のメールで提出されたCプログラミング演習のレポートの点数をニューラルネットワークによって5段階評価する実用的なシステムを開発した。その結果、次のことが明らかとなった。

- (1) 2005年度は教員1名に対して次の補助作業を行うTA 3名でレポートを採点していた。(a)メールで届いたレポートを紙に印刷し、課題毎に分類する。(b)教員の採点した5段階評価をファイルに書き込む。(c)書式が整っていないレポートは理由を付けてメールで返信する。しかし、2007年度以降は本システムによって点数が予測できるため、教員1名で細部まで労力と時間をかけなくても確認のみで採点できる。このことは手作業による採点では0.5通/分に対して、本システムでは2通/分であり、評価に要する時間が従来の1/4ですむことから確認できる。
- (2) 採点用XMLファイルにスタイルシートを適用することでレポートの書式が統一され、閲覧しやすい形式で公平に採点できる。スタイルシートによって、コピーレポートや採点の未入力は赤字で警告されるので、採点ミスや未入力の採点が防止できる。
- (3) ニューラルネットワークによる予測値は教員の採点時の重みを考慮しているため、実際の手作業による迷い(例えば、3点か4点か)を補助でき、同一レポートを二度提出した場合、同じ採点結果が得られる。その誤差は10%以内である。
- (4) これまで1ページ/人のレポートがあり、10課題100人では合計1,000ページの紙を使用していた。し

かし、本システムはブラウザを用いて採点するため、これらの紙の消費削減が可能である。

- (5) z 得点はレポート数によって平均と標準偏差が異なるので、採点結果もレポート数によって異なる。このため、全てのレポートが集まった後に一度に採点する場合や、平均値がレポート数によって変化しない場合はよい。しかし、毎週ごとに再提出を含むレポートを評価するような場合は、同じレポートでも比較レポート数によって点数が変化する問題がある。これに対して、ニューラルネットワークの場合は、過去の採点を基に総合的に評価できるので、比較レポート数によって採点結果が異なることはない。
- (6) ニューラルネットワークによる点数予測は、教師データに変更がない限り、共通の採点基準で年度の違いによる学力の評価が可能である。
- (7) 採点結果に個別メッセージを付加してメールで返信できるので、コピーレポートが減少し、質の高いレポートの書き方に貢献できる。

本研究ではCプログラミング演習を対象に結果を考察したが、電子工学実験レポートにおいても同様な結果が得られており¹⁰⁾、演習・実験の課題に依存しない活用が可能である。

参 考 文 献

- 1) Moodle : <http://download.moodle.org/>
- 2) 高橋, 宮川, 小高, 白井, 黒岩, 小倉 : Webサイトからの剽窃レポート発見支援システム, 信学論D, 90-D, 11, pp.2989-2999, 2007
- 3) 渡邊博之 : 1 対比較法とTF-IDF法によるコピーレポートの検索に関する一考察, 平19東北連大, 1D11, p.135, 2008
- 4) 祖, 大山, 若林, 木村 : 統計的分類手法による英文新聞記事のテキスト自動分類, 電学論C, 124, 3,

pp.852-860, 2004

- 5) 渡邊博之 : z 得点を用いたレポートの成績評価に関する一考察, 2007年信学ソサイエティ大, A-13-12, p.203, 2008
- 6) 遠西, 大作, 五十嵐, 長谷川, 中村 : eLearningにおけるレポート評価支援システムの試み, 信学技報, ET2005-93, IEICE2006-03, pp.31-34, 2006
- 7) 松浦佐江子 : プログラミングレポート採点支援ツールと課題設計による評価方法の改善, IT活用教育方法研究, 9, 1, pp.36-40, 2006
- 8) 津森, Alieu, 磯本 : 自由記述式レポートの自動採点と教員評価による総合評価, 信学技報, ET2003-16, pp.37-42, 2003
- 9) 情報センター : <http://202.16.123.5/jess/>
- 10) 渡邊博之 : 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, 信学技報, ET2008-18, pp.7-12, 2008
- 11) 渡邊博之 : レポート自動採点システムを用いたプログラミング演習の評価, 平成20年度工学・工業教育研究講演会, 10-109, pp.192-193, 2008
- 12) 中野 馨 : Cでつくる脳の情報システム, 近代科学社, 1997

著 者 紹 介

渡 邊 博 之

1979年 日本大学工学部電気工学科卒業
1981年 日本大学大学院工学研究科博士前期課程修了, 同大学助手
1991年 同大学専任講師
2000年 同大学准教授
専門分野 教育情報工学
学 位 博士 (工学)
所属学会 電子情報通信学会, IEEE, 日本工学教育協会
連絡先 hiroyuki@ee.ce.nihon-u.ac.jp