

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 22 日現在

機関番号：34506

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500134

研究課題名(和文) 非定常的感情を考慮した情報検索基盤の研究

研究課題名(英文) Analysis and Extraction of Non-stationary emotion from Social media

研究代表者

灘本 明代 (Nadamoto, Akiyo)

甲南大学・知能情報学部・教授

研究者番号：30359103

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：非定常的感情分析として、TwitterとSNSを対象とし、感情分析を行った。具体的には、(1)Twitter感情分析用感情語辞書の構築(2)Twitter上で用いられている顔文字の感情及びその役割に着目した、顔文字付きツイートの感情値抽出手法の提案、(3)Twitter用の感情軸の決定、(4)SNSからの耳より情報抽出手法の提案を行った。研究成果として、国際会議査読付き6本、国内査読付き1本、国内会議15本、論文誌3本の論文と書籍1編(分担)を発表した。

研究成果の概要(英文)：We had studied the analysis and extraction of emotion from social media such as Twitter and SNS. We studied as follows: (1)Creating emotion dictionary for Twitter, (2)Role of Emoticons for Multidimensional Sentiment Analysis of Twitter,(3)Emotion axis for Twitter, (4)Extracting tip information from SNS.

Our outputs are published 6 International Conference papers, on refereed domestic conference paper 15 non-refereed domestic conference papers, and one book.

研究分野：データ工学

キーワード：感情分析 感情軸 Twitter SNS 顔文字

1. 研究開始当初の背景

現在の情報検索の研究では、情報検索と情報推薦を融合し、検索を行っているユーザに適した情報を提示する研究が多く行われている。これらはユーザの過去の振る舞いに基づいた人の定常的な好みにあわせた検索手法であり、そのユーザが検索を実行しようとしている瞬間の意向をとらえた検索を行うことができない。人のその時々に変化するものとして「感情」があげられる。我々はこの瞬間瞬間で変わる人の感情に合わせた情報検索を行うことにより、よりユーザがその時に欲している情報を的確に提示できると考える。

2. 研究の目的

ソーシャルメディアである、SNS や Twitter は気楽に思ったこと感じたことを投稿することが出来るため、その時その時のユーザの感情が表現されている場合が多い。そこで我々は、瞬間瞬間で変わる人の非定常的感情は、SNS と Twitter から取得できると考え、これら SNS と Twitter を対象とし、これらの大規模データから感情を抽出する手法を提案する。そして、今後の情報検索に役立つ基盤技術を確立する。

3. 研究の方法

以下の5つの研究より非定常的感情抽出を行う。

(1) Twitter 感情分析用感情語辞書の構築

これまでの感情語辞書は新聞を対象とした辞書であった。しかしながら、話し言葉の多い Twitter や SNS では既存の感情語辞書に不足している単語が多数存在する。そこで、話し言葉に対応した Twitter 感情分析用感情語辞書を構築する。

(2) 顔文字付きツイートの感情抽出手法の提案

Twitter には顔文字が多く存在している。これらは様々な感情を持つと共に、付属している文の感情も変える役割を持っていると考えられる。そこで、ユーザ実験を行い、これら顔文字の感情及び役割を抽出し、顔文字付き文の感情抽出手法を提案する。

(3) Twitter 用の感情軸の決定

上記3つの研究では10軸の感情軸を用いて感情抽出を行ってきた。しかしながら、Twitter では10軸の感情を決定するのは困難である事が判明し、次元削減を行って、Twitter 用の感情軸を決定する事を行う。

(4) SNS からの耳より情報抽出手法の提案

SNS 上にはお得と感じる様々な耳より情報が存在する。そこで、SNS を対象として、ユーザがお得と感じる耳より情報の抽出手法を提案する。

4. 研究成果

(1) Twitter 感情分析用感情語辞書の構築

ツイートから感情を推定する為、まずはじめに単語毎に感情の定量化を行い、感情語辞書の構築を行う必要がある。ここでは感情語辞書構築の手法として、熊本ら[1]が提案している感情語辞書構築システムを用いる。熊本らの手法は、辞書を構築する為に大量のデータに現れる任意の単語と、あらかじめ定義してある感情語群との共起関係を調べ、その結果に基づき辞書構築を行っている。熊本らが感情語辞書構築システムを用いて構築した感情語辞書は、用いる感情軸を新聞記事を読んだ人々がその記事からどのような印象を受けたかをアンケート調査により調べた結果に基づき、「楽しい 悲しい、うれしい 怒り、のどか 緊迫」の3つに決定している。しかしながら、本研究では Twitter を対象としているため熊本らの提案している感情語辞書をそのまま用いるのは困難である。そこで本研究では、ツイートに含まれる様々な感情に対応できるように中村の感情表現辞典[2]で分類されている「喜」、「好」、「安」、「哀」、「厭」、「怖」、「怒」、「恥」、「昂」、「驚」の10軸の感情を用い熊本らの感情語辞書構築システムにより感情語辞書を構築する。感情語辞書構築システムでは対極となる感情軸が必要だが、中村の感情軸には対極となる軸が存在しない。その為、Plutchik の感情の輪[3]で向かい合っている感情は対極関係であることを参考にして、対極の感情軸を決定する。具体的には、中村の感情軸を Plutchik の感情に当てはめて「sadness(哀) joy(喜)」、「fear(怖) anger(怒)」、「remorse(厭) love(好)」に決定した。また、それ以外の「安」、「恥」、「昂」、「驚」の感情軸に関しては Plutchik の感情の輪に対応した感情が無かった為、手動で「恥 安」、「驚 昂」に決定した。さらに、新聞記事と違い Twitter ではユーザが自由にツイートしている為、口語的な表現が存在する。その為、Twitter 用に感情語辞書を再構築する必要がある。しかしながら、ツイートは崩れた表現や文法の乱れが多くあり、単語間の共起関係が正しく現れてない場合が多い。そこで、ツイートと同様に口語的な表現が多く、表記ゆれや文法の乱れが少ない Yahoo!映画のレビューデータ 74,000 文書を用いて熊本らのシステムにより辞書構築を行った。その結果、今回用いる各感情軸に対して約 5,600 語~7,500 語の感情語とその感情値を感情語辞書に登録した。

(2) 顔文字付きツイートの感情抽出手法の提案

Twitter には顔文字と呼ばれる表情を持った顔のように見える記号を使用することで表情を表す表現が多く用いられている。しかしながら、同じ顔文字でも文との関係によって

その役割が異なると言える。そこで本研究では、まず始めに、ユーザ実験により顔文字辞書を構築した。そして、文の感情と顔文字の感情に着目し、その関係を分類することで、顔文字の役割を定義するとともに、顔文字の役割を考慮した感情値抽出手法を提案した。

(2.1) 顔文字辞書の構築

顔文字の役割は顔文字のみで決まっているのではなく、文と顔文字の感情の関係から顔文字の役割が決定する。したがって、感情軸ごとに顔文字を追加する必要がある。しかし、既存の顔文字辞書では本論文の感情軸ごとに顔文字を分類していないため、以下の手順で顔文字辞書の構築を行った。

1. 教師データとする顔文字を集める。
2. 人手により教師データの顔文字の感情を決定する。
3. 教師データの顔文字をパーツに分類し、パーツごとの感情を決定する。
4. 3で抽出したパーツをもつ顔文字 400 個を収集し、顔文字辞書に追加する

まず、教師データ作成のため、Twitter から収集した顔文字付きの 17,647 ツイートの中で出現頻度が高い 100 個の顔文字を教師データに用いた。また、今回用いる 100 個の顔文字で収集した顔文字の 92% の割合を占める。教師データの感情を決定するためにユーザ実験を行った。被験者は顔文字を利用したことがある 20 代の男性 10 人である。実験の方法は、顔文字を見てもらいその顔文字がどの感情軸に当てはまるか最大 3 つの感情から選んでもらい、1 番に選択した感情を 3 点、2 番に選択した感情を 2 点、3 番に選択した感情を 1 点としてアンケート結果を収集した。次に、ユーザ実験で各感情に割り当てられた顔文字のパーツに着目し分析を行った。顔文字は人の表情を表現するという性質上、人の顔のパーツを表していることが多い。そこで、本論文では顔文字のパーツを「目」、「口」、「眉」、「頬」の 4 つとし分析を行った。その結果、パーツの中には各感情と強い共起関係にあるものがあることがわかった。抽出したパーツを参考に顔文字辞書に顔文字を手で追加した。その結果、顔文字辞書に登録した顔文字を 500 個まで増やした。このようにして顔文字辞書を構築した。

(2.2) 顔文字の役割を考慮したツイートの感情抽出手法

ツイートから感情を抽出するために、構築した感情語辞書と顔文字辞書を用いて、顔文字の役割を考慮したツイートの感情値算出手法を提案した。以下に算出手順を示す。

1. Twitter からツイートを取得する。
2. 取得したツイートを形態素解析し、単語を抽出する。
3. 感情語辞書とのマッチングにより各単語の感情値を算出する。
4. 顔文字を含んでいるツイートでは、文の感情と顔文字の感情から顔文字の役割を決定

し、ツイート全体の感情値を算出する。また、顔文字を含んでいない場合は(3)の感情値をツイート全体の感情値とする。

< 文の感情値の決定 >

Twitter から取得したツイートに対して形態素解析エンジン Juman を用いて形態素解析し、ツイート内の形態素と感情語辞書に登録されている感情語のマッチングを行った。しかしながら、表記ゆれが発生する場合は感情語が正しく抽出できない。表記ゆれを解決するために Juman の代表表記を用いた。また、ツイートの否定語が入っている場合は正しい感情値を抽出することが出来ない。そこで、熊本ら [1] の提案する否定語についてのルールを適用することで否定語の判定を行い、否定語を含む場合の感情語を打ち消すことで問題を解決した。

< 顔文字の役割の決定 >

文の感情と顔文字の感情から役割を決定するために実験を行った。被験者は 10 代から 50 代までの男女各 10 人の合計 100 人である。実験データは顔文字を含んだ任意のツイート 100 件とそのツイートから顔文字を除いた文字のみのツイート 100 件の合計 200 ツイートである。被験者はこの 200 件のツイートを見て各々のツイートに対して当てはまると感じた感情を 10 軸の感情の中から最大 3 つの感情を選び、各々 10 段階で評価した。実験において、あるツイートのある感情における被験者のつけた点数の合計値をその感情の感情値とする。感情値がある閾値以上で且つ最大のものをそのツイートの感情とする。このときすべての感情が閾値以下である場合は、そのツイートの感情は無感情として扱う。

実験結果より、文のみのツイートの感情と顔文字を含んだツイートの感情の変化より顔文字の役割を決定した。ここで実験では 10 軸の感情によりツイートの感情を判定したが、顔文字の役割を決定するに当たり、感情の変化が大きいほど顔文字の役割が決定しやすいため、10 軸の感情を「positive」、「negative」、「neutral」の 3 軸に分類し、この 3 軸の感情の変化により役割を決定した。具体的には、「喜」、「好」、「安」、「昂」を「positive」、「哀」、「厭」、「恥」、「怖」、「驚」、「怒」を「negative」、10 軸に判定されなかったツイートを「neutral」とした。文のみで判定した感情（実験より）とすでに求めてある顔文字の感情（辞書より）と顔文字を付加したツイートで判定した感情（実験より）との違いを見て役割を決定した。実験の結果より文の感情と顔文字の感情の組み合わせが、ツイートの感情極性に与える影響について分類した。その結果を表 1 に示し、以下に顔文字の役割を示す。

・強調：顔文字が文の感情をより強めている。positive の文に positive の顔文字が付加された場合と negative の文に negative の顔文字が付加された場合。

- ・弛緩：顔文字が文の感情を和らげている．positive の文に negative の顔文字が付加されたときツイートの感情が弱い positive になる場合と，negative の文に positive の顔文字が付加されたときツイートの感情が弱い negative になる場合．
- ・転換：顔文字の感情が文の感情と入れ替わる positive の文に negative の顔文字が付加されたときツイートの感情が弱い negative になる場合と，negative の文に positive の顔文字が付加されたときツイートの感情が弱い positive になる場合．
- ・付加：文の感情がないときに顔文字の感情が付加される．neutral の文に positive の顔文字が付加された場合と neutral の文に negative の顔文字が付加された場合．

表 1. 顔文字の役割

文字 + 顔文字	ツイートの感情		
	Positive	Negative	Neutral
positive+positive	強調	-	-
positive+negative	弛緩	転換	-
positive+neutral	-	-	-
negative+positive	転換	弛緩	-
negative+negative	-	強調	-
negative+neutral	-	-	-
neutral+positive	付加	-	-
neutral+negative	-	付加	-
neutral+neutral	-	-	-

(3) Twitter 用の感情軸の決定

これまで我々は中村明の感情語辞書[2]に基づき，10 軸の感情を用いて Twitter の感情抽出を行ってきた．それらの感情軸に対して振り分けられている単語の感情値を求めることによってツイートの感情を抽出する方法を提案した．また，既存の 10 軸の感情におけるツイートの感情抽出手法の有用性を示すために実験を行った結果，システムの判断と被験者の判断には大きな違いがあった．その原因の 1 つとして 10 軸の感情では軸の数が多く，Twitter に合った感情軸ではないことが分かった．そこで本研究ではさらに，既存の 10 軸でツイートの感情を抽出した際に，ある軸が他の軸に対してどのような影響を与えているかを「感情の関係」と定義し，この感情の関係から Twitter に適した感情軸を決定する手法の提案を行った．感情の関係を分析するために，感情をノードとして共起確率が閾値以上の感情を接続して有向グラフを作成する．この有向グラフを感情グラフと呼び，この感情グラフを用いて感情の関係を分析する．まず，2 つのノード間を比較し，双方向に共起確率が閾値以上の場合，共起関係として

接続する．共起頻度の閾値 を 0.35 として 10 軸の感情グラフを作成した．その結果を図 1 に示す．

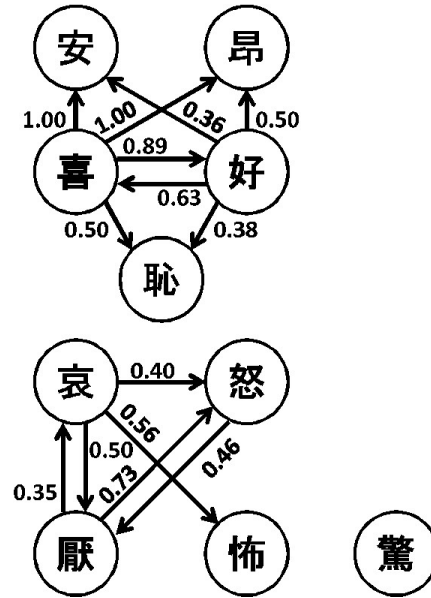


図 1. 感情グラフ

図 1 の感情グラフを用いて感情の関係を分析する．まず，「喜」と「好」の感情のペアを例に感情の結合の流れを説明する．「喜」を対象とすると，「好」との共起確率が 0.89 となり，閾値以上なので，共起関係「喜」「好」として接続する．逆に，「好」を対象とした場合，「喜」との共起確率が 0.63 となり，閾値以上のため，共起関係として接続する．よって，「喜」と「好」は双方向に接続し，このような場合に共起感情として結合する．同様の処理をすべてのノードに対して行った結果，「喜」と「好」の感情のペアと「厭」と「怒」の感情のペアが結合する．この感情のペアが結合したのは，ツイートという短い文において複雑な感情を読み解くことが困難な為と考えられる．例えば，「今日は楽しかった」というツイートの場合，ツイートが短すぎるため複雑な感情が表現できずポジティブな感情である「喜」や「好」が共起したと考えられる．また，「安」と「昂」の感情は「喜」の感情と片方向にのみ強い接続があった．具体的には，「喜」「安」と「喜」「昂」は共起確率が 1.0 となったが，「安」「喜」と「昂」「喜」は共起確率が 0.35 以下となった．これは安らいでいるツイートは必ず喜んでいるが，喜んでいるツイートは必ずしも安らいでいるわけではないという結果である．このことから「安」と「昂」の感情は単独で存在することがなく，他の感情を補助する役割があるのではないかと考えられる．また，共起頻度の閾値 が 0.35~0.10 までの感情の結合結果， の値を小さくすると多くの感情が結合している．また， の値を

0.10 にすると大きく分けてポジティブとネガティブに感情が2分割されている。しかし、=0.10 の第1結合はポジティブを表す感情が結合しているが、感情「恥」が第1結合に含まれている。これは閾値を下げすぎたため不要な感情まで結合してしまったのではないかと考えられる。

結果、Twitter用の感情軸として「喜、好」、「安」、「昂」、「哀」、「怖」、「怒、厭」、「驚」、「恥」の8軸を決定した。

(4) SNSからの耳より情報抽出手法の提案

本研究では、口コミ情報のSNSから耳より情報を含む文を抽出する手法を提案した。具体的には、以下の手順で耳より情報を抽出する。

- 1) 口コミ情報を文単位で分割
- 2) 1)の中で我々の提案する耳よりキーワードが含まれている文すべてを耳より情報の候補とする。
- 3) 2)で抽出された耳より情報の候補を形態素解析する。
- 4) 3)の結果から、耳よりキーワード以外の名詞を用いてクラスタリングを行う。
- 5) 各クラスタ内でクラスタの中心ベクトルとそのクラスタを構成する文の類似度を求める。
- 6) 類似度がある閾値の幅内にある文を耳より情報とする。

同じユーザの口コミの内容には感想等様々な情報と耳より情報の文が混在しているので、本研究では各口コミを文単位に分割する。文章中に「。」、「.」、「!」、「!」、「?」、「?」、「」または改行が一つ以上の場合にそこが文の終わりと考え分割した。

本研究では、旅行の耳より情報を抽出するための、分類及びそのキーワードを人手によって決定した。表1に分類と各々の耳よりキーワードの一部を示す。

表1 耳より情報の分類とキーワード

分類	耳よりキーワード例
提案耳より	おすすめ、方がいい、良いと思、いいかと、いいのでは、良いでしょう、良くない、すべき、するべき、いかが、間違いない、よさげ、どうでしょう
お得耳より	無料、タダ、安い、売り切れ、必須、必見、ポイント、スムーズ、混雑、役立つ、見所、渋滞、禁止、裏道、近道、特典、注意、警告、ルール、手軽、セオリー、コツ、開始、解消、
時間情報	春、夏、秋、冬、朝、昼、夕
天気	晴れ、曇り、雨

公知でないレアな情報はユーザが得をしたと感じる耳より情報ではないかと考え、このレアな情報を抽出することを行った。この時、あまりにもレアな情報は旅行者にとって有用でない場合がある。そこで本論文では、他の

耳より情報の候補と類似する話題について述べながら、他の候補と少し異なる事を述べている口コミを耳より情報と定義した。つまりは、話題ごとにクラスタリングを行い、そのクラスタ内の中心ベクトルからある程度離れている範囲にある文を耳より情報とした。

具体的には、手順2にて抽出した耳より情報の候補となるすべての文を形態素解析し名詞を抽出した。そして、耳よりキーワード以外の名詞を対象としてクラスタリングした。ここでクラスタリング手法として、Repeated Bisectionを用いた。クラスタ数は実験より50とした。ここでクラスタの中心ベクトルとクラスタ内の文各々の類似度を求め、類似度がある閾値の幅内にある文を耳より情報とした。このとき閾値の幅は、ある値との間であるとした。本研究では、実験より、を0.5を0.6とした。

<参考文献>

1. T. Kumamoto, "Design of Impression Scales for Assessing Impressions of News Articles", Lecture Notes in Computer Science, LNCS6193, Springer, pp.285-295, In International Workshop on Social Networks and Social Media Mining on the Web (SNSMW'10), 2010.
2. 中村明 .感情表現辞典 .東京堂出版 ,1993
3. R. Plutchik, "The nature of emotions, American Scientist", Vol.89, pp.344-355, 2001.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 6件)

1. Yuki Yamamoto, Tadahiko Kumamoto, Akiyo Nadamoto, "Multidimensional Sentiment Calculation for Twitter based on Emoticon", International Journal of Pervasive Computing and Communications, 10pages, 2015, 査読有り。
2. Tadahiko Kumamoto, Hitomi Wada, Tomoya Suzuki, "Proposal of a system for visualizing temporal changes in impressions from tweets", International Journal of Pervasive Computing and Communications, 10pages, 2015, 査読有り。
3. Yuki Hattori, Akiyo Nadamoto, "Tip information from social media based on topic detection", International Journal of Web Information Systems, pp.83-94, 2013, 査読有り。
4. 熊本 忠彦, "記事の閲覧順序がその印象に及ぼす影響の分析", 知能と情報 (日本知能情報ファジィ学会誌), Vol. 25, pp. 636-640, 2013, 査読有り。

5. **Akiyo Nadamoto** and Koichi Takaona, "Words-of-wisdom Search based on User's Sentiment", Intelligent Control and Innovative Computing, Springer, pp. 217-232, 2012, 査読有り.
6. Kouichi Takaoka, **Akiyo Nadamoto**, "Words-of-Wisdom Search System based on User's Desired Sentiment", International Journal of Business Intelligence and Data Mining , Vol. 7 No.3, pp. 172-185, 2012, 査読有り.

〔学会発表〕(計 10 件)

1. Yuki Yamamoto, **Tadahiko Kumamoto**, **Akiyo Nadamoto**, "Role of Emoticons for Multidimensional Sentiment Analysis of Twitter", The 16th International Conference on Information Integration and Web-based Applications & Services(iiWAS2014), pp.107-115,2014, 査読有り.
2. **Tadahiko Kumamoto**, Tomoya Suzuki, and Hitomi Wada, "Visualizing Impression-Based Preferences of Twitter Users", Proc. of the 16th International Conference on Human-Computer Interaction (HCII '14), LNCS 8531, pp. 209-220, Heraklion, Crete,Greece,(June 2014), 査読有り.
3. **Tadahiko Kumamoto**, Hitomi Wada, and Tomoya Suzuki, "Visualizing Temporal Changes in Impressions from Tweets, The 16th International Conference on Information Integration and Web-based Applications & Services(iiWAS2014), pp.116-125,2014, 査読有り.
4. 山本 湧輝, **熊本 忠彦**, **灘本 明代**, "顔文字の役割に着目したツイートの多次元感情抽出手法の提案", ARG 第 4 回 Web インテリジェンスとインタラクシオン研究会, pp. 27-32., 2014.
5. 服部 祐基, **灘本 明代**, "ソーシャルメディアからのタイプ別耳より情報の抽出手法の提案", 第 5 回データ工学と情報マネジメントに関するフォーラム, online <http://db-event.jpn.org/deim2013/proceedings/pdf/c1-6.pdf>, 7 pages, 2013.
6. 若井 祐樹, 田中 美羽, **熊本 忠彦**, **灘本 明代**, "顔文字を考慮したニュースに対するツイートの感情抽出手法の提案", 第 5 回データ工学と情報マネジメントに関するフォーラム, online <http://db-event.jpn.org/deim2013/proceedings/pdf/d9-5.pdf>, 6 pages, 2013.
7. 若井 祐樹, **熊本 忠彦**, **灘本 明代**, "ツイートの感情抽出の為に顔文字

- の役割分類", ARG 第 2 回 Web インテリジェンスとインタラクシオン研究会, pp.59-60, 2013.
8. 若井 祐樹, **熊本 忠彦**, **灘本 明代**, "映画に対する実況ツイートの感情抽出手法の提案",情報処理学会研究報告: データベースシステム, Vol. 2013-DBS-158, No. 16, pp. 1-6, 2013.
9. Yuki Hattori, **Akiyo Nadamoto**, "Extracting Tip Information from Social Media", The 14th International Conference on Information Integration and Web-based Applications & Services(iiWAS2012), LNCS7235, pp.158-169, 2012, 査読有り.
10. 服部 祐基, **灘本 明代**, "トピック推定に基づくソーシャルメディアからの耳より情報抽出手法の提案", 第 5 回 Web とデータベースに関するフォーラム (WebDB Forum 2012), 2012, 査読有り.

〔図書〕(計 1 件)

Akiyo Nadamoto and Yuki Hattori
Chapter 43. Categorized tip information of social media based on topic detection
IAENG Transactions on Engineering Sciences,
ISBN-13: 978-1138001367, CRC press,
pp. 404-410, 2014

〔産業財産権〕
出願状況 (計 0 件)
取得状況 (計 0 件)
〔その他〕
ホームページ等

6. 研究組織
(1) 研究代表者
灘本 明代 (Akiyo Nadamoto)
甲南大学・知能情報学部・教授
研究者番号: 30359103
- (2) 研究分担者
熊本 忠彦 (Tadahiko Kumamoto)
千葉工業大学・情報学部・教授
研究者番号: 30358890