# 科学研究費助成事業　　研究成果報告書

平成 ２６ 年　 ６ 月　 ９ 日現在

機関番号：１６１０１
研究種目：基盤研究(A)
研究期間：2010～2013
課題番号：２２２４００２０

研究課題名（和文）人間が感じる「不自然さ」感性の多属性文脈解析法とＷｅｂ有害情報判定への応用

研究課題名（英文）Multiattribute context analysis to "unclarity" sensibility and applications to malicious information judgment on the Web

研究代表者
青江　順一（AOE, Jun-ichi）

徳島大学・ソシオテクノサイエンス研究部・教授

研究者番号：９０１０８８５３

交付決定額（研究期間全体）：（直接経費）　　31,300,000円、（間接経費）　　9,390,000円

研究成果の概要（和文）：インターネットは広く社会に普及してきているが、掲示板などには多くの有害情報（誹謗中傷、危険、犯罪など）が存在し、社会問題となっている。この有害情報を人手で抽出することは不可能であり、自動検出手法が必要である。
本研究では、「自然な表現」を大規模に収集し、その知識を活用して、不自然な有害情報を自動検出する手法を確立した。実験結果により、70%から80%の自動検出成果を得た。

研究成果の概要（英文）：Internet has spread widely in the world.  On the other hand, there is much malicious information such as abuse, danger and crime in bulletin boards and this becomes a social problem. An automatic method of detecting the malicious information should be required, because it is impossible to extract it by hand.
In this research, "natural and clear expressions" is collected on a large scale and the detecting rules built from these expressions is utilized to extract malicious information automatically.  According to experimental results, the achievement of automatic detection is from 70% to 80%.

研究分野：総合領域

科研費の分科・細目：情報学，感性情報学・ソフトコンピューティング

キーワード：感性情報処理　文脈感性情報　有害情報フィルタリング

## １．研究開始当初の背景

円滑な感性コミュニケーション技術の確立は感性情報処理の重要な課題であるが、国内外の研究動向は、感情「喜怒哀楽驚き」や感性「良い・悪い」の研究が中心である。感性「自然さ」は「円滑な」コミュニケーションの規範となり、逆に、「不自然さ」は「驚きや逸脱感」の規範となる新しい感性情報処理の学術研究の位置づけとなる。

また、社会問題である電子メールや掲示板（広義のコミュニケーションと捉える）の有害情報への「不自然さ」の適用は、「メディア情報学」、「知能情報学の自然言語処理と知能情報処理」の研究に関係する。特に、「安全・安心な社会」を目標とする国の科学技術重点戦略に関係する基礎研究の位置づけとなる。

## ２．研究の目的

（１）人間が感じる「不自然さ（自然さ）」に対する新しい感性規範の学術研究を確立する。

（２）感性規範要素として、常識表現判定の意味共起感性、焦点一貫性判定の話題感性、真実判定の固有実体感性を導入し、これら多属性感性による文脈解析手法を提案する。

（３）人間生命に関わる違法（薬物・銃器販売)、危険（殺人・爆破告知）のインターネット上の有害情報判定に的を絞り、「不自然さ」を判定する多属性感性の知識構築法と定量化を研究する。

（４）既設の大規模分散解析装置を活用し、また従来から継続する大規模コーパスに対する感性情報処理の研究成果を利用して、Ｗｅｂ上約１億件規模に対する実験と改善を実施する。

## ３．研究の方法

### （１）「不自然さ」感性と有害度を結びつける多属性感性要素の知識構築手法の提案

違法売買では対象を隠語（「麻薬」は「クリスタル、スピード」、「拳銃」は「レンコン、北京ダック」など）で書くが、これらを含めて図１の例で説明する。図１では、例の 1A-1C が違法販売を表し、例の 2A-2C が危険告知を表す。また、例の前に付記した記号●と▲が有害情報を表し、○と△が非有害情報を表す。本報告で使用する、非有害情報とは、有害でない情報の総称として使用する。

図１の例●1A では、「麻薬」の感性表現「爽快な」を隠語に使った非常識な共起「爽快な＋クリスタル」、「会社名や連絡先」の実体が実存しないこと（仮定)、分散した話題＜材料＞、＜食べ物＞の焦点ぼけによって「不自然さ」を感じる。

逆に、例▲2A では、非常識で危険な共起「新幹線＋爆破」を明確な実体「日時や場所」で「告知」している点において、「不自然さ」を感じる。以上、本研究では、意味共起、話題、実体情報の感性要素の知識構築手法を違法売買と危険告知の判定に的を絞って明らかにする。

### （２）多属性感性情報を用いた「不自然さ（有害度）」の文脈解析による決定法の確立

文脈解析による文書 $d$ の定量的有害度を $\xi(d)=\alpha(x,d)+\beta(y,z)+\gamma(\mathrm{F}(d))+\delta(\mathrm{N}(d))$ とする。

①$\alpha(x)$：$d$ に含まれる $n$ 個の有害語 $x$ による語彙有害度を定義する。

具体的には、拳銃やナイフなどの表現が中心となり、直接的な危険表現を広く収集して、危険度を判定する場合に適している。特に、「新幹線を爆破する」のような告知的な有害情報には、この語彙的な有害度が重要な情報となる(図１の例▲2A と▲2B)。

例(●1A)
AB 商会 **@ab.com 爽快なクリスタル。レンコンも強力だ。
*{隠語の不自然な共起表現、話題分散の焦点ぼけ、実体なし}*
例(○1B)
美しいクリスタル製品を特価売出し。CDF（株）電話 123-456。
*{自然な共起、話題が明確で焦点ぼけなし、実体あり}*
例(○1C)
レンコンの繊維は健康に良い。新宿駅地下のテンプラ試食はおいしい。*{同上}*
例(▲2A)
新幹線を爆破する。5 日午後、品川発だ。
*{危険で異常な共起表現の爆破告知、日時、場所の実体も明確に告知}*
例(▲2B)
サバイバルナイフを買った。明日、秋葉原で決行する。
*{危険物を所有する共起表現、日時、場所の実体も明確に告知}*
例(△2C)
道路の爆破工事です。明日、新天王山で決行です。
*{危険物ワードを含むが、危険でない道路工事の話題内容}*

**図１　有害と非有害情報の例**

②$\beta(y,z)$：$d$ に含まれる意味共起(y,z)による感性有害度を定義する。

本研究で中心となる意味共起による知識表現であり、「爽快な＋気分」のような人間の感情や感性に相当する単語と対象単語の意味的な関係である。日常的によく利用される意味共起表現は、一般的で「自然な」表現

であり、この自然な表現を大規模に収集することで、造語により変化する隠語表現「クリスタル」に対する不自然な表現が検出できる。例えば「爽快なクリスタルあるよ」の「爽快な＋クリスタル」が頻度的、あるいは意味的に「不自然さ」が生じる可能性が高く、有害候補として浮かび上がらせる効果がある（図1の例●1A）。

③γ(F(*d*))：*d*に含まれる連想語から判定した話題集合 F(*d*)による話題有害度を定義する。

具体的には、語彙的な有害表現（殺人や爆破など）は、マンガや映画のシーンでは日常的に使用される表現である。この情報が掲示板や SNS(ツイッターなどのソーシャルネットワークサービス)に掲載されると、語彙的な有害判定が強くなる。この場合は、文脈的な情報である話題知識を利用して、有害度判定の強さを弱める処理を実施する。従って、この話題有害度は、語彙的に間違った解釈を話題によりフィルタリングすることが必要となる（図1の例○1C)）。

④δ(N(*d*))：*d*に含まれる固有実体表現集合 N(*d*)から決定した実体有害度を定義する。

有害情報で大きな問題は、密売であるが、実体として存在しない組織名や電話番号を記載することで、情報の信憑性を確保しようとする場合があり、有害情報判定では、これら固有表現が実際に存在するかどうかを確認する必要がある。図1の例●1A における組織名「AB 商会」やメールアドレス「**@ab.com」などがこの事例に相当する。本研究では、試験的に固有表現を収集し、組織や企業名の辞書を構築し、解析結果に実体のある固有表現情報を出力できるようにして、実体判定を行うものである。

**（3）大規模分散解析装置により、1億件規模で有効性を実験実証する。**

既設の大規模分散解析装置を活用し、また従来から継続する大規模コーパスに対する感性情報処理の研究成果を利用して、4 年間でＷｅｂ上約１億件規模に対する実験と改善を実施する。このコーパスに対して、上記の意味共起による感性情報を収集し、大規模な「自然さ（不自然でない)」の表現を収集する。

表１には、意味共起による感性情報の結果の一部を頻度の多い順位に示す。「悪い」感性が多く、地震や災害などに詳細化できるものもあるが、これらは後述の多属性照合の概念規則の分類として構築する。

**表１　感性の自然な表現頻度結果（一部）**

| 感性 | 頻度 |
|---|---|
| 悪い | 24,649 |
| 良い | 11,514 |
| 不満 | 6,551 |
| 賞賛 | 4,406 |
| 体調が悪い | 3,393 |
| 批判 | 3,156 |
| 満足 | 2,536 |
| 不快 | 2,444 |
| 悲しい | 2,427 |
| 嬉しい | 1,884 |
| 怒り | 1,517 |
| 快い | 1,394 |
| 不安 | 1,211 |
| 好き | 1,067 |
| 美味しい | 1,034 |
| 楽しい | 1,017 |
| 恐怖 | 996 |

## ４．研究成果

**（１）大規模コーパスからの感性意味共起の構築**

本研究では、まず大規模コーパスを段階的に収集し、意味共起を中心とした「自然な表現」を収集した。コーパスとしては、目標の1億文収集結果から約 1,000 万文の掲示板などに使用される対話的な表現を抜粋した。さらに、この中から感性情報による意味的な分類として、約1万の詳細化意味分類を行った。この詳細化分類に対する意味共起関係で照合できるかどうかで、「自然さ」と「不自然さ」を判定する多属性照合手法を確立した。

以上の「自然さ」と「不自然さ」を判定に関連する発表論文は、5,7 が関係する。

**（２）不適切表現検出精度**

データ１（２ちゃんねる)、データ２(Yahoo! 掲示板)、データ３（学校裏サイト）の総投稿数約 2,000 に対する、不適切表現（誹謗中傷、差別、猥褻、苛立ちなど）の検出精度を図２と図３に示す。図２のクローズデータでは平均適合率 85.5%、平均再現率 94.3%となり、図３オープンデータでは、それぞれ 81.5%と 83.0%となった。

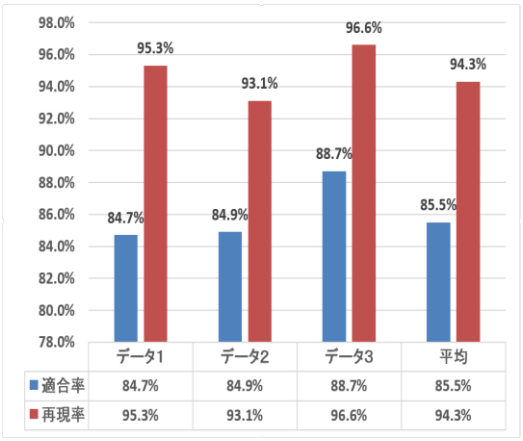


**図２　不適切表現検出クローズデータ精度**

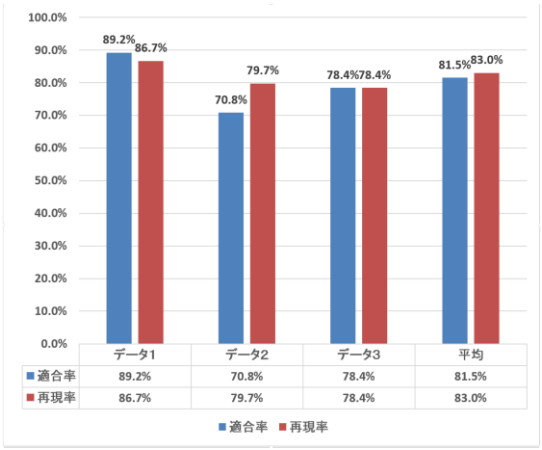


**図３　不適切表現検出オープンデータ精度**

以上、提案手法では、約 80%以上の不適切表現が自動抽出でき、適合率も 80%以上であるので、膨大なデータから人手で探す労力は大きく軽減できると思われる。

適合率が下がった要因としては、以下の過剰検出が挙げられる。この投稿は、“君は氏ね”という表記により不適切表現として検出されていた。しかし、このような投稿の場合、不適切表現の直後に“もアウトですよ”と続き、相手に対する注意を促す内容になっていることから、≪不快≫として検出すべきではない。

１．無党派さん：
君は氏ね もアウトですよ
【検出結果】：**不適切表現≪不快／誹謗中傷≫**
：君は氏ね

２．ばかがいこつ：
＞＞94 バーカ
はーぁ，子供だね
【検出結果】：**不適切表現≪不快／苛立ち≫**
：バーカ

また、２つ目の投稿でも、“バーカ”という表記により不適切表現として検出されている。

以上の問題に対する解決策として、不適切表現≪不快≫の前後に存在する注意や引用符などにより、≪不快≫を打ち消す規則を導入した。これら打ち消し情報を定義することで、不適切情報の検出を無効にする仕組みを試作し、方法論として解決した。

### （３）犯罪表現検出と危険判定の評価

実験データとして、犯罪表現が含まれる事件簿（http://netjikenbo.no.land.to/hy/矢野さとる：犯行予告の収集・通報サイト 予告.in　http://yokoku.in/）のクローズデータ 250 文とオープンデータ 540 文を評価した。

犯罪表現検出の実験結果を図４に示す。犯罪表現の検出精度に関して、クローズデータで関連手法（本研究の前の科学研究（B）の手法）は適合率 100%、再現率 65.4%という結果になった。これに対して、提案手法による適合率は 98.3%、再現率は 87.7%という結果となり、適合率が若干低いものの、再現率に関して関連手法より 20%以上高い結果を得られた。

また、図５のオープンデータに関しては、関連手法では適合率 99.3%、再現率 55.6%という結果になった。それに対して、提案手法による適合率は 95.2%、再現率は 76.7%となり、適合率が低下しているものの、関連手法と比較して再現率が 20%以上向上しており、大規模な WEB 空間から人手で有害情報を漏れなく探す目的として、効果が出ているといえる。

なお、以上の実験における検出知識は表１に示す「自然な感性情報」の意味共起知識が基盤として利用されており、本実験のクローズデータによる学習は、この意味共起知識を活用したものである。

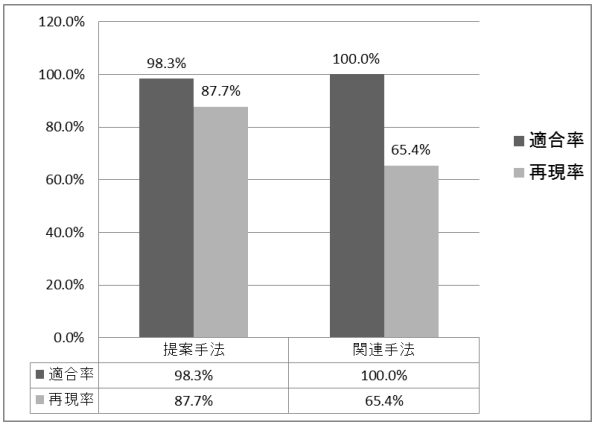


図４　**犯罪・危険判定のクローズデータ精度**

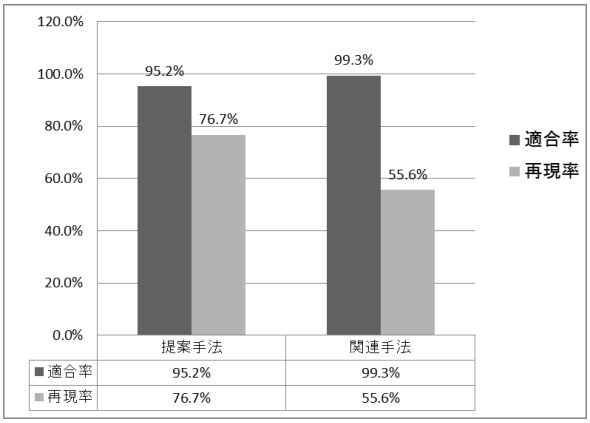


図５　**犯罪・危険判定のオープンデータ精度**

適合率が下がった要因としては、犯罪表現≪殺人・暴力≫に関する過剰検出が挙げられる。以下に過剰検出の例を示す。

**≪殺人・暴力≫**
名無しさん@九周年：○○(金) 19:11:00 ID:dRk2uqD+0
鬼畜殺人レストランの○○は直ちに全国の系列店の営業を停止しろ。さもないと○○に爆弾を仕掛ける。明日からの連休で全国の系列店を回る。客に被害を出したくなかったら営業をやめろ。
【検出結果】：犯罪分類：殺人，爆破／人物・日付・場所指定
**過剰検出表現≪殺人・暴力≫**：殺人
犯罪対象≪場所/組織名≫：○○
犯罪表現≪爆破・放火≫：爆弾を仕掛ける
犯罪対象≪日付/日付≫：明日，犯罪対象≪人物≫：客

この例の網掛け部の“殺人”という表記を、犯罪表現≪殺人・暴力≫として検出していた。しかし、このような投稿の場合、レストラン“○○”に対する批判の意味を込めた内容であるため、人に対する殺意の意味はないので、詳細な打ち消し規則の構築が必要となる。

全体の評価として、犯罪表現検出はオープンデータでは再現率が低下しているものの、組み合わせや当て字、隠語を用いた表現も検出できているので、本手法の有効性を示せたといえる。一方、危険判定は犯罪対象の検出方法に課題が残ったものの、逮捕・書類送検の判断材料となる犯罪対象の数を用いた判定は 70%前後の正解率となっており、犯罪対象の未検出は少ない。以上の結果から、犯罪対象の過剰検出を抑えることにより、本手法における危険判定の有効性を高めることができると考えられる。

以上、有害情報に関する発表論文は、4,6 が該当する。

### （４）有害表現の抽出知識

提案手法の最大の特徴は、多属性照合規則の効率的な構築手法であり、嫌がらせ、わいせつ、薬物、事件を含む表現を検出する概念規則である。表２は、概念中の要素である表層的な表現を組み合わせた表現パターンで算出した知識数である。例えば、概念＜乗り物＞の要素”バス”などの数が 100 個、＜破壊＞の要素”爆破”などの数が 50 個の場合、二つの概念による意味共起規則＜乗り物＞＋＜破壊＞のパターン数は 5,000 個算出している。表２では、以下の略称を利用する。

WORD：語表現数
RULE：多属性ルール数
PAT：パターン数

**表２　有害情報（不適切、犯罪・危険）で活用する概念規則の内容**

| | WORD | RULE | PAT |
|---|---|---|---|
| 不適切 | 16,239 | 1,281 | 12,875,138 |
| 犯罪・危険 | 12,681 | 1,378 | 9,486,523 |
| 合計 | 28,920 | 2,659 | 22,361,661 |

表２は、多属性照合手法による概念規則の効率的な記述を評価するものであり、RULE の多属性概念ルール数は非常に少なくとも、多くの PAT（パターン数）が定義できることが分かる。

以上の多属性照合エンジンのアルゴリズムに関係する発表論文は、1,2,3,8 が該当している。

### （５）SNS 情報に対する有害情報処理

本研究期間の途中からツイッターなどの SNS(ソーシャルネットサービス)の情報が爆発的に増加し、ストーカー犯罪の検出遅れなどが社会的な問題になってきた。

研究の最終年度では、有害情報を検出するために非有害情報をフィルタリングする手法の取り組みをした。この手法では、キーワード表現（提案手法の語彙的有害度）を採用した。友人同士の雑談であれば有害情報から除外できる新しい知見が得られた。また、映画や漫画の話題であれば、提案手法の話題有害度と同様に有害情報から除外できることがわかった。例えば、キーワード「ストーカー」で収集されたツイッターのフィルタリングでは、友人同士の雑談において、単純に文末に「w（笑いの意味)」が付く場合は除外でき、映画や漫画の話題でも除外できたので、提案手法の有用性は、媒体が異なっても効果的であることが分かった。短期間の小規模な実験であるが、キーワード「ストーカー」によるツイッターに対するフィルタリング評価では、適合率 84.9%、再現率 72.9%となった。

## 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 8 件）


(1) Masao Fuketa, Toshiyuki Tamai, Kazuhiro Morita, Jun-ichi Aoe, Effectiveness of an implementation method for retrieving similar strings by trie structures, International Journal of Computer Applications in Technology, 査読有, Vol.48, No.2, 2013, pp.130-135
DOI: 10.1504/IJCAT.2013.056019

(2) Kazuhiro Morita, Hiroya Kitagawa, Masao Fuketa, Jun-ichi Aoe, An incremental construction method of a large-scale thesaurus using co-occurrence information, International Journal of Computer Applications in Technology, 査読有, Vol.48, No.2, 2013, pp.120-129
DOI: 10.1504/IJCAT.2013.056018

(3) Kazuhiro Morita, Takuki Ogawa, Hiroya Kitagawa, Masao Fuketa, Jun-ichi Aoe, A method of extraction and visualisation for relationships among objects on web, International Journal of Intelligent Systems Technologies and Applications, 査読有, Vol.12, No.3/4, 2013, pp.316-327
DOI: 10.1504/IJISTA.2013.056541

(4) Toshihiro Satomi, Atlam EL-Sayed, Kazuhiro Morita, Masao Fuketa, Jun-ichi Aoe, A Context Analysis Scheme of Detecting Personal and Confidential Information, International Journal of Innovative Computing, Information and Control, 査読有, Vol.8, No.5(A), 2012, pp.3115-3134
http://www.ijicic.org/ijicic-11-03075.pdf

(5) Takuki Ogawa, Hiroaki Bando, Kazuhiro Morita, Masao Fuketa, Jun-ichi Aoe, A Method of Combination of Language Understanding with Touch-Based Communication Robots, International Journal of Intelligence Science, 査読有, Vol.2, No.4, 2012, pp.71-82
DOI: 10.4236/ijis.2012.24010

(6) Hiroshi Hanafusa, Kazuhiro Morita, Masao Fuketa, Jun-ichi Aoe, A method of extracting malicious expressions in bulletin board systems by using context analysis, Information Processing & Management, 査読有, Vol.47, No.3, 2011, pp.323-335
DOI: 10.1016/j.ipm.2010.08.003

(7) Li Wang, Masao Fuketa, Kazuhiro Morita, Jun-ichi Aoe, Context Constraint Disambiguation of Word Semantics by Field Association Schemes, Information Processing & Management, 査読有, Vol.47, No.4, 2011, pp.560-574
DOI: 10.1016/j.ipm.2011.01.001

(8) Masao Fuketa, Atlam EL-Sayed, Nobuo Fujisawa, Hiroshi Hanafusa, Kazuhiro Morita, Jun-ichi Aoe, A fast search method of similar strings from dictionaries, International Journal of Computer Applications in Technology, 査読有, Vol.40, No.4, 2011, pp.265-272
DOI: 10.1504/IJCAT.2011.041655


〔学会発表〕（計 1 件）


(1) Kazuhiro Morita, Masao Fuketa, Jun-ichi Aoe, A Cloud Based Communication System for Elders Using Dialogue Control, 6th International Conference on Computer Science and Information Technology(ICCSIT2013), 2013年12月20日～12月21日, Timhotel Berthier Paris (France, Paris)


## ６．研究組織

(1)研究代表者


青江 順一（AOE, Jun-ichi）
徳島大学・ソシオテクノサイエンス研究部・教授
研究者番号：９０１０８８５３


(2)研究分担者


泓田 正雄（FUKETA, Masao）
徳島大学・ソシオテクノサイエンス研究部・准教授
研究者番号：１０３０４５５２

森田 和宏（MORITA, Kazuhiro）
徳島大学・ソシオテクノサイエンス研究部・講師
研究番号：２０３２５２５２