科学研究費助成事業　　研究成果報告書

平成　２６　年　　６　月　１３　日現在

機関番号：１０１０１
研究種目：若手研究(B)
研究期間：2012～2013
課題番号：２４７００２５３

研究課題名（和文）ウェブマイニングによる人間行動の倫理性推論アルゴリズムの構築

研究課題名（英文）Algorithm for an automatic moral judgement of human behavior by using web-mining

研究代表者
ＲＺＥＰＫＡ　Ｒａｆａｌ（Rzepka, Rafal）

北海道大学・情報科学研究科・助教

研究者番号：８０３９６３１６

交付決定額（研究期間全体）：（直接経費）　　2,700,000円、（間接経費）　　810,000円

研究成果の概要（和文）：申請者が本研究において提案したのは，人間行動の倫理性推論アルゴリズムであった．システムの行動は，従来のアルゴリズムのようにプログラマの指示によるのではなく，多くのインターネットユーザが「民衆の知恵」として蓄積してきたネット上のリソースから収 集可能な「倫理的常識」により決定するシステムを構築した．これによりユーザからシステムへのフィードバックは 最小限化され，自動的に日常の行動パターンとその結果の抽出を行う「行為の社会的な結果の評価」及び「行為の感情的な結果の評価 」アルゴリズムを実装し，入力された行動（話題に制限無し）が人間にとって「善」か「悪」かをシステム自身が判断できるようになりました．

研究成果の概要（英文）：The idea proposed in this study was an ethical reasoning algorithm of human behavior. The behavior of the system, is determined by a collection possible "ethical common sense" from Internet resources rather than by the instructions of the programmer as in conventional systems. Many Internet users have accumulated their "folk wisdom" which is the base for the proposed system. Feedback to the system is minimized by utilizing "automatic emotional evaluation of an act's results" algorithm and "automatic social evaluation of an act's results" to extract and the resulting reaction patterns of human behaviors. This helps the system to determine whether the given act is regarded "good" or "bad" by humans (no limit to the topic in the input). For deeper situation understanding five senses simulation and instincts lexicon was added and methods for calculating vectors of Bentham's Felicific Calculus were proposed. Precision achieved was 70-75% on average for each recognition step.

研究分野：情報学

科研費の分科・細目：人間情報学・知能情報学

キーワード：知識獲得　マシンエシックス　人工汎用知能　自然言語処理　コモンセンス

１．研究開始当初の背景

賢くなりつつ機械は我々の生活に不可欠な存在になっているが，選択肢の拡大と共に，文脈の理解不足によっての失敗の可能性も高くなる．その問題に対して，人間の行動をより深く処理するための知識を獲得するアルゴリズムの構築が望ましい．実世界の日常に関する知識が膨大なため，手動入力が不可能であり，コンピュータの処理スピードとビッグデータへのアクセスを利用して自動抽出の手法を研究している．

２．研究の目的

今回は人間の行動の因果関係を獲得する試みであった．哲学と心理学の分野のアイデアを調査し，行動 y の一般的な原因となる本能 x 及びその行動 y とつながる普段の結果 z をブログコーパスから抽出するための辞書

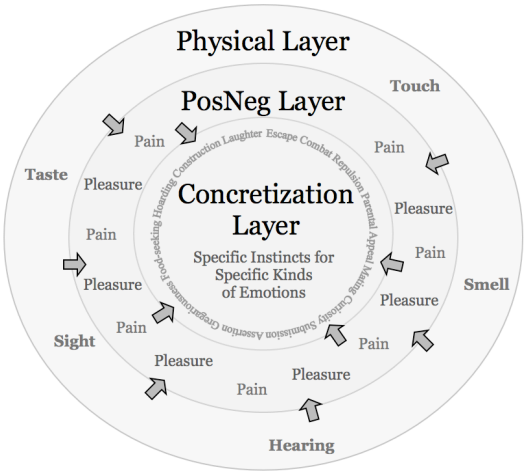

図 1　五感，Pos/Neg 及び本能レイヤーの関係

（lexicons）を提案した．機械はなんで人間がある行動を起こし，その後どういうことが起きうるか，そしてその結果は感情的と社会的に望ましいかどうか予測するのは研究の目的である．最終目的は機械の行動選択が最善になるための計算の要素を抽出することである．

３．研究の方法

（１）人間及び人間が生きる世界を理解するために知恵を経験に基づいて獲得することが有効な方法だと考えられるが，現在のセンサーテクノロジーや画像処理のレベルは不十分であるため，実世界の情報もシチュエーションの理解に不足している．そこでテキストデータで五感からの情報を補い，物理的な物の特徴（形，触感，色，音，匂い）の知識，その物に関する動作（人間の行動），その動作を起こしうる本能，その動作につながる結果，及びその結果の感情的と社会的な影響

図 2　感情結果抽出の例

（ポジティブかネガティブか）の判断を自然言語処理の技術によって行った（図１参照）．以上の知識カテゴリーのパターンマッチングのために作成した辞書を以下の理論に基づいた．

① ５感からの入力の代わりに物理的な特徴を表すのは主に形容詞及びオノマトペだと判断し，「Adj N が見えた」のようなパターンで抽出した単語を５つのカテゴリーに分けた．

② McDougall の本能タイプの研究に基づいて，各タイプに合わせた１４カテゴリーの単語はシソーラスなどを使用して決定した．

③ 以前の研究に使用した中村感情表現辞典の１０カテゴリー及び単語を再利用し，Bentham の幸福計算のアイデアに基づいて感情的な結果の lexicon 及び時間表現を処理する A-Dur システムを用いて各ベクターの計算アルゴリズムを提案した．

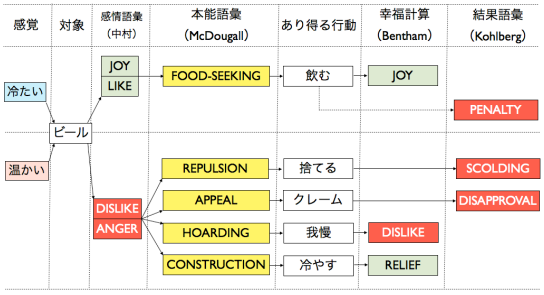

図 3　決定判断の可能な理由と結果の例

④ 社会結果の辞書は Kohlberg の道徳性発達理論を参考にして，感情結果とのバランスを向上させるために１０カテゴリーに分けた．

（２）以前の研究で構築したアメバブログのコーパスに対して自然言語処理の基本的な手法のみを用いてパターンマッチングを行い，原因と結果を抽出し，評価実験を行った．新たに提案した lexicon は以下のものであった．

① McDougall の本能カテゴリー

escape = 怖，心配，恐，不安，悲鳴，鳥肌，不気味,冷や汗，震え，怯，ひゃっ，ぞっと，心細，胸騒ぎ，おびえ,寒心,逃げた方がいい,逃げないと,逃げなきゃ,やばい,危ない
combat = 怒，むかつ，腹が立，腹立，ぷんぷん，腹を立て，むかむか
repulsion = いや，嫌，不愉快，不快，悪，吐き気，嘔吐，気に入らない，恨，むかつく，腐，憎，忌，好かない，厭,気持ち悪い,キモい,まさか
parental = お気に入り，好み，好きで，愛着，好意，愛情，憧れ，好む，親しみ，好ましい，好きだ，愛しい
appeal = 弱い,傷つきやすい，もろい，脆弱,落ち込む,落ち込んで，何もできない，力がない，どうにもできない，無力，動けない，困った，助けのない
mating = 奇麗,ハンサム,セクシー,色っぽい,美人,美男子,性的魅力,つきあいたい,結婚したい，色気
curiosity = 知りたい，調べたい，面白い，普通じゃない，異常，まれな，珍しい，変わった
submission = 真っ赤，恥，屈辱，はずかしい，赤面，照れ，照れ臭，劣っている，粗悪，劣等
assertion = 嬉，笑，爽やか，喜，ありがたい，気楽，快楽，幸，楽しい，ほのぼの，輝，楽しさ，ご機嫌，気持ちがいい，にやり，優れ，誇らしい
gregariousness = 悲しい，泣き，泣く，寂しい，号泣，痛み，悲惨，泣ける，涙，
かわいそう，気の毒，追悼，悲しみ，哀れ，嘆き，傷つけ
food_seeking = 美味しい，おいしい，美味しそう，食べたい，食いたい，食べたがる
hoarding = 欲しい，欲しがる，持ちたい，買いたい，もらいたい，集めたい，失いたくない
construction = 作ってみたい，作りた，創造したい，造りた，創作したい，創設したい，開発したい，生み出したい，産みた,生みた，創作したい，作ってよかった，創造してよかった，造ってよかった，創作してよかった，創設してよかった，開発してよかった，生み出してよかった，産んでよかった,生んでよかった
laughter = 可笑しい，笑える，笑った，笑えた，笑って，すっきり，楽だ，落ち着，さっぱり，のびのび，ほっと，安心，気楽，落ち着き，平安，平穏，安ら，微笑ましい，微笑ん

② Kohlberg の理論に基づいた社会的な結果のカテゴリー

Praises = ほめる，ほめた，ほめられ，ほめて，褒める，褒めた，褒められ，褒めて，誉める，誉めた，誉められ，誉めて，称え，たたえ，愛でる，愛でて，嘉する，嘉し
Reprimands = 叱る，叱られ，叱った，叱って，しかる，にしかられ，をしかって，とがめる，とがめて，戒め，誡め，警め，縛め，叱責，警戒，禁止，戒告，訓戒，諭旨，教戒，勧戒，注意を，譴責，怒，叱咤，叱責，譴責，叱り，一喝，大喝，お目玉，大目玉，諫言
Awards = 賞する，称する，賛する，賞し，称し，賛し，奨励，報酬，ボーナス，持て囃，称賛，称美，称揚，推賞，嘉賞，褒美，賞品，景品，賞金，報奨，景品，なでる，撫で，可愛が，慈し，
Penalties = 刑罰，制裁，実刑，執行猶予，私刑，リンチ，おしおき，お仕置き，折檻，見せしめ，懲らしめ，こらしめ，罰金，懲役，死刑，
Society approval = 正しい，正しか，正しくて，批判できない，同意する，同意している，同意していた，同意できる
Society disapproval = 正しくない，正しくない，正しくなく，正しくなか，批判する，批判した，同意でない，同意できません
Legal = 合法，適法，合憲，リーガル，公認される，公認された，公認されている，公認されていた
Illegal = 違法，不法，違反，非合法，違憲,公認されない，公認されません，公認されていない
Forgivable = 許される，許された，許した，許す，許せた，許せる
Unforgivable = 許せない，許されな，許さな，許せませ，許せな

③　５感シミュレータに利用された単語

touch = 固い,硬い,堅い,べたべたする,冷たい,柔らか,重い,軽い,涼しい，浅い，深い，暖かい,温かい,熱い,暑い,厚い,痛い,薄い,寒い，鋭い，とがっている，鋭利な，ぬるい，乾いた，濡れた,ぬれた，ねばねばした,ぬるぬるした,じめじめした,むんむんする,ざらざらした，つるつるした，ごつごつした，ふわふわした,ぐちゃぐちゃした
hearing = うるさい,静かな,ぽきぽきする,ざあざあする,びりびりする,ごろごろする,ぶくぶくする,カサカサする,カタカタする,ガタガタする,ガチャガチャする,ガチャンする,カランする,ガンガンする,キーキーする,ゴロゴロする,ゴーンゴーンする,サラサラする,ジャンジャンする,チャリンする,チリンする,チリンチリンする,チンチンする,トントンする,バタンする,パタンする,パチ

パチする, パチパチする, パチンする, パチンする, ピシャリする, ポキポキする, リンリンする
sight = 明るい, 暗い, 大きい, 小さい, 赤い, 黄色い, 黒い, 白い, 青い, 緑の, 浅い, 深い, 薄い, 多い, 少ない, 遅い, 速い, 輝かしい, 茶色の, 長い, 古い, 真っ黒な, 真黒な, まっ黒な, 汚い, きれいな, 奇麗な, 四角い, 近い, 遠い, 高い, 低い, 広い, 狭い, 太い, 細い
taste = 甘い, 苦い, 酸っぱい, 辛い, すっぱい, しょっぱい, からい, あまい, にがい, 渋い, しぶい, 甘酸っぱい, 脂肪の多い, 油っこい
smell = くさい, 香ばしい, 生臭い, 芳しい, いい匂いの, いい香りの, 臭い

４．研究成果

各 lexicon に対して評価実験を行い，適合率と再現率を確認した．割と高い適合率はアプローチの正しさを照明したが，利用したコーパスの希望は小さかったため，再現率は全体的な f 値を下げた．今度より大きいデータ及びより深い文脈処理を導入し，適合率と再現率を向上したい．結果は表１とおりである．図２は，感情結果の例を表している．社会的に判断が難しい「安楽死」の意見の割合，「殺す」の対象の違いによる判断の差，問題の知名度も（日本でイルカが殺される情報に対する「驚き」の数字）はっきり抽出されることがわかる．より信頼度の高いシステムを構築すれば，様々な分野（社会学，心理学，経済学など）の研究者に重要なツールになる可能性があると考えられる．

lexicon	適合率	再現率	f 値
５感	0.96	0.43	0.60
本能	0.75	0.46	0.57
感情結果	0.74	0.32	0.45
社会結果	0.70	0.31	0.43

表１ 各辞書の評価実験の結果

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 １ 件）

① Rafal Rzepka and Kenji Araki, “ELIZA Fifty Years Later - The Possibility of an Automatic Therapist Using Bottom-up and Top-down Approaches to Artificial Morality”, to appear in Machine Medical Ethics" in Intelligent Systems, Control and Automation, Springer, 2014.
http://www.springer.com/engineering/robotics/book/978-3-319-08107-6

〔学会発表〕（計 ７ 件）

① Rafal Rzepka and Kenji Araki, “Experience of Crowds as a Guarantee for Safe Artificial Self”, in proceedings of AAAI Spring Symposium on Implementing Selves with Safe Motivational Systems & Self-Improvement, pp. 40-44, Stanford, USA, March 25, 2014.
http://www.aaai.org/ocs/index.php/SSS/SSS14/paper/viewPaper/7730

② Rafal Rzepka, “Pros and Cons of Borrowing Morality from P2P Civilization”, Invited Talk at AAAI Spring Symposium on Implementing Selves with Safe Motivational Systems & Self-Improvement, Stanford, USA, March 26, 2014.
http://www.digitalwisdominstitute.org/SS14Self/Presentations.aspx

③ Rafal Rzepka and Kenji Araki, “Possible Usage of Sentiment Analysis for Calculating Vectors of Felicific Calculus”, In proceedings of IEEE 13th International Conference on Data Mining Workshop “SENTIRE”, pp. 967 - 970, Dallas, USA, December 7, 2013.
DOI: 10.1109/ICDMW.2013.70

④ Rafal Rzepka and Kenji Araki, “Society as a Life Teacher - Automatic Recognition of Instincts Underneath Human Actions by Using Blog Corpus”, in proceedings of SocInfo 2013, pp. 370-376, Kyoto, Japan, November 25, 2013.
DOI: 10.1007/978-3-319-03260-3_32

⑤ Rafal Rzepka and Kenji Araki, “Using Empathy of the Crowd for Simulating Mirror Neurons Behavior”, Invited presentation at the 4th International Workshop on Empathic Computing IWEC' 13, Beijing, China, August 4, 2013.
http://www.ai.sanken.osaka-u.ac.jp/IWEC2013/

⑥ Rafal Rzepka and Kenji Araki, “Web-Based Five Senses Input Simulation - Ten Years Later”, 人工知能学会第 2 種研究会 ことば工学研究会, SIG-LSE-B301-5, pp. 25-33, Sapporo, Japan, July 27, 2013.
http://ultimavi.arc.net.my/banana/Workshop/Programs/rafal_abst2.html

⑦ Marek Krawczyk, Yuki Urabe, Rafal Rzepka and Kenji Araki, “A- dur:

Action Duration Calculation System”, 人工知能学会第２種研究会 ことば工学研究会, SIG-LSE-B301-7, pp. 47-54, Sapporo, Japan, July 28, 2013. http://ultimavi.arc.net.my/banana/Workshop/Programs/krawczyk_abst.html

６．研究組織

(1)研究代表者
ジェプカ ラファウ（RZEPKA, Rafal）北海道大学・大学院情報科学研究科・助教 研究者番号：80396316