

個人 Blog からのエゴグラム推定手法に関する考察

南川 敦宣 横山 浩之

KDDI 研究所

1. はじめに

近年、情報システムやコンテンツのパーソナライズ化が進み、行動履歴からユーザの特性を抽出する技術が注目されている。コミュニケーションや対人・対物行動における振る舞いの特徴を表現する指標の一つにエゴグラム[1]がある。我々は、ユーザのエゴグラムを抽出する行動履歴データとして個人 Blog に着目し、Blog 文章から同執筆者のエゴグラムを推定する手法を検討している。本稿では、500 人の個人 Blog データと、エゴグラムを収集し、使用語句、表現、記号などの特徴量とエゴグラムとの関係を統計的に調査した結果を述べ、エゴグラム推定に有効な Blog 特徴量を考察する。

2. エゴグラムとは

エゴグラムとは、精神科医エリック・バーンによって提唱されジョン・デュセイによって確立された人の自我状態を表すモデルで、交流分析と呼ばれる分野で発達し、使用されている。

エゴグラムは人の自我状態を以下の 5 つの要素で表現する。

- ・ 批判的親 (CP : Critical Parent)
- ・ 養育的親 (NP : Nurturing Parent)
- ・ 大人 (A : Adult)
- ・ 自由な子供 (FC : Free Child)
- ・ 従順な子供 (AC : Adapted Child)

上記要素間の大小関係と、ユーザの対人時、及び対物時における行動の特徴との間には強い相関があることが知られており、臨床心理学の分野では、ユーザ行動の分析やモデル化のための指標として良く用いられている。

エゴグラムの抽出は従来、専門医によるカウンセリングから行うものであったが、近年では東大式エゴグラム (TEG) [2] に代表される質問紙法が主流であり、約 50 問以上の質問への回答結果から、CP, NP, A, FC, AC を 5 段階の数値で評価している。どちらの手法もエゴグラムを取得するために、対象となる人物自身が何らかの試験を受ける必要があり、コンテキストとして使用するには大きな手間となる。エゴグラムの自動抽出手法が必要となる。

3. Blog からのエゴグラム推定手法

カウンセリングによってエゴグラムを抽出する際には、被験者の観察から、エゴグラムの最も強い自我、及び最も弱い自我を決定し、残りの 3 つの要素の相対値を決める [1]。[1] によれば、観察の際に使用する情報は、聴覚、

及び視覚から得られる特徴をさす。聴覚は、話し言葉、表現から得られる特徴、および視覚は見た目や表情、しぐさ等から得られる特徴を指す。

我々はエゴグラム抽出に重要なこれらの情報を取得可能な行動履歴データとして、個人 Blog に着目した。個人 Blog は執筆者が自身の体験や考えを自分の言葉で表現する空間であり、執筆者の個性が現われ易い。具体的にはエゴグラム抽出にあたり使用される視覚、聴覚からの情報は、Blog の以下の特徴に対応づけられる。

- ・ 聴覚：話し言葉 => Blog の文章、表現
- ・ 視覚：見た目、表情 => Blog の文体、絵文字、記号による感情表現

個人 Blog から上記の特徴量を抽出することでエゴグラム推定に必要な指標を得ることが可能である。

より詳細にエゴグラムを推定するために、エゴグラムの各要素が担う特性を以下に示す。

- ・ CP : 義務, 責任感, 批判的, 保守的,
- ・ NP : 受容的, 思いやり, 過保護
- ・ A : 理性的, 客観的, 事実優先
- ・ FC : 自由奔放, 自己中心的, 本能的
- ・ AC : 従順, 協調性, 依存的

つまり、これらの特性に対応する表現を Blog 文章の特徴量として抽出し、エゴグラムの各要素の値の推定に利用することが有効であると考えられる。そこで、我々は、上記特性に対応する表現として、以下の文末表現、感嘆詞、接続詞に着目した。

- 断定表現 (“である”等)
- 記号の使用 (“♪”, “☆”, “!!”等)
- 義務表現 (“べき”等)
- 協調表現 (“ですね”, “でしたね”等)
- 感嘆詞 (“ええ”, “ああ”等)
- 接続詞 (“だが”, “しかし”)

実際にエゴグラム推定を行うには、Blog 文章をまず形態素解析し、これら表現の出現頻度を特徴量とする。各特長量を入力として単純ベイズ分類器等で、エゴグラムの各要素についての値を推定するか、又はエゴグラムのどの要素が支配的かを推定する。

4. 検証

エゴグラム推定に有効な語句を選択するため、上記に挙げた語句の幾つかに関して、エゴグラムの各要素の値にどの程度影響があるかを統計的に調査した。まず 500 人の Blog データを収集すると共に、各 Blog の執筆者に TEG を受診させ、エゴグラムの各要素の値を評価した。次に、各語句の効果を評価する指標として以下に述べる相互情報量と平均値の変化量をエゴグラムの 5 要素に対

“Discussion about the framework for egogram estimation using Personal Blog Text”

A. Minamikawa and H. Yokoyama
KDDI R&D Laboratories, Inc.

表 1 : エゴグラムに対する各語句の出現に関する相互情報量、及び平均値増減

語句	CP		NP		A		FC		AC	
	相互情報量	Avg 増減	相互情報量	Avg 増減	相互情報量	Avg 増減	相互情報量	Avg 増減	相互情報量	Avg 増減
である	0.0091	0.44	0.0012	0.033	0.014	0.45	0.002	0.0012	0.0095	-0.39
す☆	0.0043	-0.07	0.0094	0.77	0.0016	0.24	0.0028	0.22	0.0057	0.18
た!	0.002	-0.07	0.0017	0.05	0.0126	0.16	0.0008	-0.01	0.0062	0.15
(笑)	0.0021	-0.06	0.0013	0.05	0.0036	-0.06	0.0027	-0.006	0.0093	0.13
orz	0.0102	-0.43	0.0045	-0.39	0.0064	-0.32	0.0072	-0.49	0.005	0.35
ええ、ええ	0.014	-0.32	0.0016	-0.18	0.004	-0.01	0.0018	0.23	0.0041	-0.14
うん	0.0034	-0.21	0.0081	-0.42	0.0014	-0.11	0.0033	-0.18	0.0057	0.35
しかも	0.0022	-0.24	0.0009	-0.08	0.0085	-0.08	0.0058	-0.14	0.0122	0.51
だが、しかし	0.0025	0.01	0.0053	-0.06	0.0116	0.12	0.0111	-0.12	0.0042	0.09

して、それぞれ求め、比較した。

(1) **相互情報量** : 対象語句の有無によって、エゴグラムの各要素の値のエントロピーがどのように変化するかを示す。相互情報量は以下の式で記述される。

$$InfoGain(w) = H(C) - H(C|W)$$

この値が大きいほど、エゴグラム推定に使用する効果が大きい。

(2) **平均値の変化量** : エゴグラムの各要素について、全 blog での平均値と、対象語句を含む blog 集合における平均値の差分。表現の有無によってエゴグラムの各要素が増加傾向にあるか減少傾向にあるかを示す。

上記指標の算出例を表 1 に示す。表 1 では、語句「である」に対する CP, A, AC の相互情報量が相対的に大きくなっている。これは、「である」のような断定表現が、自己を強く主張し、表現の客観性を高める際にすかわれることと対応している。平均値は CP と A で大きく増加しており、批判的かつ客観的な特徴を示す一方、AC では減少しており、依存的ではない一面を表現している。また、文末の「☆」等の記号は、NP の推定に有効で、特に NP が高くなる傾向にある。その他、「orz」というような独特の感情表現の場合、AC のみで増加傾向にあり、それ以外の要素は減少する。さらに相互情報量から CP の推定に特に有効であることがわかる。

5. 関連研究

Blog の分類に関する研究としては、出現キーワード(名詞)から、その Blog の内容を幾つかのカテゴリに分類する手法に関するものが多くある。一方、Blog の面白さを推定する手法の研究があり[3]、本手法では、各種表現手法を特徴量として面白さのスコアを推定するおり、使用する特徴量は我々のアプローチに近い。

また、エゴグラムに関する研究としては、[4]のようにマルチエージェントシミュレーションにおいて、各エージェントにエゴグラムを設定したり、ユーザ間のコミュニケーションのシミュレーションに用いたもの[5]など、シミュレーション時の利用用途に関するものがある。またシステムのユーザインタフェースとして、ヒューマノイドエージェントに適応し、利用者の発話からエゴグラ

ムを推定する研究もあるが[5]、具体的に推定手法を評価するまでには至っていない。

6. まとめ

個人 blog で使用される表現から執筆者のエゴグラムを推定するための手法に関する検討を行い、エゴグラム推定に当たって有効な語句の選択手法に関して、実際に収集した 500 人の被験者データで調査を行った。幾つかの語句に関して、エゴグラムの各要素の推定に有効であることが確認できた。今後は全 Blog に出現する全語句に対して網羅的に調査し、実際にエゴグラムの推定精度がどの程度得られるかを調査する予定である。

参考文献

- [1] ジョン・M. デュセイ, “エゴグラム”, 創元社
- [2] 東京大学医学部心療内科 TEG 研究会, “新版 TEG II 開設とエゴグラム・パターン”, 金子書房
- [3] 萩行, 柴田, 黒橋, “表層・語彙的特徴量に基づくブログの面白さ分析”, 情報処理学会自然言語処理研究会報告 2008 pp. 45-52
- [4] 谷藤, 藤田, “マルチ・エージェントへのパーソナリティモデルの適用と集団行動シミュレーション”, 信学技報 TL2002-45, pp17-22
- [5] H. Kobayashi, M. Higuchi, K. Kikuchi, “Development on Interpersonal relations simulator by Egograms”, RO-MAN 2000, pp200-205
- [6] T. Nishiyama, A. Baba, A. Terasawa, S. Hibiya, F. Obayashi, R. Nakajima, “Development of Humanoid Agent System Employing the Knowledge on Transactional Analysis”, RO-MAN 2004, pp455-460