

ラフ集合論に基づくオンライン情報統合化 ～和歌情報可視化システムの提案～

An Online Information Integration System based on Rough Set Theory

六井 淳*¹ 辻野 晃一*² 山本 桃子*¹
Jun Rokui Kouichi Tsujino Momoko Yamamoto

*1 島根大学総合理工学研究科数理・情報システム学専攻

Department of Mathematics and Computer Science Interdisciplinary Graduate School of Science and Engineering Shimane University

*2 島根大学総合理工学部数理・情報システム学科

Department of Mathematics and Computer Science Interdisciplinary Faculty of Science and Engineering Shimane University

In this paper, we propose a method that it extracts semantic structure from Japanese thirty-one syllabled verse (hereinafter collectively called "Waka") by rough set theory. Most studies has said that the syntax analysis has little influence over semantic information extraction from Waka. Waka has thirty-one syllabled japanese words and difficult grammars. Only neither language information nor background information determines the aspect of Waka. Sound information such as a rhythm and strength also affects the aspect of Waka greatly. Compose of the information is discrete, and each relation is not clear. Rough set theory is used to extract correlations from various informations. We suggest that rough set theory is available to extract semantic structure from Waka. A proposed system is expected to be a channel of various nonverbal communications including a free translation.

1. はじめに

来るべき情報爆発に対応するため、情報検索、管理、統合化に向けた研究が行われている [kitsuregawa 07]。膨大な情報の中から有用な情報を効率よく獲得する方法が求められている。特に、地域固有の表現や言語による表現の差異は識別が困難な課題である。従来、表現の差異は「意識」という形式が取られている。しかし、表現そのものが共有できない場合、「意識」では情報を伝えることが困難である。このような表現の差異を埋める手段として有効な方法が情報伝達手段そのものを共有化可能な媒体へ変換することである。例えば、音や色といった感覚表現は万国共通の表現形態であり、情報を画像に変換するビジュアルマイニングなどの研究が行われている [jack, Pak 99]。情報を画像に変換することは、情報表現において有効な手段ではある。しかし、画像への変換方法に依存するため、本質的に意識と差異がない。また、画像への変換規則は与えられる情報表現により千差万別であり、人為的に規定することは困難である。本研究では画像に変換すべき情報をあくまで予備知識と考え、多数のユーザの評価によって変換規則を変化させることで様々な変換規則に対応する枠組みを提案する。曖昧な情報を効率よく分類するためラフ集合理論を使用した。情報統合化システムには、地域や言語による表現の差異を共有化可能であり、且つ表現の変化に柔軟なシステムが求められる。本研究では和歌をタスクとしたオンライン情報統合化システムを提案する。

2. 和歌と情報規則

2.1 システム概要

和歌は日本固有の定型詩であり、複雑な文法や季語といった規則性を持つだけでなく、種々の詠唱法を用いた音声表現によ

る感性の側面を併せ持つ。このため、専門家であっても意味の解釈に大きな隔たりがあり、初学者や異文化の者が意味を理解するのは困難な情報媒体である。本研究では、単行のタスクをとりまく意味情報を視覚情報に集約することで情報を多次元かつ直感的に表現する。和歌の意味情報から意味解釈規則を抽出する方法として、ラフ集合を用いた。ラフ集合は決定表からの知識を決定規則で表現・分類する手法である [pawlak 82]。和歌の文法情報、詠唱（韻）情報、季語情報などの各種情報を条件属性とし、専門家が述べる和歌の意味を決定属性とした。条件属性は和歌の固定情報であるため不変だが、決定属性はあくまで初期値として設定し、ユーザの操作によって変化する。決定表から得られた決定規則を意味解釈規則と定め、これを視覚情報へ変換する。ユーザは視覚情報进行操作することで、意識することなく評価を行う。システムの概要を図 1 に示す。

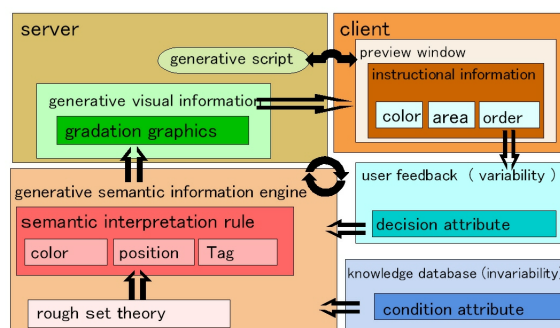


図 1: 提案システムの情報遷移図

2.2 属性の可視化

本研究では複数の和歌を特徴別に球状空間にプロットし、和歌集合を表現する。異なった条件属性の含まれる決定表を Z_x, Z_y, Z_z とする。ある和歌 G における決定表 Z_x の条件属性を a_x, b_x, c_x とする時、これらを決定表 Z_x のために用意し

連絡先: 六井 淳、島根大学総合理工学部数理・情報システム学科、〒 690-8504 島根県松江市西川津町 1060、TEL:0852-32-6842、FAX:0852-32-6489、E-mail:rokui@cis.shimane-u.ac.jp

た任意の正規関数 $f_x(\cdot)$ に代入する。

$$f_x(a_x, b_x, c_x) = I \quad (1)$$

同様に決定表 Z_y, Z_z から

$$f_y(a_y, b_y, c_y) = J \quad (2)$$

$$f_z(a_z, b_z, c_z) = K \quad (3)$$

式 (1) (2) (3) より得られた I, J, K を座標とし、各決定表の決定属性を RGB 値とした基本色を割り当てる。球状空間の位置によって彩度を定め、自然なグラデーションが生成されることになる。図 2 は実際に生成した視覚画像情報を操作する WEB ブラウザ画面である。

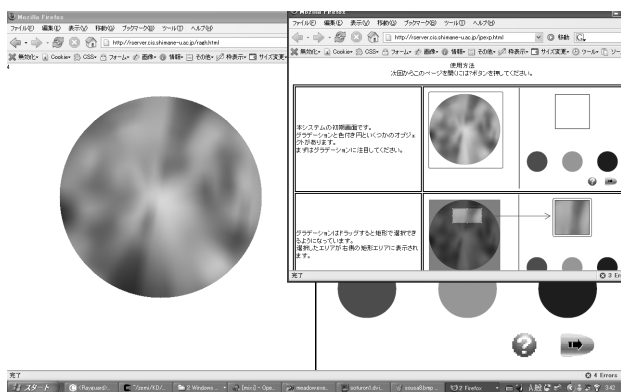


図 2: 視覚情報のクライアント操作画面

3. 理論検証

ラフ集合理論に基づく意味情報抽出エンジンの解析精度について検証を行った。条件属性に利用する学習データには代表的和歌の歌集である百人一首の出典歌集を用いた。評価データとして百人一首を用いた。

3.1 文法情報からの抽出精度

季語の種類と修辞法を条件属性とし、和歌の種類を決定属性と定義しオープン評価実験を行った。修辞法は枕詞、序詞、掛詞、縁語、本歌取り、区切れ、擬人法、見立て、体言止め、係り結び、倒置法の 11 種類である。決定属性は春の歌、夏と冬の歌、恋の歌、その他の 4 種である。学習データには 200 首を用いている。夏と冬に関する学習データ数が他の決定属性に比べて極端に少ないため、実験全体への影響を考慮し統合した。学習によって得られた条件属性との適合を行い、評価データと決定属性が一致した場合のみ正解とした。

ラフ集合を用いる際に決定属性が重複する場合が生じる。決定属性が重複したものについては、寄与率が大きい条件属性を選択した。表 3.1 は適合率である。実験結果より、季語と和歌の因果関係は比較的密接であることが確認された。評価に用いた和歌の 7.1% について、季語と修辞法から意味の分類が可能であることが確認された。

| | | | |
|-----------|-----|-----|-------|
| Attribute | 春 | 夏、冬 | 秋 |
| Accuracy | 66% | 50% | 50% |
| Attribute | 恋 | その他 | 平均適合率 |
| Accuracy | 88% | 64% | 71% |

表 1: 文法情報からの抽出精度

| | | | | |
|-----------|------|-----|-----|-------|
| Attribute | 季節の歌 | 恋の歌 | その他 | 平均適合率 |
| Accuracy | 72% | 79% | 56% | 71% |

表 2: 韻情報からの抽出精度

3.2 韻情報からの抽出精度

母音のパターンを条件属性、和歌の種類を決定属性として実験を行った。和歌の初句、第 2 句、第 3 句、第 4 句、第 5 句の各句の最後の文字の母音をそれぞれ条件属性とした。決定属性は季節、恋、その他の 3 種類である。学習データには古今集、新古今集から 300 首を用いた。表 3.2 は適合率の一覧である。実験結果より和歌の母音と和歌の種類に因果関係があることが示されている。季節の歌は第 1 句の語尾が E、第 2 句の語尾が I である場合が多く、恋の歌は第 1 句と第 4 句の語尾が O、第 5 句の語尾が A である場合が多いことが確認された。作者が韻を作為的に歌に込める事実が本実験によって示されたと考えられる。

文法情報と韻情報を組み合わせることで、9.2% 以上の精度が得られることが実験的に確認できている。これはラフ集合に基づいた条件属性の本提案システムでの高い有意性を表していると考えられる。

4. まとめ

本研究では和歌をタスクとしたオンライン情報統合化システムを提案した。情報の解析にラフ集合を用いることで、蓄積された知識を条件属性としたオンライン評価 (決定属性) を実現することが可能となった。これにより、規定の知識に基づいた統計的評価が可能となるため、和歌や絵画などの一意に評価の定まらない曖昧な情報についても統計的最適解を与えることが可能になる。本提案システムにおいて情報の提示はグラデーション画像により行われる。画像の色彩、位置情報の有意性を実験的に求めた結果、複数の情報を組み合わせて評価することにより有意な画像情報を構築できることが確認された。今後はインタフェース部分の改良を含め、より多くのユーザからの情報統合を進めていく予定である。本システムは URL: <http://rsrserver.cis.shimane-u.ac.jp/> にて公開中である。

参考文献

[kitsuregawa 07] 喜連川 優、安達 淳、田中 克己、下條 真司、黒橋 禎夫: 情報爆発時代における情報管理・融合・活用基盤、Vol22、No2、人工知能学会誌 (2007) .

[jack] ジャック・ベルダン、森田喬訳: 図の記号学-視覚言語による情報の処理と伝達

[Pak 99] Pak Chung Wong : Visual Data Mining, (Vol. 19, No. 5, pp. 20-21 , IEEE Computer Graphics (1999).

[pawlak 82] Z . Pawlak : Rough sets, Internal. J. Inform. Comput. Sci, Vol.11, No.5, pp341-356 (1982).