

# 住民参画 Web プラットフォームにおける コンサーン・アセスメント支援機構

白松 俊 大園 忠親 新谷 虎松

名古屋工業大学 大学院工学研究科

{siramatu,ozono,tora}@nitech.ac.jp

## 1 はじめに

本研究では、地域社会における住民参画 (Public Involvement) の基盤となる住民参画 Web プラットフォーム O<sub>2</sub>\*<sup>1</sup>を開発中である。オープンガバメントや Gov 2.0[1] といった文脈で重要視されている透明性、参画、協働のうち、特に地域社会における透明性と参画に焦点を当て、その基盤としての情報インフラの実現を目指している。

住民参画とは、住民公聴会など公的討議から住民意見を集約し、意思決定に反映させる取り組みである[2]。住民参画が求められる事柄は、災害対策、経済、環境、福祉など多岐に渡るため、全ての問題に通じている住民は少ない。さまざまな問題の解決に向けて議論を深めるためには、住民の参画を通じ、増加していく意見を問題意識・懸案事項ごとにまとめる構造化と、その結果を共有可能にする仕組みが重要となる。このように問題意識を構造化する作業を、コンサーン・アセスメントと呼ぶ。

コンサーン・アセスメントにより、地域に関する問題意識を共有できるようになれば、若い世代等、地域との関わりがあまり深くない住民も新たに参画しやすくなると期待される。このような、地域に浅く関わる人々の参画促進は、地域活性化のために重要である。例えば名古屋市は、H22年度に「名古屋市中期戦略ビジョン」の策定過程にタウンミーティングやパブリックコメント制度を取り入れており、住民参画に積極的に取り組んでいる。タウンミーティングへの参加者によると、積極的に発言する若者は必ずしも多くない。文献[3]によると、『「若者の活力」による魅力あふれるまちづくり』が施策として提言されており、若者の参画の重要性が再確認されている。また行政側でも Web を活用して住民の意見を得る取り組み[4]が実

験的に始まっており、コミュニケーション支援のための技術開発は重要視されている。その一方で、住民行政-専門家間での情報共有を十分に促進できるコンサーン・アセスメント支援技術は未だ実用化に至っておらず、開発が急務となっている。

そのため本稿では、地域コミュニティにおける問題意識や懸案事項をわかりやすく整理する仕組みとして、意見アーカイブを蓄積するための仕組みについて述べる。幅広い意見を収集するため、Twitter 上のツイートを意見収集対象としている。また、意見の構造化に関連して、意見の背景となるイベントを Web 上のニュース記事から検出し、イベントと意見を紐付けるための仕組みを実現している。

これらの Web コンテンツを図1に示すようなサイクルによって住民参画に活用する。地域に関連する Web コンテンツを議論の「種」として収集し、蓄積する。この際、自動処理および人手編集による構造化を行い、相互可用性の高い Linked Open Data (LOD)[5]として公開することで、コンサーン・アセスメントへの活用を可能にする。

## 2 住民参画 Web プラットフォーム

図2は、本研究において開発した住民参画 Web プラットフォームを表している。本システムの中核



図1 住民参画促進のためのサイクル

\*1 <http://open-opinion.org/>

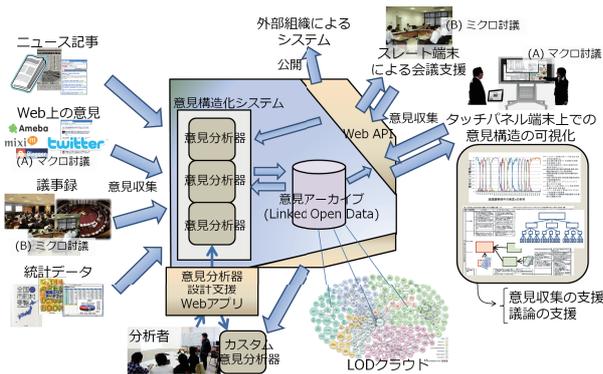


図2 住民参画 Web プラットフォーム O<sub>2</sub>

は、意見構造化システム、および、LOD データセット SOCIA (Social Opinions and Concerns for Ideal Argumentation) から構成されている。意見構造化システムは、自動分析結果と人手のコンサーン・アセスメントの結果を LOD データセットとして蓄積・公開するためのシステムである。本システムは複数の意見分析器から構成されており、例えば、意見構造化システムに含まれる意見分析器の一つは、Twitter から取得したツイートが言及している地域を推定し、ツイートの内容に関する位置情報をツイートに付加した上で、LOD データセットとして蓄積・公開する。これにより、特定の地域に対するツイートを収集することが可能になる。

Web 上から収集した関連情報は、意見という観点から整理するためのスキーマである SOCIA に従って構造化される。意見アーカイブでは、ツイート、ニュース記事、Web クリップ、および、議事録等を、問題意識や懸案事項という観点から整理・蓄積することを狙っている。例えば、Web クリップでは、ユーザの Web 閲覧における観点を考慮した Web ページ分割技術を利用することにより、特定の問題意識や懸案事項を考慮した Web 情報の再構造化が可能になる。

### 3 コンサーン・アセスメント支援のための意見構造化

本研究における意見構造化とは、地域に関する意見や背景情報を半自動的に関連付け、コンサーン・アセスメントに活用可能な形式にする処理である。例えば、ツイート等のフィルタリング、地域分類、ニュース記事からのイベント抽出、および、ニュース記事と

ツイートの関連付け等が挙げられる。さらに、入力された意見を、住民の問題意識や懸案事項に割り当てる処理も意見構造化の一種である。

#### 3.1 イベントに基づく意見構造化システム

本研究では、Web 上のニュース記事をイベントとして構造化するシステムを開発した [7, 6]。本システムは、定期的にニュース記事を収集し、かつ、イベント抽出を行う。ニュース記事の収集間隔は 30 分である、収集では、日時やタイトル、本文などの情報を取得、および、形態素解析を行う。ニュース記事とイベントの関連度を計算し、得られた情報を LOD サーバに登録する処理を、毎日 1 回実行している。ここでは全イベントを更新するのではなく、LOD サーバ上にある過去 2 週間分のイベントを対象としている。

ニュース記事の集合から抽出されたイベントと、Twitter 上の意見を関連付けるスキーマを図 3 に示す。このとき、システムによる自動解析誤りの改良や、人手による修正を考慮したスキーマも必要となる。そのための、アノテーションに関するメタ情報のスキーマを、図 4 に示す。

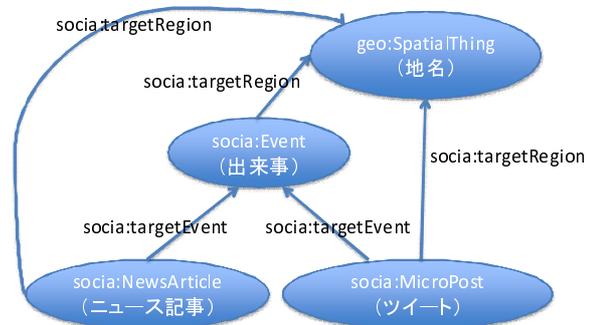


図3 地域関連情報の構造化のためのスキーマ

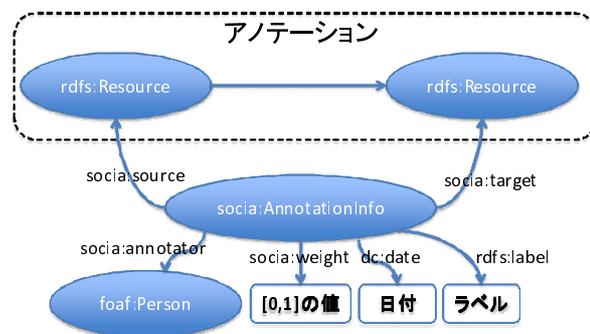


図4 アノテーションに関するメタ情報記述のためのスキーマ



図5 イベント情報の表示例

住民に提示される関連情報は、意見アーカイブから供給される。意見アーカイブには、自動的に構造化された意見として、後述するイベントに基づく意見構造化システムの出力が格納される。また、自動的に収集した情報だけでなく、手動により意見アーカイブ内の構造化された情報を、追加、更新、および、削除するための機能を実装した。

イベントとニュース記事の詳細を表示する際に、EventRelation クラスにより関連するニュース記事やイベントを表示するようにした。図5にイベント詳細の様子を示す。図中“1”の領域はイベントを表すエンティティを一覧表示している。エンティティそれぞれにその評価値が付与されており、評価値の大きなエンティティは大きく表示される。図中“2”の領域はイベントに関連付けられているニュース記事の一覧である。ニュース記事のタイトルを選択すると、各ニュース記事の詳細を表示する。図中“3”の領域はイベントとは直接関連付けられていないニュース記事の一覧を関連度の高い順に表示している。同様に、ニュース記事のタイトルを選択すると、各ニュース記事の詳細を表示する。

コンテンツの管理者、会議のモデレータ等は、関連するイベントやニュース記事を確認する必要がある。しかし、自動抽出によって関連付けられたニュース記事やマイクロブログでは、間違いや意図しない関

連付けが行われている場合がある。したがって、人手での修正機能が必要となる。簡易的な修正機能として、ニュース記事をイベントへの追加するボタンと削除するボタンを作成した。削除する場合は削除対象のニュース記事が明確であるが、追加する場合は、新たにニュース記事を探す必要がある。そこで、イベントへの関連付けは行われていないが、関連度の高いニュース記事を表示するようにして、ニュース記事の推薦を行なった。ユーザが、図5の“3”に示されたニュース記事から、イベントに関連するニュース記事として選択することで、意見アーカイブを洗練化することが可能になる。

### 3.2 意見内容に基づく地域分類

住民参画 Web プラットフォーム上では、住民-行政-専門者間で地域に関する問題意識を共有するために、Web 上のマイクロブログや新聞記事を活用する。このように地域コミュニティで活用するためには、Web コンテンツのうち地域に関連するものを抽出・分類しておき、その上で構造化・モデル化を施す必要がある。

本研究における地域分類器は、地方自治体への Web コンテンツ分類を行う。例えば日本全国の都道府県と市区町村を分類クラスとした場合、階層構造を無視すると約 2,000 クラスへの分類が必要となり、多くの訓練データが必要となる。そのための分類には、ナイーブベイズ分類器の改良である Transformed Weight-normalized Complementary Nave Bayes (TWCNB) [8] を用いる。地域分類では分類すべきクラス数が多いため、SVM のような 2 値分類器を one-vs-rest 法で適用するよりも、TWCNB のような多値分類器を用いた方が学習の手間を軽減できる。

TWCNB では、単純ナイーブベイズで計算したクラス確率  $p(c|d)$  を以下の式で補正したスコアを用い、これが最大となるクラス  $c$  に文書  $d$  を分類する。

$$\begin{aligned} \text{logscore}(d, c) = & \log p(c|d) \\ & - \sum_w \text{tf}(w, d) \log \left( \frac{1 + \sum_{k=1}^{|C|} \text{tf}(w, c_k)}{N + \sum_{k=1}^{|C|} \sum_{x=1}^N \text{tf}(x, c_k)} \right) \quad (1) \end{aligned}$$

ただし、 $C$  は地域を表すクラスの集合、 $\text{tf}(w, d)$  は文書  $d$  における語  $w$  の頻度、 $\text{tf}(w, c)$  はクラス  $c$  にお

る語  $w$  の頻度,  $N$  は異なり単語の総数である.

都道府県および市区町村というクラスへの分類のために, クラスの階層構造を考慮して TWCNB を適用する. これを, 以下では Semantically-Aware Hierarchical Balancing (SAHB) と呼ぶ. 自治体の地理的な階層構造は通常, 都道府県/市区町村/区のような形をとるが, SAHB では地理的な階層だけでなく, 訓練データの文書数も考慮して階層を設定する. すなわち, 訓練データが多いクラス (例: 東京都など) については, 地理的階層に関わらず独立した中間ノードとして扱う. これにより, 階層を考慮しなかった場合に比べて訓練データの偏りに頑健な学習が可能になる. さらに, 本手法は地域分類だけでなく様々な階層的トピックに基づく文書分類に適用でき, Web ページ推薦システムへの応用も可能である [9].

Web コンテンツの地域分類実験を行った. 訓練データとしては, Wikipedia および Twitter から収集したデータを用いた. Wikipedia 記事からの訓練データ作成にあたっては, Infobox に含まれる地域名, および, 本文の最初の文に含まれる地域名を利用した. これにより, 105,018 文書から成る訓練データを得た. 47 都道府県への分類実験では, 階層を考慮しない分類での適合率 70.3% に対し, SAHB で階層を考慮した提案手法では適合率 94.6% となった.

## 4 おわりに

本稿では, 住民参画のための基盤インフラとなる住民参画 Web プラットフォーム O<sub>2</sub> の開発について述べ, コンサーン・アセスメント支援のための意見構造化システムの概要を示した. また, 構造化された意見を蓄積・公開するための LOD データセット SOCIA を設計し, 自動構造化および人手による修正機能を実装した.

本システムは, ツイートを意見として収集・構造化する意見構造化システムをコアとしている. さらに, データセット SOCIA を活用したコンサーン・アセスメント支援機能を持つ議論システム `citispe@k`<sup>\*2</sup> を開発中である. 現在, これを用いて地域コミュニティを対象とした評価実験を計画中であり, 議論活性化の支援に関する評価を実施する予定である.

<sup>\*2</sup> <http://www.open-opinion.org/citispeak/>

謝辞 本研究の一部は, 総務省 戦略的情報通信研究開発推進制度 (SCOPE) の支援を受けたものです.

## 参考文献

- [1] E. Ferro and F. Molinari. Making sense of gov 2.0 strategies: “no citizens, no party”. *eJournal of eDemocracy and Open Government*, Vol. 2, No. 1, pp. 56–68, 2010.
- [2] A. Macintosh, T. F. Gordon, and A. Renton. Providing argument support for e-participation. *Journal of Information Technology & Politics*, Vol. 6, No. 1, pp. 43–59, 2009.
- [3] 名古屋市. 名古屋市中期戦略ビジョン, <http://www.city.nagoya.jp/somu/cmsfiles/contents/0000010/10920/saigiketsu.pdf>, 2010.
- [4] 経済産業省. オープンガバメントラボ, <http://openlabs.go.jp/>, 2010.
- [5] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011.
- [6] 平田紀史, Robin Swezey, 佐野博之, 白松俊, 大園忠親, 新谷虎松. 住民参画 web プラットフォームのためのニュース記事と意見の構造化. 人工知能学会研究会資料 SIG-SWO-A1101-02, pp. 1–6, 2011.
- [7] 平田紀史, 白松俊, 大園忠親, 新谷虎松. ユーザの観点に基づくイベント系列化を用いた web ニュース記事閲覧支援システムの実装. 人工知能学会論文誌, Vol. 26, No. 1, pp. 228–236, 2011.
- [8] J.D.M. Rennie, L. Shih, J. Teevan, and D.R. Karger. Tackling the poor assumptions of naive bayes text classifiers. In *Proceedings of the 20th International Conference on Machine Learning*, pp. 616–623, 2003.
- [9] R. Swezey, S. Shiramatsu, T. Ozono, and T. Shintani. Intelligent page recommender agents: Real-time content delivery for articles and pages related to similar topics. In *Modern Approaches in Applied Intelligence, Proceedings of IEA/AIE 2011, Part II*, pp. 173–182, 2011.