

# 電力自由化を想定した プライバシー保護データシェア手法の提案

今西 智哉<sup>†</sup> 岡田 健吾<sup>†</sup> 西 宏章<sup>†</sup>  
慶應義塾大学 理工学研究科<sup>†</sup>

## 1. はじめに

通信機能が備わった電力メータである, スマートメータ導入に伴い, 家庭の電力利用データの有効利用が考えられている. 日本では 2016 年度より電力小売り部門の自由化が始まり, 各家庭は電力を購入する電力会社を選ぶことができるようになる. そのため同じ地域の家庭でも異なる電力会社と契約する場合が予測され, 電力利用データの点在化が進む. 点在化したデータを利用する場合, 企業間でのデータシェアする必要があるが, 電力利用データはプライバシー情報と考えられており, 匿名化処理が必要である. 本研究ではプライバシーを保護しつつ, 各企業が保持する電力利用データをシェアする方法を提案する.

## 2. 前提と提案手法

本研究は, 家庭の電力データが各家庭から契約している各企業に送られている状況を想定している. 各企業は, 需要のピークカットを目的としたデマンドレスポンスなどへ, 自社のデータだけでなく他社が保持するデータを利用したいというニーズがある. しかし, 家庭の電力データは個人情報の一つであると考えられ, 企業は保持するデータを家庭の許可なく外部に実データを渡すことが出来ない. そのためデータを適切に共有する手法が必要となる.

データを共有する手法は, 匿名化, 暗号化などが考えられる. それぞれの企業が独自に匿名化 ( $k$ -匿名化[1]) を行うと, データ数が少ない場合にデータの損失が大きくなる. また, データを暗号化したまま計算処理を行う準同型暗号を適応した手法が提案されているが, 複雑な計算を行うと計算コストが高くなる. データ共有に関する問題点があるが, 本研究では, 教師なし学習の SOM (Self-Organizing Map) [2] を用いることで点在化するデータの傾向を考慮して代表パターンを作成し, パターンデータをデマンドレスポンスへの適応を考える.

扱う電力データは, 1日1軒分30分間隔で送信

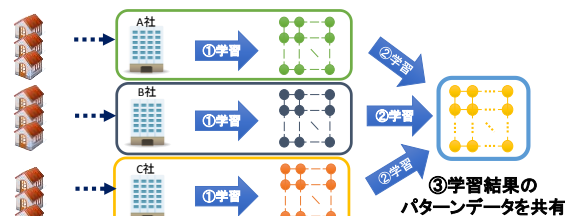


図 1 学習ステップ

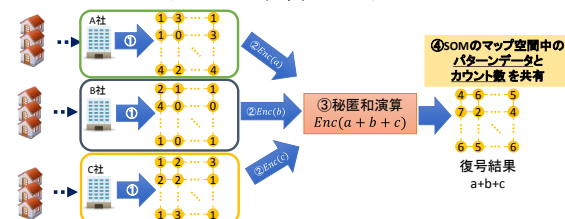


図 2 カウント情報の集計

されるスマートメータの値を1つのデータとし,  $i$ 番の家庭の電力データ  $x_i$ は,  $N_{home}$ を全体の家庭総数とすると, 以下のように表される.

$$x_i = (x_{i,00:00}, x_{i,00:30}, \dots, x_{i,23:30}) \quad i = 1, 2, \dots, N_{home}$$

本提案手法はデータの学習によるパターンデータ生成と, カウントの2つのステップに提案手法が分かれている.

### 2.1 パターンデータ生成 (図 1)

各企業が保持するデータのパターン化をそれぞれ独立に行う. 生成された独立したパターンを新たな入力データとし, それを用いて全体のデータを考慮した統合パターンデータを生成する. プライバシーを保護しつつ電力データをシェアすることが, 本統合パターンデータにより可能となる. ここで用いるパターンデータ化手法は SOM を用いる. (1)-(4)にてアルゴリズムについて簡潔に記す.

(1) マップ空間中に初期ノードを設定

$N_{map}$ を SOM のマップサイズとすると, SOM のマップ空間中のノードベクトル  $w_i$ は下記のように表され, 乱数で設定する.

$$w_i = (w_{i,00:00}, w_{i,00:30}, \dots, w_{i,23:30}) \quad i = 1, 2, \dots, N_{map}$$

(2) 入力データに最も類似度が高いノードの選択

類似度の計算はユークリッド距離に基づいて行う. 選ばれたノードを BMU (Best Match Unit)

とする.

$$BMU = \operatorname{argmin}_i \|x - w_i(s)\| \quad i = 0, 1, \dots, N_{map}$$

(3) マップ空間中のノードの持つ値の更新

BMU の周辺ノードを入力ベクトルに近づける.

$$w_i(s+1) = w_i(s) + h_{BMU,i}(s)(x - w_i(s))$$

ここで  $h_{BMU,i}$  は学習関数であり, マップ空間中のノードと BMU の距離と学習回数に依存する.

(4) (2), (3) を全入力データで行い, ノードの持つ値に変化が見られなくなるまで繰り返す.

## 2.2 データカウント (図 2)

生成した統合パターンデータは, 需要電力のピーク時間を勘案していない. そこで, 各企業は保持するデータとパターンデータの類似度を計算し, それぞれのパターンデータに近い家庭が何軒存在するかをカウントする. その後カウント情報を全体で足し合わせ, 全体としてどのパターンの家庭がそれぞれ何軒存在するかという情報を生成する.

デマンドレスポンスは地域の需用電力がピークとなる時間で行われる. 提案手法では BMU 選択の際にピーク時間の類似度を重視するための重みづけを行う. 重みづけ  $a_t$  は正規分布に従うものとする.

$$BMU = \operatorname{argmin} \left( \sqrt{\sum_{t=1}^T a_t (x_t - w_{i,t})^2} \right)$$

$$a(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-p)^2}{2\sigma^2}}$$

翌日のピーク時刻を  $p$ , 正規分布の分散値  $\sigma^2$  を調節パラメータ,  $T$  は一日のデータサンプル数 (48) である. ピーク時刻予測は時系列解析手法の季節自己回帰和分平均モデル SARIMA を用いて行った. 入力を 7 日 ( $7 \times 48$  sample) として 1 日先 (48 sample) の予測を行うものとした.

それぞれのカウントの足し合わせを秘匿を行うために, 暗号化したまま加法演算を行うことが可能である加法準同型暗号を用いる. 公開鍵で暗号処理を各企業で行い, 暗号化結果を  $\text{Enc}(a)$ ,  $\text{Enc}(b)$ ,  $\text{Enc}(c)$  として送信し, カウント数を暗号化したまま加法演算  $\text{Enc}(a+b+c)$  を行う. 最後に秘密鍵を用いて復号することによって全体のカウント合計値を得ることが可能となる.

## 3 評価

本研究に用いるデータは commission for regulation の Irish smart meter dataset [3] の 1000 軒のスマートメータ (30 分値) を用いた.

$k$ -匿名性を満たすためのクラスタリングとして Greedy K-Member Clustering [4] を用いた. クラスタリング処理後, そのクラスタの代表値

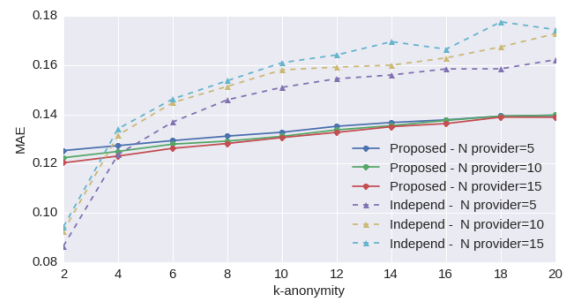


図 3  $k$  と企業数の関係

として, クラスタに含まれるデータの平均値を採用する. 情報損失度はその代表値とクラスタに含まれるデータの絶対誤差として評価した. 提案手法を用いてデータ共有後に匿名化処理を行った場合と (Proposed), それぞれの企業が独立的に匿名化処理を行った場合 (Independ) を比較する. それぞれの企業が保持するデータ数は (全データ数/企業数) として分配した.

Irish データセットを用いた場合の結果を図 3 に示す. 匿名化のレベル  $k$  が大きく, また企業数 ( $N_{provider}$ ) が増加するほど提案手法が優位であることを示している. 特に  $k=20$ , 企業数 = 15 のときはそれぞれ独立的に行った場合と比べ  $k$ -匿名化を行う際の情報損失を 20% 削減することが出来ている. しかし,  $k$  が小さい ( $k=2,3$ ) 場合においては, 独立的に匿名化を行った場合の方が良い結果を示した. 提案手法は,  $k$ -匿名性が高いデータに特に有効であると言える.

## 4 結論

自己組織化マップを用いて家庭の使用電力データシェア手法を提案した. 互いに実データを公開することなくパターンデータを生成し, 各企業がそれぞれ独立に匿名化処理を行った場合と比べ, 少ない情報損失で  $k$ -匿名化を実現した. 謝辞

本研究は, セコム科学技術振興財団研究助成, 科研費基盤 B(24360230)(25280033), 国交省住宅・建築物技術高度化の一環としてなされた.

## 参考文献

- [1] Sweeney, L.:  $k$ -anonymity: A model for protecting privacy, international Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, pp.557-570 (2002)
- [2] Kohonen, T.: The self-organizing map, Proc. IEEE, vol. 78, no. 9, pp. 1464-1480 (1990)
- [3] <http://www.ucd.ie/issda/data/commissionforenergyregulation/>
- [4] Byun, J.W., et al.: Efficient  $k$ -anonymization using clustering techniques, Advances in Databases: Concepts, Systems and Applications. Springer Berlin Heidelberg, pp. 188-200 (2007)