

平成 30 年 6 月 14 日現在

機関番号：12301

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00591

研究課題名(和文)リアルタイム水中病原体濃度レベル予測システムの開発

研究課題名(英文)Development of system for real-time prediction of pathogen level

研究代表者

加藤 毅 (Kato, Tsuyoshi)

群馬大学・大学院理工学府・准教授

研究者番号：40401236

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：水利用における微生物学的安全性を担保するには実時間病原性微生物予測法を実現した。水中における病原体濃度は下水を培養ベースの方法でモニタリングすることで観測されてきており、水由来の病気を防ぐのに有効だと知られている。一方、実時間モニタリングはいまだに実現していない本プロジェクトでは実時間で測定できる水質水門データから病原体を予測する方法論を確立した。その方法論は、水質工学における性質を利用して小標本からでも高精度に学習を可能にするものである。開発した方法論をオンラインソフトウェアとして実装した。

研究成果の概要(英文)：To ensure the microbial safety in water usage, it is meaningful to prediction pathogens on a real-time basis. The concentration of pathogens in water, which is measured by culture-dependent methods, has been used to monitor the fecal contamination in water environment, and has been proved to be effective to prevent waterborne infectious diseases in varied water usage styles. On the other hand, the real-time monitoring has not yet been achieved for most of pathogens. In this project, we exploit the fact that it is possible to measure some of the hydrological and water quality data with real-time sensors, for the purpose of the real-time prediction of pathogens by using these hydrological and water quality data as explanatory variables. The methodology was devised by combining the state-of-the-art mathematical theories with the domain knowledge in the water quality engineering. Online software was developed so that any people can use the methodology.

研究分野：機械学習

キーワード：汚染調査と評価

1. 研究開始当初の背景

水利用における微生物学的安全性は、大腸菌等に代表される「指標微生物」の水中濃度を監視・制御することにより担保されてきた。この「指標微生物」を用いた各種水質基準の運用により、様々な病原菌に起因する水系感染症の発生が抑制されていることは疑いようのない事実である。かつて江戸末期から明治初期に年間10万人超の死者を度々出したコレラ等による水系感染症の発生件数が、当時とは比較にならないほど激減していることはその証左と言えよう。その一方で、水域における微生物学的水質異常のリアルタイムモニタリングは、未だ実現していない重要課題である。そもそも「指標微生物」によって微生物汚染を「リアルタイムに監視」することは想定されていない。仮に近年急速に発展している分子生物学的な手法を水中病原体のモニタリング手法として取り入れたとしても、水中病原体の存在を検知するのは早くも数時間後であり、リアルタイムモニタリングを実現する際には明白な技術的障壁が存在する。しかも浄水場等において分子生物学的手法を用いて継続的に水中病原体濃度をモニタリングするには多大な経費とマンパワーが必要であり、昨今の税収減に苦しむ行政の立場からすれば、水処理の現場において分子生物学的手法による水中病原体のリアルタイムモニタリングシステムを導入することは、将来的な技術革新に期待をかけたとしても、実行に移す段階には残念ながら達していないのが現状である。

2. 研究の目的

以上のような背景のもと、本研究では、水環境中に存在する病原体に関し、水道原水域等における病原体の水中濃度レベルをリアルタイムで予測するシステムを開発することを目的とする。国土交通省による「水文水質データ（流量、懸濁質量等）」、気象庁による「気象データ」及び最先端の「機械学習」を組み合わせることで、高精度な「リアルタイム水中病原体濃度レベル予測システム」を構築することを目指す。

3. 研究の方法

測定した病原体濃度とそのときの水門水質データとの関係から線形識別器のパラメータを学習する。説明変数には水門水質データにはWT, pH, EC, SS, DO, BOD, TN, TP, FRが含まれる。それぞれの説明変数は、大腸菌濃度とは次の関係があることが知られている。

- WT : 水温。水温が高いと大腸菌濃度が増加する可能性がある。
- pH : 水素イオン指数。水素イオン指数が中性付近 (7.0-7.5) より高いもし

Algorithm 1 Learning Algorithm for SC-SVM.

Require: n examples $(x_i, y_i) \in \mathbb{R}^d \times \{\pm 1\}$ for $i = 1, \dots, n$, regularization parameter $\lambda \in \mathbb{R}$, and sign constraint parameter $c \in \{\pm 1, 0\}^d$.

```

1: begin
2:  $\alpha^{(0)} := \mathbf{0}_n; \bar{w}^{(0)} := \mathbf{0}_d; w^{(0)} := \mathbf{0}_d;$ 
3: for  $t := 1, \dots, T$  do
4:   Select  $i \in \{1, \dots, n\}$  at random;
5:

```

$$\Delta\alpha := \max \left\{ -\alpha_i^{(t-1)}, \min \left\{ 1 - \alpha_i^{(t-1)}, \frac{1 - y_i \langle w^{(t-1)}, x_i \rangle - \gamma \alpha_i^{(t-1)}}{\|x_i\|^2 / (\lambda n) + \gamma} \right\} \right\};$$

```

6:    $\alpha^{(t)} := \alpha^{(t-1)} + \Delta\alpha e_i;$ 
7:    $\bar{w}^{(t)} := \bar{w}^{(t-1)} + \frac{y_i \Delta\alpha}{\lambda n} x_i;$ 
8:    $w^{(t)} := \bar{w}^{(t)} + c \odot (-c \odot \bar{w}^{(t)});$ 
9: return  $\bar{w} := \frac{1}{T-T_0} \sum_{t=T_0+1}^T w^{(t-1)};$ 
10: end.

```

くは低いと大腸菌の生育速度が下がり、異常に高いもしくは低い場合には死滅する可能性がある。

- EC : 電気伝導率。水中に含まれるイオン濃度の指標で、電気の伝わりやすさを表す。ECは、下水処理場や浄化槽からの放流水のほうが河川水より高いので、ECが高いと、大腸菌と放出源となる下水処理場や浄化槽が近くに存在することを意味する。
- DO : 溶存酸素。水中に含まれる酸素量が高いと大腸菌の増殖に有利である。
- SS : 浮遊物質・懸濁物質。粒径が 1mm ~ 2mm の粒子状物質の含有量を表す。この値が高いと、大腸菌が紫外線等からの不活化ストレスから防御され、生存に有利となる可能性がある。
- BOD : 生物化学的酸素要求量。好気性微生物によって有機物が分解されるときに消費される酸素の量を表す。この値が高い場合、大腸菌の濃度も高い可能性がある。
- COD 酸性法 : 酸性高温過マンガン酸法の測定値。この値が高い場合、大腸菌の濃度も高い可能性がある。
- COD アルカリ法 : アルカリ性高温過マンガン酸法。この値が高い場合、大腸菌の濃度も高い可能性がある。
- T-N : 全窒素。窒素量が高いと大腸菌の生育に有利となる可能性がある。
- T-P : 全リン。リンが多いと大腸菌の生育に有利となる可能性がある。
- FR : 流量。流量が多いと大腸菌が希釈され、濃度が低下する可能性がある。

このように、水質工学において水門水質データと病原体濃度との関係はすでに分かっている。にもかかわらず、これまでの機械学習法は、この情報を利用せずに盲目的に機械学習のツールにデータを入力して解析が試みられてきた。本研究では、上記の知識を制約式の形で簡潔に表す数式表現を考案して、既存の機械学習法に組み込んだ。なお、pHは、

pH_+ , $\max(0, pH - 7)$ および pH_- , $\max(0, 7 - pH)$ のようにして得られた算法が Algorithm 1 である。Algorithm 1 において $d=10$ であり、ベクトル \mathbf{c} のうち WT, EC, SS, BOD, TN, TP に対応する要素は $c_n=+1$ pH_+ , pH_- ,

Algorithm 2 Learning Algorithm for SC-LR.

Require: Data matrix $\mathbf{X} := [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, class labels $\mathbf{y} := [y_1, \dots, y_n]^T \in \{\pm 1\}^n$, regularization parameter $\lambda \in \mathbb{R}$, and sign constraint parameter $\mathbf{c} \in \{\pm 1, 0\}^d$.

```

1: begin
2:  $\alpha^{(0)} := \mathbf{1}_n/2$ ;
3:  $\tilde{\mathbf{w}}^{(0)} := \mathbf{X}\mathbf{y}/(2\lambda n)$ ;  $\mathbf{w}^{(0)} := \tilde{\mathbf{w}}^{(0)} + \mathbf{c} \odot (-\mathbf{c} \odot \tilde{\mathbf{w}}^{(0)})_+$ ;
4: for  $t := 1, \dots, T$  do
5:   Select  $i \in \{1, \dots, n\}$  at random;
6:    $z^{(t)} := y_i \langle \mathbf{w}^{(t-1)}, \mathbf{x}_i \rangle$ ;
7:    $F_t := -\log \sigma(z^{(t)}) - H_{\text{bern}}(\alpha_i^{(t-1)}) + \alpha_i^{(t-1)} z^{(t)}$ ;
8:    $q_t := \sigma(-z^{(t)}) - \alpha_i^{(t-1)}$ ;
9:    $I_{q_t} := [-(q_t)_+, (q_t)_+]$ ;
10:   $\tilde{s}_i := 4\lambda n / (4\lambda n + \|\mathbf{x}_i\|^2)$ ;
11:   $\Delta\alpha := \text{Clip}_{I_{q_t}}((2q_t^2 + F_t)\tilde{s}_i / (4q_t))$ ;
12:   $\alpha^{(t)} := \alpha^{(t-1)} + \Delta\alpha e_i$ ;
13:   $\tilde{\mathbf{w}}^{(t)} := \tilde{\mathbf{w}}^{(t-1)} + \frac{\Delta\alpha}{\lambda n} \mathbf{x}_i$ ;
14:   $\mathbf{w}^{(t)} := \tilde{\mathbf{w}}^{(t)} + \mathbf{c} \odot (-\mathbf{c} \odot \tilde{\mathbf{w}}^{(t)})_+$ ;
15: return  $\tilde{\mathbf{w}} := \frac{1}{T-T_0} \sum_{t=T_0+1}^T \mathbf{w}^{(t-1)}$ ;
16: end.
```

DO, FR に対応する要素は $c_n=-1$ として用いる。Algorithm 1 はサポートベクトルマシンという機械学習の方法に事前知識から得られる制約を組み込んだものであるが、ロジスティック回帰に組み込んだ学習方法も開発した (Algorithm 2)。

4. 研究成果

本研究で開発した方法の性能を評価するため、177回測定した大腸菌のデータと対応する水文水質データを用いた。そのうち、無作為に選んだ n 個を線形識別器の訓練に用い、残りの $(177-n)$ 個を評価に用いた。評価指標には PRBEP と ROC スコアを用いた。線形識別器の学習法として、従来のサポートベクトルマシン (SVM) およびロジスティック回帰 (LR) に加えて、それぞれに事前知識を組み込んだ SC-SVM と SC-LR を比較した。図 1 は $n=10$ のときの ROC カーブを示している。サポートベクトルマシンでもロジスティック回帰でも事前知識の導入によって性能が大きく向上していることがわかる。図 2 および図 3 が n の値を変化させたときの PRBEP と ROC スコアをプロットしたものである。特に、学習用データが少ないとき、顕著に性能が向上することがわかる。次に、気象データを組み込んだ場合も比較実験を試みた。気象データは水文水質データと比べて病原体との相関が強くないため、訓練用データが少ないときは汎化性能が低下してしまうという現象が見られた。

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5件)

- (1) Toshihiro Ito, Tsuyoshi Kato, Kenta Takagishi, Satoshi Okabe, and Daisuke Sano, Bayesian modeling of virus removal efficiency in wastewater treatment processes, Water Science and Technology, Nov 2015;72(10):1789-95. doi: 10.2166/wst.2015.402. 査読有。
- (2) Tsuyoshi Kato, Ayano Kobayashi, Toshihiro Ito, Takayuki Miura, Satoshi Ishii, Satoshi Okabe, Daisuke Sano, Estimation of concentration ratio of indicator to pathogen-related gene in environmental water based on left-censored data, Journal of Water and Health, vol. 14, no. 1, pp.14--25, Feb 2016, doi:10.2166/wh.2015.029. 査読有。
- (3) Toshihiro Ito, Tsuyoshi Kato, Makoto Hasegawa, Hiroyuki Katayama, Satoshi Ishii, Satoshi Okabe, Daisuke Sano: Evaluation of virus reduction efficiency in wastewater treatment unit processes as a credit value in the multiple-barrier system for wastewater reclamation and reuse, Journal of Water and Health, , vol 14, no. 5, pp.879--889, Oct 2016, doi: 10.2166/wh.2016.096. 査読有。
- (4) Toshihiro Ito, Masaaki Kitajima, Tsuyoshi Kato, Satoshi Ishii, Takahiro Segawa, Satoshi Okabe, Daisuke Sano, Target virus log10 reduction values determined for two reclaimed wastewater irrigation scenarios in Japan based on tolerable annual disease burden. Water Research. Nov 15;125:438-448, 2017. doi: 10.1016/j.watres.2017.08.05. 査読有。
- (5) Tsuyoshi Kato, Misato Kobayashi, Daisuke Sano, Sign-Constrained Regularized Loss Minimization. 査読無, Computing Research Repository, <https://arxiv.org/abs/1710.04380>, 2017.

(学会発表)(計5件)

- (1) 下山愛祐美, 黒岩祐有美, 佐野大輔, 加藤毅, 河川水中大腸菌数予測のための符号制約回帰分析, 情報処理学会第 80 回全国大会, 第 1 分冊, 293--294, 2018 年 3

月 13-15 日，早稲田大学．情報処理学会
第 80 回全国大会学生奨励賞受賞．

- (2) 加藤毅，小林美里，佐野大輔，符号制限 SVM と水文水質データを利用した大腸菌数予測への応用，第 20 回日本水環境学会シンポジウム，pp.121--122，2017 年 9 月 27 日，和歌山．
- (3) 小林美里，佐野大輔，加藤毅，水文水質データを利用した大腸菌予測のための符号制限学習，第 16 回情報科学技術フォーラム FIT2017，第 1 分冊，103--104，2017 年 9 月 13 日，東京大学本郷キャンパス．FIT 奨励賞受賞．
- (4) 小林 美里，宮村 明帆，佐野 大輔，加藤毅，符号制限線形識別器の開発と河川水中大腸菌数予測への応用，第 15 回情報科学技術フォーラム FIT2016，第 1 分冊，149--150，2016 年 9 月 8 日，富山大学．FIT 奨励賞受賞．
- (5) 小林美里，佐野大輔，加藤毅，SVM によるリアルタイム河川水中大腸菌濃度予測，情報処理学会第 78 回全国大会，第 1 分冊，453--454，2016 年 3 月 10-13 日，慶応技術大学矢上キャンパス．情報処理学会第 77 回全国大会学生奨励賞受賞．

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

ホームページ等

<https://ts3.pl.cs.gunma-u.ac.jp/~kato/software-h30/h27kc-scsvm/h27kc-scsvm.html>

6. 研究組織

(1) 研究代表者

加藤 毅 (KATO, Tsuyoshi)
群馬大学・大学院理工学府・准教授
研究者番号：4 0 4 0 1 2 3 6

(2) 研究分担者

佐野 大輔 (SANO, Daisuke)
東北大学・大学院工学研究科・准教授
研究者番号：8 0 5 5 0 3 6 8

(3) 連携研究者

なし

(4) 研究協力者

なし

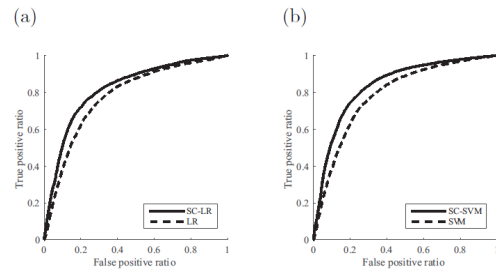


図 1 ROC カーブ

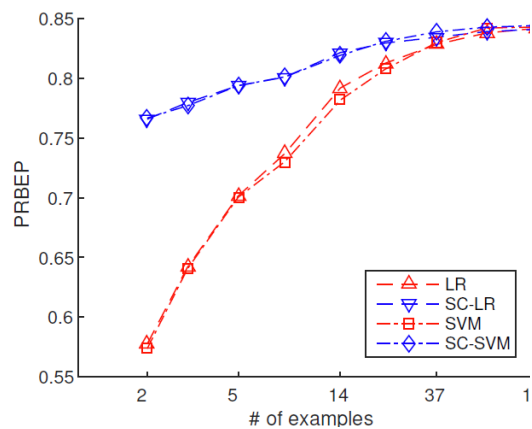


図 2 PRBEP による性能評価

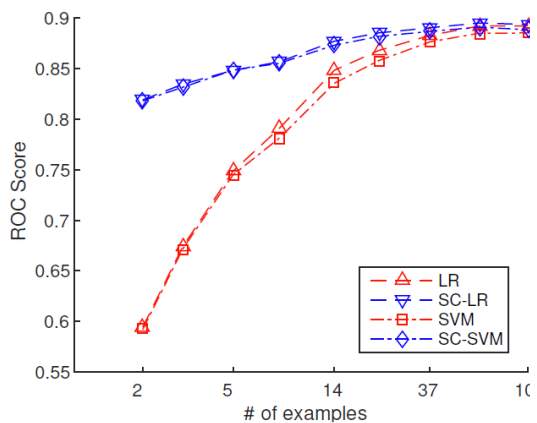


図 3 ROC スコアによる性能評価