



# データストリームの世界に挑んで

都築学<sup>1</sup> 小西修<sup>2</sup>

<sup>1</sup> 野村総合研究所 <sup>2</sup> 公立はこだて未来大学システム情報科学部

〔受賞論文〕

大規模データストリームのための履歴情報を用いたカーネル法の拡張

都築学, 小西修(公立はこだて未来大学)

情報処理学会論文誌: データベース, Vol.1, No.3 (TOD), pp.49-59 (2008)

本学会論文賞という名誉ある賞をいただき、大変光栄に思っております。また、改めてご指導、ご協力いただいた皆様に感謝申し上げます。本論文の提案に先立ち、我々は2006年の前研究において、実際の大规模クレジットカードデータを正常利用と不正利用に識別する実験を行いました。その結果、クレジットカードデータのような大規模データストリームが持つ「過去の履歴情報とその振る舞い」が、データを特徴づける重要な要素であることを示しました。思い返せば、この知見は本論文の構想の大きな礎になっていると感じます。その後、木構造や配列、グラフなどの構造データの特徴を分析に導入できるよう、カーネル法を拡張した研究が多く発表されていることを受け、「大規模データストリームが持つ特徴を、カーネル法で分析に導入できないか?」と考えました。この大規模データストリームの特徴とは、前述した「過去の履歴情報とその振る舞い」を指します。我々はこの構想に基づき、検討を進め、このたび本論文にあるストリームカーネルを提案することができました。提案手法は、履歴情報にある新しいデータであるほど分析結果に与える影響が大きく、古いデータであるほど影響が小さくなるようにデータを扱うことで、データストリームの時間的変化を分析の特徴に用いることができました。前述と同様の大規模な実データを用いた識別実験では、従来手法を上回る識別精度を得ることができました。一方、この実験においては苦勞の連続であったと感じています。本論文では大規模な実データを実験に使用したため、(a) データクレンジング、(b) データサンプリング、

さらに(c) データストリームの過去の履歴情報からの特徴抽出等、事前作業に多量の時間と労力を費やしました。これら前処理の計算時間の短縮や効率的処理法については本論文では扱えなかったのですが、大規模データストリームに適したデータベースモデルの提案などとも関連して今後の重要な研究課題であると思われま

す。また、本論文で提唱したストリームカーネル法では識別精度において大きな成果を得ましたが、多くの計算時間を要したことが課題でした。小西らは、その後これらの問題をパラレルブースティングにより解決し、学習時間・検証時間を大幅に削減する成果を得ています<sup>☆1</sup>。

データストリームというダイナミクスを扱う本研究分野はいまだ発展を続け、先は長く末広がりです。今後の本分野のさらなる発展を願うとともに、本研究がその発展に寄与できれば幸いです。

(平成22年4月30日受付)

都築学 (正会員) manabu.t124@gmail.com

1985年生。2007年公立はこだて未来大学システム情報科学部複雑系科学科卒業。2009年同大学院システム情報科学研究科システム情報科学専攻博士前期課程修了。現在、野村総合研究所勤務。

小西修 (正会員) okonishi@fun.ac.jp

公立はこだて未来大学システム情報科学部教授。京都大学工学博士。京都大学、名古屋大学、文部省核融合科学研究所、高知大学を経て、2002年より現職。2008年より公立大学法人公立はこだて未来大学理事・副学長。

<sup>☆1</sup> 小西他: ストリームカーネルマシンによるパラレルブースティング, 情報処理学会論文誌: データベース, Vol.2, No.4, pp.13-23 (Dec. 2009).