

Deep learning from big data on cancer

Kenji Satou

*Faculty of Electrical and Computer Engineering
Institute of Science and Engineering
Kanazawa University
Kakuma-machi, Kanazawa 920-1192, Japan*



Today it is well known that cancer is fundamentally a genomic disease. However, it is also known that cancer is quite a complicated phenomenon related to various molecular, cellular, and tissue factors. To unveil all the mechanisms of cancer, huge amounts and wide varieties of data are now available. Next-generation sequencing (NGS) can generate sequence and expression data in high throughput and low cost. Also, microarrays can yield large amounts of expression data. In addition, data on more than millions of chemical molecules are available. Using such data, so many computational studies on cancer have been conducted or are undergoing: statistical analysis, network simulations in systems biology, molecular simulations for drug discovery, machine learning for analysis and prediction, etc. However, still we need deeper analysis on cancer to understand and control it.

In the field of machine learning, deep learning is one of the hottest topics in recent years. Though it is a technology developed from (artificial) neural networks, it has achieved remarkable success in many application domains including speech recognition, image recognition, prediction of chemical molecule compounds, etc. It is said that deep learning can be a breakthrough of once-paradecades, and its application field is rapidly extending. In this talk, I will introduce some basics of deep learning and published studies about deep learning application to biomedical data. After presenting some preliminary studies in my lab, I'd like to discuss about the applicability of deep learning technology to big data on cancer.

Kenji Satou

Professor, Institute of Science and Engineering, Kanazawa University, Japan
E-mail: ken@t.kanazawa-u.ac.jp

EDUCATIONS/TRAINING

1987 B.E. Computer Science and Communication Engineering, Kyushu University
1989 M.E. Computer Science and Communication Engineering, Kyushu University
1996 D.E. Computer Science and Communication Engineering, Kyushu University

POSITIONS AND HONORS

1989-1994 Research Associate, Kyushu University
1995-1998 Research Associate, the University of Tokyo
1998-2007 Associate Professor, Japan Advanced Institute of Science and Technology
2007-2011 Associate Professor, Kanazawa University
2011-Present Professor, Kanazawa University

RECENT PUBLICATIONS

1. Nguyen TLA, Hirose O, Dang XT, Le TTK, Saethang T, Tran VA, Kubo M, Yamada Y and Satou K. Improving the Prediction of Protein-Protein Interaction Sites Using a Novel Over-Sampling Approach and Predicted Shape Strings. *Annual Review & Research in Biology* 3: 92-106, 2013.
2. Le TD, Liu L, Liu B, Tsykin A, Goodall GJ, Satou K and Li J. Inferring microRNA and transcription factor regulatory networks in heterogeneous data. *BMC Bioinformatics* 14: 92-104 2013.
3. Dang XT, Hirose O, Saethang T, Tran VA, Nguyen TLA, Le TTK, Kubo M, Yamada Y and Satou K. A novel over-sampling method and its application to miRNA prediction. *Journal of Biomedical Science and Engineering* 6: 236-248, 2013.
4. Dang XT, Hirose O, Bui DH, Saethang T, Tran VA, Nguyen TLA, Le TTK, Kubo M, Yamada Y and Satou K. A Novel Over-Sampling Method and its Application to Cancer Classification from Gene Expression Data. *Chem-Bio Informatics Journal*, 13: 19-29, 2013.
5. Saethang T, Hirose O, Kimkong I, Tran VA, Dang XT, Nguyen TLA, Le TTK, Kubo M, Yamada Y and Satou K. PAAQD: Predicting immunogenicity of MHC class I binding peptides using amino acid pairwise contact potentials and quantum topological molecular similarity descriptors. *Journal of Immunological Methods* 387: 293-302, 2013.
6. Yamada Y, Sawada H, Hirotsu K, Oshima M and Satou K. Validation of MIMGO: a method to identify differentially expressed GO terms in a microarray dataset. *BMC Research Notes* 5: 680-691, 2012.
7. Saethang T., Hirose O., Kimkong,I., Tran,V.A., Dang,X.T., Nguyen,T.L.A., Le,T.T.K., Kubo,M., Yamada,Y., Satou,K., EpicCapo: epitope prediction using combined information of amino acid pairwise contact potentials and HLA-peptide contact site information. *BMC Bioinformatics* 13: 313-324, 2012.
8. Tran VA, Clemente JC, Nguyen DT, Li J, Dang XT, Le TTK, Nguyen TLA, Saethang T, Kubo M, Yamada Y and Satou K. IMPACT: A Novel Clustering Algorithm based on Attraction. *Journal of Computers* 7: 653-665, 2012.
9. Yoshida M, Ishimura A, Terashima M, Enkhbaatar Z, Nozaki N, Satou K and Suzuki T. PLU1 histone demethylase decreases the expression of KAT5 and enhances the invasive activity of the cells. *Biochemical Journal* 437: 555-564, 2011.
10. Tran DH, Ho TB, Pham TH and Satou K. MicroRNA expression profiles for classification and analysis of tumor samples. *IEICE Transactions on Information and Systems* E94-D: 416-422, 2011.