

PENERAPAN PCA PADA KLASIFIKASI DATA MIKRO ARAY GEN MENGGUNAKAN JARINGAN SYARAF TIRUAN

Prisca Deviani Pakan

Departemen Mikrobiologi, Universitas Nusa Cendana Kupang

Jl. Adisucipto – Penfui Kupang NTT

E-mail: priscapakan@yahoo.com

Abstrak

Penelitian ini dilakukan dengan melakukan analisa terhadap data micorarray, selanjutnya ekspresi dari ribuan gen yang merepresentasikan suatu jaringan pada manusia, akan diklasifikasikan sebagai jaringan kanker atau bukan setelah dilakukan proses reduksi dimensi menggunakan algoritma PCA. Tahapan Klasifikasi dilakukan menggunakan algoritma jaringan syaraf tiruan

Kata kunci: mikro aray, kanker, ekspresi gen, jaringan syaraf tiruan, pca.

PENDAHULUAN

Teknologi DNA microarray digunakan untuk menentukan tingkat ekspresi ribuan gen yang dilakukan dalam sekali percobaan, dan secara simultan memantau proses biologi yang sedang berlangsung [6]. Teknologi ini mampu menghasilkan data yang dibutuhkan untuk proses prediksi dan klasifikasi gen yang diambil dari beberapa jaringan tertentu pada manusia untuk digolongkan ke dalam kanker atau bukan.

Berdasarkan [3], yang menjadi kendala utama dalam suatu data microarray adalah besarnya dimensi, yang menyebabkan beban komputasi menjadi tidak stabil. Untuk menangani masalah itu, Genetics Algorithm (GA) digunakan sebagai seleksi fitur untuk mereduksi dimensinya. GA akan membangkitkan populasi awal secara random, kemudian fungsi fitness digunakan untuk mengevaluasi populasi tersebut, dan selanjutnya dilakukan operasi genetika seperti seleksi orang tua, *crossover*, dan mutasi hingga kondisi kriteria dicapai. Kemudian data yang dimensinya sudah direduksi akan diklasifikasi menggunakan metode Jaringan Syaraf Tiruan (JST)

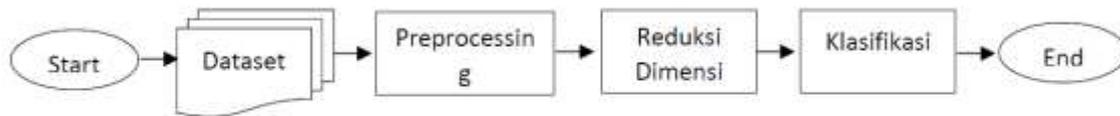
Aplikasi microarray banyak digunakan dalam deteksi kanker, dimana sel kanker mengalami abnormalitas dalam meng-ekspresikan gennya. Pada dasarnya, DNA microarray akan mengamati ekspresi genetik dari berbagai gen dengan jumlah yang mencapai ratusan bahkan ribuan secara bersamaan. Dalam analisis kanker, digunakan DNA normal dan DNA kanker yang kemudian akan di hibridisasi lalu akan dibaca dengan menggunakan laser (scanner). Klasifikasi data microarray merupakan salah satu kajian dalam

bidang bioinformatika yang digunakan, salah satunya yaitu untuk analisis suatu kanker. Klasifikasi pada deteksi kanker, dapat dilakukan dengan mengelompokkan data ke dalam kelas yang sudah ditetapkan.

Micorarray mampu menentukan ekspresi ribuan gen dan secara simultan memantau proses biologis yang sedang berlangsung [3]. Microarray merupakan bagian dari high dimensional data karena memiliki ratusan sampai dengan ribuan fitur [4]. Namun permasalahan yang terjadi adalah data microarray mempunyai jumlah variabel yang lebih besar dibandingkan dengan jumlah observasinya. Oleh karena itu, perlu dilakukan seleksi variabel dengan menggunakan metode fast correlation based filter untuk menyelesaikan permasalahan klasifikasi.

METODE PENELITIAN

Data yang telah melewati tahap *preprocessing* dalam hal ini adalah normalisasi, selanjutnya akan dipartisi untuk menjadi data latih dan data uji menggunakan CVpartition dengan teknik *holdout* berdasarkan proporsi tertentu yang akan disesuaikan dengan kebutuhan. Lalu bilangan biner dibangkitkan random sebanyak fitur. Jika sebuah bit sama dengan 0 artinya fitur tersebut tidak dipilih, namun jika bit sama dengan 1 maka fitur tersebut akan dipilih untuk selanjutnya diklasifikasikan menggunakan algoritma JST. Data yang digunakan diperoleh dari Kent-Ridge Biomedis data Repository [10], dimana pada tabel data terdapat beberapa tipe kanker. Pada tabel jumlah kelas yang akan diklasifikasikan. Pada tabel samples, merupakan jumlah data sampel yang tersedia.



Gambar 1. Metode Penelitian

- Preprocessing, dalam proses ini hanya beberapa dataset saja yang menjalaninya yaitu cleansing data dari fitur yang kosong atau sama untuk seluruh kelasnya [11]. Tidak hanya itu, namun pada preprocessing ini juga untuk memisah data training dan testing menggunakan *Cross Validation* untuk beberapa dataset yang belum dipisahkan.
- Reduksi dimensi diperlukan karena Microarray data memiliki dimensi serta kompleksitas sangat tinggi, untuk itu diperlukan suatu proses yang dapat mengurangi kompleksitas dari microarray data tersebut. Pengurangan kompleksitas tersebut bertujuan untuk meminimalisir kesalahan. Reduksi dimensi bertujuan untuk mengambil fitur terpenting dengan *threshold* tertentu yang akan menghasilkan input untuk dilakukan klasifikasi. Reduksi dimensi bisa dilakukan dengan melalui *feature selection* atau *feature extraction*. Beberapa metode reduksi dimensi akan digunakan seperti Minimum Redundancy Maximum Relevance (mRMR), Mutual Information, PCA, Relief Method, dan lain sebagainya.
- Klasifikasi dilakukan setelah data sudah siap proses. Klasifikasi ini dibagi menjadi tiga proses. Pada proses yang pertama yaitu *training* akan dilakukan proses *learning* terhadap datasetnya yang menghasilkan suatu model. Proses selanjutnya yaitu *testing* akan dilakukan proses pengklasifikasian yang menghasilkan *class* dan dibandingkan dengan *class* target. Terakhir *class* hasil klasifikasi dan *class* target akan diproses untuk menghitung akurasi. Klasifikasi dilakukan dengan beberapa metode seperti Neural Network, Bayesian, Decision Tree dan lain-lain.

HASIL DAN PEMBAHASAN

Pada penelitian ini akan dilakukan pengujian terhadap dua jenis JST yakni JST Backpropagation dan JST Elman Recurrent Neural Network.

Tabel 1. Hasil Pengujian

Algoritma	MSE
Backpropagation	2.2251e214
ERNN	1.1192e443

Berdasarkan data pada table di atas, terlihat bahwa penerapan algoritma reduksi dimensi PCA bekerja lebih baik pada JST Elman Recurrent Neural Network. Hal ini karena memiliki MSE yang lebih rendah dibandingkan dengan JST algoritma backpropagation

KESIMPULAN

Hasil pengujian dalam penelitian ini menunjukkan bahwa algoritma PCA cukup handal diterapkan pada arsitektur JST ERNN dalam melakukan klasifikasi data kanker berdasarkan mikro array

DAFTAR PUSTAKA

- [1]. World Health Organization, Cancer fact sheet, <http://www.who.int/mediacentre/factsheets/fs297/en/index.html>, ccessed at March 2nd 2016.
- [2]. Vanitha, C.D.A., Devaraj, D. and Venkatesulu, M., 2015. Gene Expression Data Classification Using Support vector machines and Mutual Information-based Gene Selection. *Procedia Computer Science*, 47, pp.13-21.
- [3]. Siang, T.C., Soon, T.W., Kasim, S., Mohamad, M.S., Howe, C.W., Deris, S.,

- Zakaria, Z., Shah, Z.A. and Ibrahim, Z., 2015. A Review of Cancer Classification Software for Gene Expression Data. *International Journal of Bio-Science and Bio-Technology*, 7(4), pp.89-108.
- [4]. KR, S., 2011. Microarray Data Classification Using Support Vector Machine. *International Journal of Biometrics and Bioinformatics (IJBB)*, 5(1), p.10.
- [5]. Nurfalalah A., Adiwijaya, Suryani A.A., "Cancer Detection Based On Microarray data Classification Using PCA And Modified Backpropagation", *Far East Journal of Electronics and Communications*. 2015
- [6]. Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M. and Haussler, D., 2000. Support vector machines Classification And Validation Of Cancer Tissue Samples Using Microarray Expression Data. *Bioinformatics*, 16(10), pp.906-914.
- [7]. Purbolaksono, M.D., Widiastuti, K.C., Mubarak, M.S. and Ma'ruf, F.A., 2018, March. Implementation of mutual information and bayes theorem for classification microarray data. In *Journal of Physics: Conference Series* (Vol. 971, No. 1, p. 012011). IOP Publishing.
- [8]. Aydadenta, H., 2018, March. On the classification techniques in data mining for microarray data classification. In *Journal of Physics: Conference Series* (Vol. 971, No. 1, p. 012004). IOP Publishing.
- [9]. Nhita, F. and Adiwijaya, U.N., 2014. Forecasting Indonesian Weather through Evolving Neural Network (ENN) based on Genetic Algorithm. In *The Second International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE2014)* (pp. 78-82). The Society of Digital Information and Wireless Communication.
- [10]. Jinyan Li, Kent-ridge bio-medical data set repository, School of Computer Engineering Nanyang Technological University, Singapore, Downloaded at January 2013 from [URL: http://levis.tongji.edu.cn/gzli/data/mirror-kentridge.html](http://levis.tongji.edu.cn/gzli/data/mirror-kentridge.html)
- [11]. Adiwijaya, Wisesty, U.N. and Nhita, F., 2014. Studi of line search techniques on the modified backpropagation for forecasting of weather data in Indonesia. *Far East Journal of Mathematical Sciences*, 86(2), p.139.
- [12]. Bai Anita and S. K. Rath. Classification and Clustering using Intelligence Techniques: Application to microarray cancer data. Master Thesis of Department of Computer Science and Engineering National Institute of Technology Rourkela, India, 2013.
- [13]. Hieu Trung Huynh, Jung-Ja Kim, Yonggwon Won., 2009. Classification Study on DNA Microarray with Feed Forward Neural Network Trained by Singular Value Decomposition.
- [14]. *International Journal of Bio-Science and Bio Technology*, 1(1).
- [15]. Moorthy, K. and Mohamad, M.S., 2012. Random forest for gene selection and microarray data classification. In *Knowledge Technology* (pp. 174-183). Springer, Berlin, Heidelberg