

**ANALISIS DAN IMPLEMENTASI EVOLVING ARTIFICIAL NEURAL NETWORKS
DALAM KLASIFIKASI DATA MINING YANG MEMILIKI MASALAH IMBALANCE
CLASS DENGAN PENDEKATAN SAMPLING ANALYSIS AND IMPLEMENTATION
OF EVOLVING ARTIFICIAL NEURAL NETWORKS IN DATA MINING
CLASSIFICATION W**

Danang Risang Djati¹, Suyanto², Moch Arif Bijaksana³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Dunia nyata sering menyediakan data dimana distribusi suatu kelas lebih dominan dibandingkan kelas-kelas yang lain. Hal ini menimbulkan masalah tersendiri dalam dunia data mining dimana menyebabkan suatu classifier gagal mengenali kelas minoritas yang terkadang lebih berharga dibandingkan kelas mayoritas. Pendekatan **sampling** yang mampu mengoptimalkan kemampuan classifier dalam mengenali kelas minoritas dapat menjadi solusi bagi masalah ini. Jaringan Syaraf Tiruan (JST) sebagai salah satu classifier dalam data mining memiliki masalah tersendiri dalam menentukan struktur dan bobotnya. Evolutionary Programming (EP) dapat menjadi solusi untuk mengoptimalkan JST baik dalam struktur koneksi antar node maupun bobot yang menyertai koneksi-koneksi tersebut. Perpaduan antara EP dan JST ini disebut dengan Evolving Artificial Neural Networks.

Kata Kunci : evolutionary programming, evolving artificial neural networks, jaringan syaraf tiruan, klasifikasi kelas imbalance, sampling.

Abstract

Real world often providing data which has unbalanced distribution between each class. In data mining, this problem caused a classifier fail to classify minority class, and sometimes this minority class is more valuable than the majority class. Sampling approach which can optimize classifier in classifying minority class can be a solution for this problem. Artificial Neural Networks as a classifier in data mining has its own problem in deciding its structure and connection weights. Evolutionary Programming can be a solution for optimizing the structure and the connection weights. The combination of Evolutionary Programming and Artificial Neural Networks is called Evolving Artificial Neural Networks.

Keywords : artificial neural networks, evolutionary programming, evolving artificial neural networks, imbalanced class classification, sampling.

Telkom
University

1. Pendahuluan

1.1 Latar belakang

Perkembangan dunia industri yang menghasilkan data dalam jumlah yang besar mendorong manusia untuk mencari informasi yang berguna di balik data tersebut. Salah satu bentuk informasi yang bisa dicari adalah klasifikasi suatu data menjadi kelas-kelas tertentu. Jaringan Syaraf Tiruan (JST) sebagai salah satu teknik klasifikasi dalam *data mining* menawarkan berbagai keunggulan seperti mampu menangani klasifikasi dimana kelas targetnya lebih dari dua kelas (*non binary classification*), mampu menangani masalah *non linear*, mampu menangani atribut data yang *redundant*, dan mampu mengklasifikasikan suatu obyek dengan cepat apabila model dari hasil pembelajaran telah didapat.

Akan tetapi seringkali data yang disediakan dalam dunia nyata memiliki distribusi yang tidak seimbang antar kelas. Dalam *data mining*, hal ini disebut dengan masalah *imbalance class*. Hal ini menimbulkan masalah tersendiri dalam teknik klasifikasi yang digunakan dalam *data mining*. JST yang pada awalnya bertujuan untuk memaksimalkan akurasi dalam mengenali pola menjadi cenderung untuk mengenali pola yang dominan dan mengabaikan pola minoritas. Padahal terkadang dalam suatu data, kelas minoritas justru memiliki informasi yang bernilai lebih daripada kelas mayoritas.

Oleh karena itu dibutuhkan suatu pendekatan yang mampu mengoptimalkan kemampuan JST untuk lebih sensitif terhadap kelas minoritas. Pendekatan ini dapat berupa pendekatan *sampling*. Ide dasar dari *sampling* adalah untuk memodifikasi distribusi dari data sehingga kelas minoritas dapat direpresentasikan dengan baik dalam data *training* [7]. Beberapa jenis *sampling* yang ada antara lain *undersampling*, *oversampling*, dan gabungan dari *undersampling* dan *oversampling*.

Untuk menemukan struktur dan bobot JST yang optimal tidaklah mudah. *Evolutionary Algorithms* (EA) dapat menjadi solusi bagi masalah ini. Ketertarikan penggunaan EA untuk mendesain arsitektur JST telah berkembang dalam tahun-tahun belakangan karena EA mampu berevolusi mencari arsitektur optimal tanpa intervensi dari luar dan menghilangkan pencarian struktur optimal secara manual yang membosankan dan melelahkan [1]. Algoritma-algoritma yang termasuk dalam EA antara lain Algoritma Genetika, *Evolution Strategies*, *Evolutionary Programming*, dan *Genetic Programming*. *Evolutionary Programming* (EP) sebagai salah satu EA memiliki keunggulan dibandingkan EA lainnya dalam pencarian struktur dan bobot JST. Dengan karakter EP yang hanya menggunakan mutasi tanpa rekombinasi dalam menghasilkan anak, membuat EP dapat mengurangi efek negatif dari masalah permutasi [12].

1.2 Perumusan masalah

Masalah pokok yang akan diteliti adalah :

1. Bagaimana menerapkan EP untuk mencari bobot dan arsitektur JST.
2. Bagaimana menerapkan metode *sampling* untuk menangani masalah *imbalance class*.

Dalam penyusunan tugas akhir ini permasalahan dibatasi dalam beberapa hal, yaitu :

1. Perangkat lunak yang dihasilkan untuk menangani studi kasus Data Mining Cup 2007 (DMC 2007).
2. JST yang digunakan adalah *Feed Forward Neural Networks* dengan *supervised learning*.
3. EP digunakan untuk menentukan bobot dan arsitektur JST saja.

1.3 Tujuan

Tujuan dari penelitian tugas akhir ini adalah sebagai berikut :

1. Mengimplementasikan EANN dengan menggunakan EP untuk mencari struktur dan bobot JST yang optimal.
2. Menerapkan *sampling* untuk memodifikasi data *training* agar EANN dapat mengklasifikasikan kelas *imbalance* dengan baik.
3. Menghitung performansi EANN menggunakan *recall*, *precision*, *F-measure*, akurasi, dan parameter yang telah ditetapkan dalam DMC 2007.

1.4 Metodologi penyelesaian masalah

Metodologi yang akan digunakan dalam merealisasikan tujuan dan pemecahan masalah di atas adalah dengan menggunakan langkah-langkah berikut.

1. Studi Literatur :
 - a. Pencarian referensi, mencari referensi dan sumber-sumber lain yang layak dan berhubungan dengan *data mining*, JST, EP, dan *sampling*.
 - b. Pendalaman materi, mempelajari dan memahami materi yang berhubungan dengan tugas akhir.
2. Mempelajari konsep dari JST dan EP serta *sampling* yang akan digunakan dalam implementasi perangkat lunak.
3. Melakukan analisis terhadap parameter- parameter JST, EP, dan *sampling*.
4. Melakukan implementasi perancangan perangkat lunak.
5. Melakukan pengujian perangkat lunak dengan memasukkan data yang akan dievaluasi serta mencatat hasil keluaran program.
6. Pengambilan kesimpulan dan penyusunan laporan tugas akhir.

5. Penutup

5.1 Kesimpulan

1. Nilai *fitness* yang semakin besar belum tentu mendapatkan *profit* yang semakin besar pula.
2. Nilai akurasi yang semakin besar belum tentu mendapatkan *profit* yang semakin besar pula.
3. Nilai akurasi tidak dapat dijadikan tolak ukur keberhasilan EANN dalam klasifikasi *imbalance class*.
4. Penggunaan *sampling* membantu EANN untuk mengenali kelas *imbalance* lebih baik tanpa mengorbankan kelas mayoritas.

5.2 Saran

1. Selain menggunakan pendekatan *sampling*, untuk mengatasi masalah *imbalance class* dapat juga menggunakan *cost-sensitive learning*.
2. Untuk mengubah bobot arsitektur JST dapat juga digunakan *hybrid training*, algoritma genetika, atau *evolution strategies*.

Referensi

[1]	Abraham, A, <i>Meta-Learning Evolutionary Artificial Neural Networks</i> , Department of Computer Science, Oklahoma State University, Oklahoma, 2004.
[2]	Alexander Yun-chung Liu, <i>The Effect of Oversampling and Undersampling on Classifying Imbalanced Text Datasets</i> , University of Texas, 2004.
[3]	Data Mining Cup Contest 2007, Germany.
[4]	Haykin, S, <i>Neural Networks: A comprehensive Foundation</i> , Prentice-Hall Inc, New Jersey, 1999.
[5]	Hermawan, A, <i>Jaringan Saraf Tiruan: Teori dan Aplikasi</i> , ANDI, Yogyakarta, 2006.
[6]	Jong Jek Siang, <i>Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab</i> , ANDI, Yogyakarta, 2005.
[7]	Pang Nin Tan, Michael Steinbach, Vipin Kumar, <i>Introduction to Data Mining</i> , Pearson Education Inc, Boston, 2006.
[8]	Pressman, Roger S., <i>Software Engineering, A Practitioner's Approach</i> . McGraw-Hill Companies Inc., New York, 2001.
[9]	Suyanto, <i>Algoritma Genetika Dalam MATLAB</i> , ANDI, Yogyakarta, 2005.
[10]	Weiss, G.M., <i>Mining With Rare Cases</i> , Department of Computer and Information Science, Fordham University.
[11]	Xin Yao, <i>A Gentle Introduction to Evolutionary Computation</i> , x.yao@cs.bham.ac.uk, didownload pada tanggal 10 Juli 2007.
[12]	Xin Yao, <i>Evolving Artificial Neural Networks</i> , Proceedings of the IEEE, 87(9):1423-1447, September 1999.