**Review Automated Feature Design for Numeric Sequence Classification by Genetic Programming**

*By: Dustin Y. Harvey and Michael D. Todd/ 2015*

*Review by: Anna Dina Kalifia / 15/388458/PPA/04897*

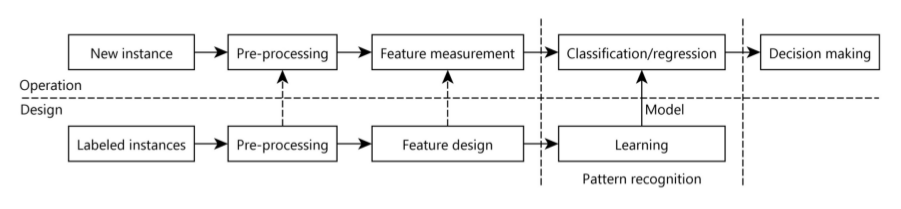
Pendekatan umum yang sering dilakukan diperlukan untuk memperluas pengenalan pola untuk memasukkan desain fitur dan seleksi untuk pengurutan numerik, seperti time series, dalam proses belajar itu sendiri.

Makalah ini mengusulkan Genetic Programming (GP) sebuah pendekatan baru untuk desain fitur otomatis yang disebut Autofill. Dalam metode ini, varian GP berkembang populasi fitur kandidat dibangun dari sebuah perpustakaan fungsi urutan penanganan. metode optimasi numerik, termasuk melalui pendekatan hybrid, memastikan bahwa nilai fitness untuk kandidat algoritma diukur menggunakan parameter nilai optimal

Autofeed merupakan fitur desain sistem otomatis pertama untuk urutan numerik untuk memanfaatkan kekuatan dan efisiensi dari kedua optimasi numerik dan algoritma pengenalan pola standar.

Potensi aplikasi termasuk pemantauan sinyal elektro kardiogram untuk indikasi gagal jantung, analisis lalu lintas jaringan untuk sistem deteksi intrusi, pengukuran getaran untuk bantalan kondisi penentuan pada perputaran mesin, dan aktivitas kartu kredit untuk deteksi penipuan.

Pada makalah ini menjelaskan bahwa data sequence didefinisikan sebagai satu set sampel numerik memberi perintah pada time series dalam domain waktu atau pengukuran spektral dalam domain frekuensi.



Kebanyakan algoritma pengenalan pola mengikuti paradigma statistik ditunjukkan pada Gambar. 1 dimana setiap contoh diwakili oleh vektor fitur menggambarkan satu titik dalam multivariat, ruang fitur skalar. Literatur pengenalan pola mengandung banyak metode untuk melakukan klasifikasi dan regresi tugas pada ruang fitur serta meningkatkan ruang fitur melalui seleksi dan konstruksi fitur.

Dalam operasi, fitur vektor diukur dan model belajar yang digunakan untuk mengklasifikasikan atau mundur kasus baru. Masalah desain dapat didekomposisi menjadi empat komponen sebagai berikut, tanpa pendekatan tunggal optimal yang sesuai untuk setiap componennya;

* Komponen 1: algoritma Fitur.
* Komponen 2: Parameter dalam algoritma fitur.
* Komponen 3: Temukan Fitur dan konstruksi.
* Komponen 4: algoritma pengenalan pola.

Latar Belakang Desain AutoFead

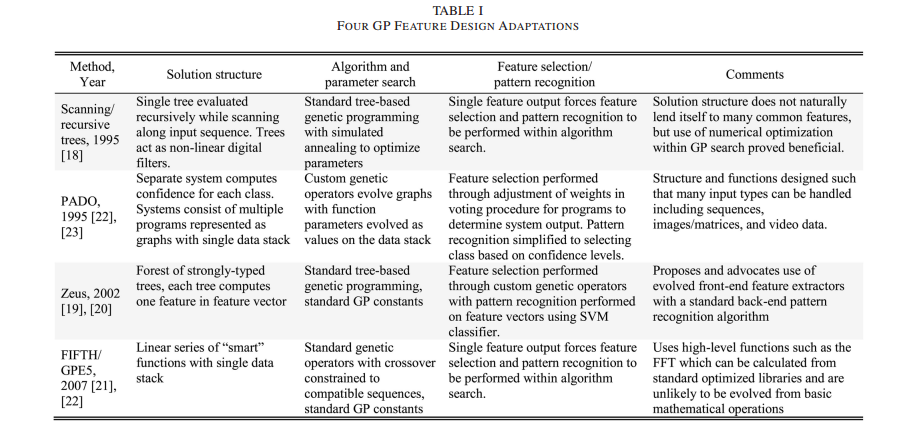
1. Pengantar genetik Pemrogramming

Secara khusus, metode berbasis GP telah menunjukkan keberhasilan pada pemecahan masalah dengan karakteristik sebagai berikut

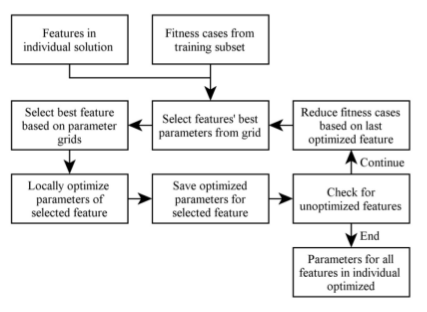
* 1. Pengetahuan Minimal dan pemahaman tentang domain masalah.
  2. Ukuran dan struktur solusi yang tidak diketahui yang diinginkan.
  3. Sejumlah besar data digital yang tersedia atau dapat diperoleh.
  4. Solusi yang baik dengan mudah diuji tapi tidak langsung diperoleh.
  5. Solusi analitis tidak tersedia.
  6. Solusi optimal diinginkan tetapi dengan perkiraan yang dapat diterima
  7. Peningkatan kinerja Kecil yang penting

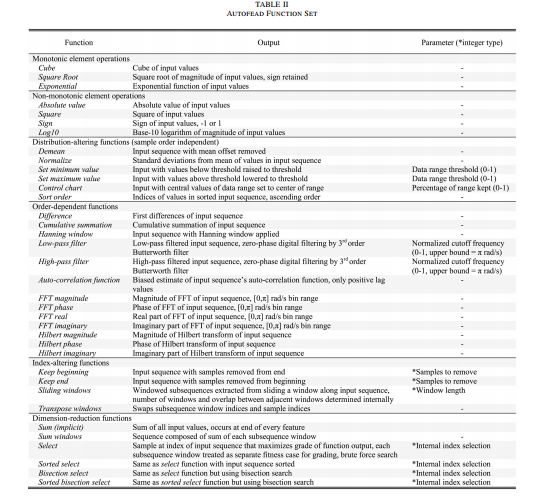
Makalah ini membahas masalah yang difokuskan pada identifikasi panjang dari suatu peristiwa secara berurutan dan menemukan fitur sederhana dari hubungan yang multivariat. Makalah ini mengusulkan metode baru yang disebut Autofead untuk desain fitur otomatis.

Tujuan keseluruhan dari makalah ini adalah untuk mengembangkan sebuah sistem yang dapat mengambil kesimpulan dari inputan data optimal dari sequensial numerik , Firur berbasis minimal di set untuk proses pengambilan keputusan.



Autofeed menggunakan varian GP berevolusi fitur untuk input ke algoritma pengenalan pola standar melalui pendekatan wrapper. Misalnya, fitur yang menghitung energi dari urutan time series akan menggunakan dua fungsi algoritma persegi kemudian sum. Setiap algoritma fitur berakhir dengan penjumlahan untuk memastikan hasil. Tidak seperti GP tradisional, struktur solusi tidak memiliki terminal konstan; sebaliknya, fungsi memiliki parameter yang numerik dioptimalkan dalam setiap generasi. Pada alur dibawah ini dapat dilihat set functionnya



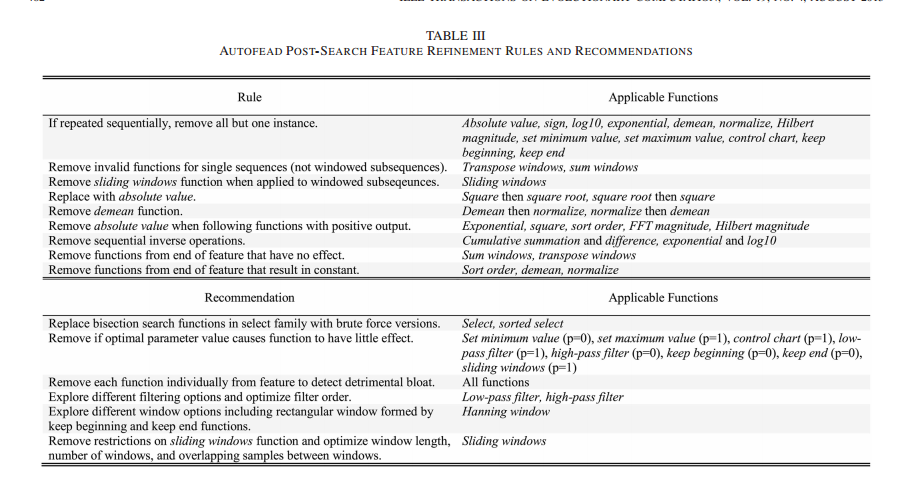


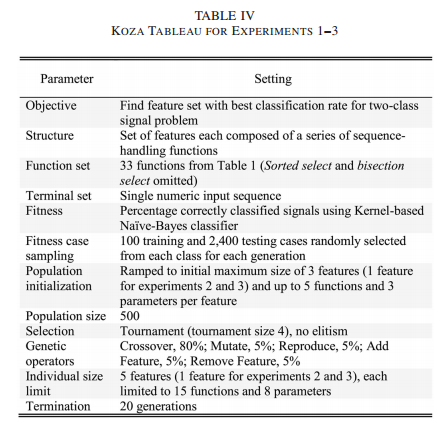
Terminal set pada masalah ini hanya terdiri dari satu atau lebih inputan sequensial. Sementara sistem ini dirancang untuk masukan langsung dari data pengukuran baku seperti time series, pengetahuan domain dapat menunjukkan bahwa representasi alternatif yang bermanfaat untuk aplikasi tertentu.

Proses berikut merangkum semua langkah optimasi untuk fitur yang berisi dua parameter integer, I1 dan I2, dan dua parameter terus menerus, C1 dan C2.

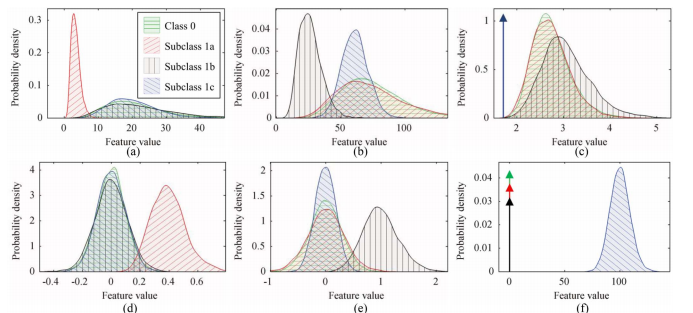
* 1. Mengatur C1 dan C2 ke nilai default.
  2. Membangun adaptif parameter grid untuk I1 dan I2 dan set ke nilai-nilai terbaik dari grid.
  3. Membangun parameter grid untuk C1 dan set ke nilai terbaik dari grid.
  4. Membangun parameter grid untuk C2 dan set ke nilai terbaik dari grid.
  5. Lokal mengoptimalkan I1 dan I2 menggunakan pencarian pembelahan.
  6. Lokal mengoptimalkan C1 dengan metode Brent.

Lokal mengoptimalkan C2 dengan metode Brent.

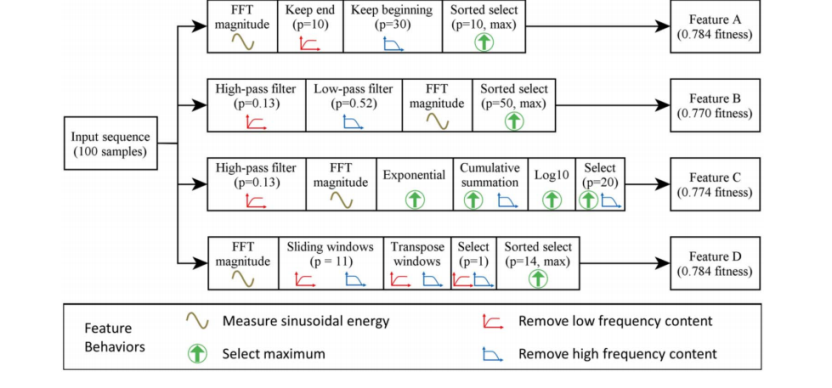




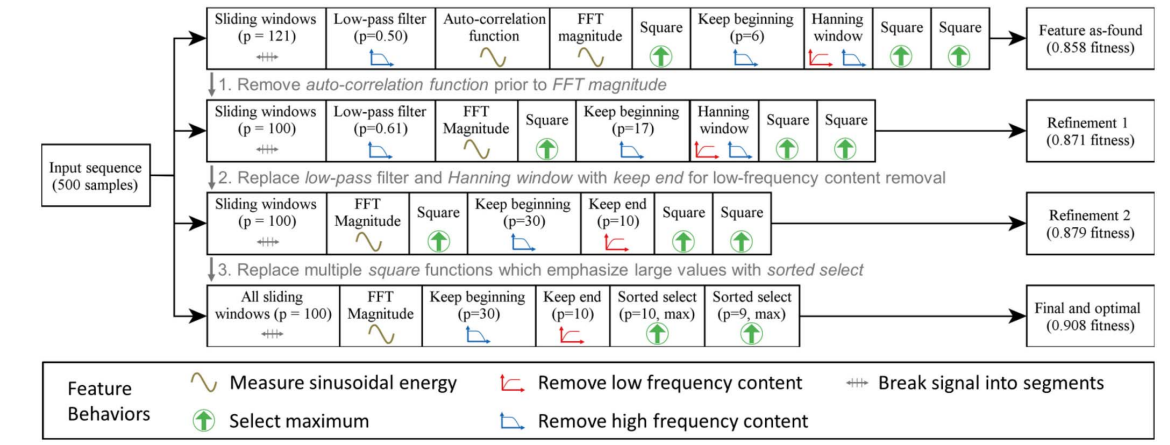
Experiment 1 comparison of feature probability density functions for optimal and Autofead solutions. (a) Autofead, subclass 1a. (b) Autofead,subclass 1b. (c) Autofead, subclass 1c. (d) Optimal, subclass 1a. (e) Optimal, subclass 1b. (f) Optimal, subclass 1c.



Experiment 2 feature behavior analysis for four Autofead solutions



Experiment 3 feature behavior analysis for Autofead solution, two intermediate levels of refinement, and final solution.



Dalam makalah ini, metode baru GP berdasarkan diusulkan untuk mengotomatisasi desain fitur untuk klasifikasi urutan numerik melalui sistem sepenuhnya berbasis data disebut Autofead. Hasil disajikan untuk beberapa waktu percobaan klasifikasi series.