

Perancangan struktur model hidden markov untuk prediksi ekson DNA gen plasmodium falciparum

Suhartati Agoes, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20277907&lokasi=lokal>

Abstrak

Pemrosesan sinyal genom (Genomic Signal Processing) seperti Deoxyribonucleid Acid (DNA) dan protein dapat dilakukan untuk memprediksi ekson atau coding region suatu gen. Metoda yang paling banyak digunakan adalah Hidden Markov Model (HMM) yang kaya akan struktur matematik dan berpotensi untuk mengetahui lebih banyak tentang sumber sinyal tanpa harus tersedia sumber tersebut.

Pada penelitian ini dirancang struktur model yang menggunakan metoda HMM dengan struktur dasar model sesuai struktur ekson pada coding sequence (CDS) sehingga dapat memprediksi ekson DNA Plasmodium falciparum. Jumlah state model pada struktur dasar adalah 5, 7 dan 9 sedangkan untuk struktur pengembangan ditentukan secara acak yaitu 20, 30, 50 dan 100 state. Proses training HMM menggunakan algoritma Viterbi dan proses testing HMM menggunakan kedua algoritma yaitu Viterbi dan Baum-Welch sedangkan kinerja model menggunakan parameter Correlation Coefficient (CC). Sekuen yang digunakan adalah 152 sekuen DNA Plasmodium falciparum dengan panjang minimum 684 pb dan maksimum 10095 pb.

Hasil simulasi pada umumnya menghasilkan nilai CC rata-rata lebih baik dengan menggunakan algoritma Viterbi dibandingkan dengan algoritma Baum-Welch. Pada struktur dasar model 9 state menghasilkan nilai CC paling baik dibandingkan dengan struktur dasar model lainnya yaitu 0,7289 dengan menggunakan algoritma Viterbi dan 0,7166 dengan menggunakan algoritma Baum-Welch. Sedangkan untuk pengembangan model diperoleh nilai CC rata-rata paling baik untuk Model 2 dengan 100 state yaitu 0,7827 dengan menggunakan algoritma Baum-Welch dan 0,7820 dengan menggunakan algoritma Viterbi. Waktu proses testing HMM rata-rata seluruh model hampir dua kali lebih lama dengan algoritma Baum-Welch dibandingkan dengan algoritma Viterbi.

<i>Genomic signal processing like as Deoxyribonucleid Acid (DNA) and protein can be done for exon prediction or coding region of the gene. The most used method is Hidden Markov Model (HMM) which has various mathematical structures and potentially capable of learning a great deal of signal source without having to have the source available.

The model structure designed in this research is using the HMM method with based structure model in accordance with exon structure in the coding sequence (CDS) in order to predict of DNA Plasmodium falciparum. The state number of model in the basic structure are 5, 7 and 9 states, meanwhile the expansion structure was randomly defined having 20, 30, 50 and 100 states. The HMM training process are using the Viterbi algorithm and the HMM testing are using both algorithms, Viterbi and Baum-Welch, meanwhile the performance indicator of the model are using the Correlation Coefficient (CC). It is using 152 sequences -of

DNA *Plasmodium falciparum* with the minimum length of 634 base-pair (bp) and maximum length of 10095 bp.

In general, the simulation results produced the best average of CC value by using Viterbi algorithm rather than Baum-Welch. In the basic structure of 9 states model produced the best CC value compared with the other basic structure models with 0.7289 using the Viterbi Algorithm and 0.7166 using Baum-Welch. Meanwhile, for the model expansion, the best CC average value is for Model 2 with state number 100 with 0.7827 using the Baum-Welch algorithm and 0.7820 using Viterbi. The average processing time of the HMM tests for all models using the Baum-Welch algorithm are almost two times slower than using Viterbi algorithm.