

ABSTRAK

Sebuah alat bantu dibutuhkan untuk mempelajari kitab-kitab yang menyimpan kekayaan budaya Jawa. Alat bantu ini dibuat untuk mengenali dan menterjemahkan aksara Jawa menjadi bahasa yang lebih dikenal masyarakat luas. Hidden Markov Models (HMM) menjadi pendekatan yang diajukan, mengingat keberhasilannya mengenali banyak jenis huruf maupun pada pengolahan sinyal.

HMM merupakan metode stokastik. HMM memproses data diskret aksara Jawa yang telah dipindai dan mengalami *pre-processing*, menjadi sebuah rangkaian rantai Markov. Rantai ini terdiri dari banyak *state*, dengan transisi probabilitas yang menghubungkan *state* satu dengan yang lain. Permodelan rantai Markov sekuens kiri-kanan merupakan model dari data diskret aksara Jawa ini.

Terdapat 20 jenis aksara yang ingin dikenali, sehingga untuk setiap aksara dibuat satu model dari 50 sampel aksara tersebut dengan algoritma Baum-Welch yang mencari *means* dan *varians* dari data sampel yang didapat. Di sinilah fase pelatihan HMM dilaksanakan.

Fase pengujian dilakukan dengan algoritma Viterbi. Angka kemiripan terbesar dicari dengan membandingkan data baru dengan model-model yang ada. Setelah itu, seluruh data pengujian ini dievaluasi dengan metode *5-fold cross validation* untuk mencari nilai akurasi dari 1000 data yang diberikan.

Penelitian ditujukan untuk mencari nilai akurasi dan mempelajari cara HMM mengenali data aksara Jawa. Pengujian 8 bentuk pengolahan *feature* dan sejumlah *state* juga dilakukan. Hal ini ditujukan untuk mencari bentuk *feature* terbaik dengan jumlah *state* terkecil. Akurasi yang baik harus didapatkan dari model yang memiliki kompleksitas waktu paling sedikit.

Dari penelitian, didapati bentuk *feature* terbaik pengenalan aksara Jawa dengan metode HMM ialah bentuk 1-Vertikal, dengan akurasi sebesar 85,7 %. Walaupun bukan *feature* dengan angka akurasi tertinggi, 1-V dipilih menjadi bentuk *feature* terbaik dengan mempertimbangkan jumlah *state* terendah, yaitu sebanyak 16 *state* saja.

ABSTRACT

A tool is needed to learn from the books which treasure the Javanese culture. This tool is made to recognize the Javanese characters to be a better known language by human society. Hidden Markov Models (HMM) is the proposed approach to recognize Javanese character, consider from its succeed to recognize many kind of characters and signal processing.

HMM is a stochastic method. It process the scanned and pre-processed discrete Javanese characters, to be a sequence of chain called Markov chain. This generated chain consists of a number of states with hidden parameters and also transition probabilities which make these states are connected to others. Left-right sequence Markov chain will be the model to the discrete Javanese character.

There are 20 characters to be recognized, as one model is generated for each character from 50 sample data using Baum-Welch algorithm. Means and variance are generated to represent the character from the data samples. This is the training phase of HMM.

Viterbi algorithm is used at the testing phase. Maximum similarity found by comparing models to the given character. After the testing and training phase are completed, 5-fold cross validation performed to find accuracy from 1000 given data.

The main goal of this study is to find accuracy. The secondary goal is to know how HMM recognize the Javanese characters. As the study is held, we also experienced 8 features to be handled with a number of state. This experiment is to find the best feature extraction with least state. Good accuracy should be gained with a model which has less time complexity.

Experimental result shows that 1-Vertical is the best feature, considering from the least state number and good accuracy. Tough it is not represent the best accuracy, but 1-Vertical feature needs only 16 state to recognize 85,7% from total number of the given data.