

IMPLEMENTASI DAN ANALISIS BIASED REGULARIZATION SUPPORT VECTOR MACHINE (BRSVM) PADA KASUS PENGENALAN HURUF TULISAN TANGAN

Laily Fikriyah Shochibah A.¹, Retno Novi Dayawati², Tjokorda Agung Budi Wirayuda³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Berkembangnya teknologi komputerisasi membuat metode penulisan konvensional bergeser, namun cara ini akan terus bertahan selama menulis menggunakan pena dan kertas lebih memberikan kenyamanan dan kesenangan dibanding menggunakan keyboard dan PC table. Untuk menjembatani hal tersebut dengan perkembangan teknologi, maka diperlukan suatu sistem pengenalan tulisan tangan yang mampu mengubah tulisan tangan kertas menjadi format teks komputer. Hal inilah yang mendorong munculnya berbagai penelitian mengenai pengenalan tulisan tangan. Dan hingga saat ini tingkat kesempurnaan dari penelitian tersebut belum ada yang mampu menggantikan kemampuan semantik manusia secara sempurna. Penggunaan SVM sebagai metode pembelajaran dalam sistem tersebut dapat membantu penentuan hyperplane terbaik yang memisahkan kelas-kelas karakter huruf tulisan tangan. Permasalahannya tingkat akurasi klasifikasi data tersebut menjadi kurang optimal ketika data tulisan tangan yang diuji berbeda jauh dengan sistem yang ada. Dan dengan menggunakan Biased Regularization pada SVM (BRSVM), tingkat akurasi dari pengujian terhadap data tersebut dapat ditingkatkan dengan cara membiaskan resiko umum yang muncul dari pengenalan tulisan tangan SVM.

Kata Kunci : Hand Writing Recognition , BRSVM, SVM, Biased Regularization

Abstract

Computerization technology development takes over the conventional writing method, but this way will survive as long as paper writing is more comfortable and gives the happiness than keyboard writing or PC table writing. For bridging this case with technology development, we need a system to recognize hand writing which is able to change this hand writing become a computer text format. This case encourages a lot of research about hand writing recognition. For this moment, none of those researches are able to replace the perfection rate of human semantic ability.

Using SVM as learning method in the system will helps determination of the best hyperplane which is separating word character classes of hand writing. The problem is the accuracy rate of data classification become not optimal when the testing hand writing data is very different from the existing system.

With using Biased Regularization on SVM (BRSVM), accuracy rate from testing for the data can be increased with refracting the general risk that is appear in hand writing recognition SVM.

Keywords : Hand Writing Recognition, BRSVM, SVM, Biased Regularization

1. Pendahuluan

1.1 Latar Belakang

Tulisan tangan menggunakan media kertas dan pena merupakan alat komunikasi dan penyimpanan informasi yang cukup *konvensional* dan mampu bertahan dalam beberapa dekade hingga saat ini. Seiring dengan berkembangnya teknologi *komputerisasi*, metode penulisan yang menggeser metode *konvensional* mulai bermunculan seperti penggunaan *keyboard* dan *PC table*. Dilain pihak metode *konvensional* masih banyak diminati. Dimana saat tersedia *keyboard* dan *PC table* sebagai pengganti pena dan kertas, kebanyakan orang masih memilih catatan/ kertas manakala mereka ingin mencatat [4]. Cara ini akan terus bertahan selama menulis menggunakan pena dan kertas lebih memberikan kenyamanan dan kesenangan dibanding menggunakan *keyboard* dan *PC table* [8][4]. Namun tidak dapat dipungkiri bahwa tulisan tangan yang *terkomputerisasi* lebih mudah diolah dibandingkan dengan tulisan tangan *konvensional*. Menanggapi hal ini, diperlukan suatu teknologi yang mampu menjembatani antara penggunaan tulisan tangan dengan dunia komputerisasi. Teknologi tersebut adalah teknologi pengenalan tulisan tangan. Dengan harapan, kedepan teknologi ini mampu disertakan pada suatu alat yang sering digunakan sebagai media cetak seperti printer misalnya.

Pengenalan tulisan tangan adalah penyalinan data tulisan tangan ke dalam suatu format digital, dengan tujuan memproses data tulisan secara elektronik agar mampu di kenali menggunakan kemampuan yang sama atau hampir sama dengan kemampuan manusia [11]. Dan dalam perkembangannya pengenalan tulisan tangan ini memiliki banyak manfaat seperti pengenalan alamat pada bagian atas amplop, verifikasi tandatangan dan identifikasi penulis, penyalinan (*offline*) arsip bersejarah historos, arsip kartu, data silsilah dan jurnal[5]. Hal inilah yang mendorong munculnya perkembangan penelitian mengenai teknologi pengenalan tulisan tangan. Hingga saat ini tingkat kesempurnaan dari penelitian tersebut belum ada yang mampu menggantikan kemampuan semantik manusia secara sempurna.

Adapun proses pengenalan tulisan tangan tersebut secara umum terdiri dari berbagai tahapan proses pengolahan dan pengenalan citra. Proses yang memicu permasalahan ketidak sempurnaan dalam sistem pengenalan tulisan tangan adalah proses pengenalan (*learning*) tulisan. Pada tahap tersebut terjadi proses pengenalan pola karakter dari tulisan tangan tangan yang diproses. Dan untuk mencapai kesempurnaan, dibutuhkan suatu proses pengenalan karakter tulisan tangan yang handal. Sehingga ketika ada tulisan tangan yang kualitasnya buruk bisa dideteksi menggunakan sistem tersebut dengan benar. Dan proses ini akan terwujud jika metode pembelajaran dalam proses tersebut mampu mengenali tulisan tangan tersebut dengan baik.

Saat ini, salah satu metode pembelajaran yang marak digunakan pada pengenalan pola adalah *Support Vector Machine* (SVM). Adapun kemampuan evaluasi dari algoritma ini dalam berbagai aplikasinya yang menempatkannya sebagai *state of the art* dalam pattern recognition, yang memicu maraknya penggunaan algoritma ini[3]. Selain itu SVM juga merupakan metode pembelajaran yang mempunyai prinsip kerja *Structural Risk Minimation* (SRM)

dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah kelas pada *input space*-nya [3]. Dengan adanya prinsip tersebut, ketika SVM tersebut sudah menemukan *hyperplane* terbaik pada *input space*-nya maka solusi yang tercapai adalah *global optimum* [10]. Namun dalam persoalan pengenalan tulisan tangan, tingkat pencocokan/penyelarasan itu menjadi kurang optimal ketika proses penyelarasan data tulisan tangan seorang penulis (*single-user*) dengan data training (*multi-user*) yang ada itu berbeda jauh. Hal ini yang melatarbelakangi penggunaan pendekatan *Biased Regularization* pada SVM untuk persoalan pengenalan tulisan tangan. Dengan pendekatan tersebut resiko umum dari penyelarasan data yang ada di SVM akan terminimalisasi dengan membiarkan resiko umum yang muncul [13]. Sehingga tulisan tangan *single-user* yang berbeda jauh dengan data *multi-user*, mampu dikenali ketika menggunakan algoritma SVM dengan pendekatan *Biased Regularization SVM* (BRSVM).

1.2 Perumusan masalah

Berdasarkan latar belakang di atas, maka rumusan masalah pada penelitian tugas akhir ini adalah:

1. Bagaimanakah tahap *preprocessing* citra tulisan tangan agar mampu diolah dan dilatih menggunakan algoritma BRSVM.
2. Bagaimana membangun suatu sistem yang dapat mengenali tulisan tangan berupa citra digital tulisan tangan menggunakan algoritma BRSVM.
3. Bagaimana menguji dan menganalisis kerja sistem.

1.3 Tujuan

Penelitian tugas akhir ini bertujuan untuk:

1. Merancang dan membangun sistem pengenalan tulisan tangan dengan menggunakan algoritma BRSVM.
2. Menguji dan menganalisis kerja sistem dengan memperhitungkan tingkat akurasi kebenaran pengenalan dari sejumlah data tulisan tangan

1.4 Batasan Masalah

Batasan penelitian tugas akhir ini adalah:

1. Sistem yang dibangun adalah sistem pengenalan tulisan tangan *offline*.
2. Input dari sistem adalah citra digital tulisan tangan latin hasil scan tulisan tangan yang bertipe *bitmap* dan *jpeg*.
3. Jenis Tulisan tangan yang diuji adalah citra tulisan tangan *spaced discrete character*
4. Karakter dari citra digital masukan adalah a-z/A-Z,0-9, operasi aritmatik (+, -, x, /).
5. Data *multi-user* diperoleh dari contoh tulisan tangan dari beberapa penulis (*user*) dengan jumlah minimal 50 penulis.
6. Dokumen input minimal berukuran 100 – 300 dpi.
7. Kernel yang digunakan adalah *kernel Gaussian*.

1.5 Metodologi penyelesaian masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini adalah :

1. Studi Literatur
 - a. Pencarian referensi
Mencari referensi dari sumber-sumber yang layak yang berhubungan dengan pengolahan citra digital (*preprocessing*), SVM, BRSVM dan sumber-sumber lain yang membantu dalam memberikan gambaran dan detail dari landasan teori pada pembangunan sistem ini.
 - b. Pendalaman materi
Mempelajari materi dengan cara mempelajari dan memahami materi yang berhubungan dengan tugas akhir.
2. Analisis kebutuhan sistem
Pada tahap ini terjadi analisis kebutuhan dari sistem perangkat lunak yang akan dibangun berupa perangkat lunak pengenalan karakter tulisan tangan menggunakan algoritma BRSVM.
3. Perancangan sistem
Hasil analisis kebutuhan sistem akan dijadikan dasar dalam tahap perancangan sistem.
4. Implementasi sistem
Hasil dari perancangan sistem akan diimplementasikan menggunakan bahasa pemrograman Matlab 7.
5. Evaluasi
Evaluasi terhadap hasil implementasi dilakukan untuk mengukur hasil rancangan. Proses dari evaluasi tersebut berbentuk konsultasi dengan dosen pembimbing dan pihak lain yang dapat membantu pencapaian tugas akhir ini secara sempurna.
6. Analisis hasil Implementasi
Proses implementasi akan menghasilkan keluaran data yang selanjutnya akan dianalisis berdasarkan rancangan kerja sistem. Hasil analisis tersebut digunakan untuk menarik kesimpulan dari penelitian tugas akhir ini.
7. Pengambilan kesimpulan dan penyusunan laporan tugas akhir
Hasil penelitian akan disusun menjadi laporan yang didalamnya akan ada perbandingan antara kesimpulan dan hipotesa awal.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan penelitian yang dilakukan pada tugas akhir ini, diperoleh beberapa kesimpulan bahwa :

1. Dalam merancang dan membangun sistem pengenalan menggunakan algoritma BRSVM, sistem tidak bisa dibangun secara independent. Mengingat dalam membangun BRSVM dibutuhkan informasi klassifier *generic recognizer* (algoritma SVM) yang dibiaskan kedalam data training baru berupa data training single-user.
2. Sistem BRSVM mampu meningkatkan akurasi pengenalan tulisan tangan suatu user bias yang telah ditraining terlebih dahulu menggunakan BRSVM dengan jumlah data training minimum 10 data tiap karakter. Dengan model BRSVM yang memberikan akurasi terbaik adalah model BRSVM dengan jenis multiclass OAO dan jumlah faature 1x122.
3. Suatu user dikatakan sebagai user bias ketika suatu user memiliki nilai akurasi dibawah 55 % pada saat diuji menggunakan sistem pengenalan tulisan tangan SVM *generic recognizer*.

5.2 Saran

Tugas akhir ini masih jauh dari sempurna, olehkarena itu masih dibutuhkan pengembangan lebih lanjut agar sistem ini menjadi lebih baik dan dapat menghasilkan output yang lebih memuaskan. Adapun saran pengembangan tugas akhir ini adalah :

1. Sistem BRSVM pada penelitian belum mencakup penentuan batas minimum dan maksimum jumlah data training yang digunakan untuk tiap-tiap karakter.
2. Penentuan huruf besar, huruf kecil dan angka pada tahap Recognize saat penggunaan feature 1x122 pada sistem ini masih masih belum sempurna dan membutuhkan proses manual. Sehingga perlu adanya sistem otomatisasi penentuan kelas karakter untuk pengembangan lebih lanjut agar proses pengenalan lebih mudah dan terhindar dari kesalahan.
3. Sistem ini hanya mampu bekerja pada tulisan tangan *Spaced discrete characters*. Untuk pengembangan lebih lanjut diharapkan sistem mampu bekerja pada semua jenis tulisan tangan.

Daftar Pustaka

- [1] Abe, shigeo.2005. *Support vector Machine for Pattern Classification*. United States of America : Springer
- [2] Ahmad, Usman. 2005. *Pengolahan Citra Digital dan Teknik Pemrogramannya*. Yogyakarta: Graha Ilmu.
- [3] Anto Satriyo Nugroho, Arief Budi Witarto, and Dwi Handoko. *Support Vector Machine*.2003.Follow : IlmuKomputer.com.
- [4] C`dric Schaller. *An offline handwriting Recognition system using Constant Density Hidden Markov Models (CDHMM)*. Technology and Medicine Departement Of Computing, University of London Imperial Collage of Science.
- [5] Caesar Ogole. *Offline Cursive Handwriting Recognition using Holistic Features and Learning Vector Quantization*. Departement of Computing Science, University of Groninge, Netherlands.
- [6] Campbell, Colin and Nello Cristianini. *Simple Learning Algorithms for training Support Vector Machines*. Unevirsity of Brisbol. United Kingdom.
- [7] Dharma, Eddy Muntina. 2005. *Diktat Kuliah Digital Image Processing [DIP] Fundamental*. Jurusan Teknik Informatika STT Telkom Bandung.
- [8] N.Shanthi, and K.Duraiswamy. *Performance Comparison of Different Image Sizes for Recognition Unconstrained Handwritten Tamil Characters using SVM*. Depastement of Information Technology, Tiruchengode, India.
- [9] Novelianty.S. Tugas Akhir. 2007. *Implementasi Jaringan Syaraf Tiruan Learning Vector Quantization dalam Kasus Pengenalan Karakter Tulisan Tangan*. Jurusan Teknik Informatika STT Telkom Bandung.
- [10] Santosa, Budi. 2007. *Data mining teosri dan aplikasinya*. Yogyakarta: Graha Ilmu.
- [11] Simon Gunter. *Multiple Classifier System in Offline eCursive Handwriting Recognition*. Departement of Philosophisch-naturwissenschaftlicen, Bern University.
- [12] Sugiharto, Aris. 2006. *Pemrograman GUI dengan MATLAB*. Yogyakarta : Andi.
- [13] Wolf Kienzle, and Kumar Chellapilla. *Personalized Handwriting Recognition via Biased Regularization*. Microsoft Research, One Microsoft Way, Redmond, WA, 98052, USA