

Conditional Neural Fields untuk Pengenalan Fase Gerak

Intan Nurma Yulita^{1,*}

¹Program Studi S1 Teknik Informatika, Fakultas MIPA, Universitas Padjadjaran, Sumedang

¹Kampus UNPAD Jatinangor, Jalan Raya Bandung-Sumedang KM21 Jatinangor, 45363, Indonesia

email: ¹intan.nurma@unpad.ac.id

Abstract – This paper presented an implementation of the pattern recognition that related to the gesture phase. In particular, this paper focuses on pattern recognition for data with sequential form. The recognition can ignore the sequential factor, but it will degrade the accuracy to be obtained. Therefore, to overcome this challenge, then the paper proposed the use of Conditional Neural Fields (CNF). This method is a combination of Conditional Random Fields (CRF) and Artificial Neural Networks (ANN). The representation of ANN is presented in some gate on the middle layer of the CRF. This layer aims to map the nonlinear relationship between input and output in the data. As a result, it is obtained that the CNF is more effective and efficient than CRF based on the accuracy and the number of required iterations. However, the use of too many gates was ineffective due to the convergence of recognition models is not easy to achieve. On the other hand, if only one gate is used, then the convergence achieved but its accuracy is low. So it is necessary to find the optimal number of required gates.

Intisari > Pengenalan pola merupakan area informatika yang banyak dikaji hingga saat ini. Hal ini dikarenakan pemanfaatannya yang luas diterapkan dalam kehidupan sehari-hari. Di dalam makalah ini disajikan pengenalan pola untuk gerakan khususnya fase gerak. Secara khusus pengenalan fase gerak di dalam makalah ini menitik beratkan pada pengenalan pola pada data berbentuk sekuensial. Pengenalan ini dapat saja mengabaikan faktor sekuensialnya, namun tentu akan menurunkan akurasi yang akan diperoleh. Oleh karena itu untuk mengatasi tantangan tersebut, maka ditawarkan penggunaan *Conditional Neural Fields* (CNF). Metode ini merupakan gabungan antara *Conditional Random Fields* (CRF) dan *Artificial Neural Networks* (ANN). Representasi ANN disajikan dalam bentuk *gate* pada lapisan tengah dari CRF. Lapisan ini bertujuan untuk memetakan hubungan non-linear antara input dan output yang terdapat di dalam data. Sebagai hasilnya diperoleh bahwa CNF terbukti lebih efektif dan efisien dibandingkan CRF berdasarkan akurasi dan banyaknya iterasi yang dibutuhkan. Namun penggunaan terlalu banyak *gate* ternyata tidak efektif dikarenakan konvergensi dari model pengenalan semakin sulit tercapai. Di sisi lain, jika hanya satu *gate* yang digunakan maka konvergensi tercapai namun akurasi yang diperoleh rendah. Sehingga diperlukan upaya untuk menemukan banyaknya *gate* optimal yang diperlukan.

Kata Kunci > *Conditional Neural Fields, Pattern Recognition, Artificial Intelligence, Conditional Random Fields*

*) penulis korespondensi (Intan Nurma Yulita)
Email: intan.nurma@unpad.ac.id

I. PENDAHULUAN

Pengenalan pola (*pattern recognition*) banyak dikaji untuk penelitian dewasa ini. Pemanfaatan yang banyak dan luas digunakan menjadikan salah satu alasan atas hal ini. Salah satu kajian pengenalan pola yang banyak dibahas diantaranya adalah pengenalan pola dengan data berbentuk sekuensial.

Pola sekuensial dapat saja diproses dengan menggunakan metode yang mengabaikan faktor sekuensial yang terdapat dalam data. Namun tentunya dengan menjadikan faktor ini sebagai bagian dalam proses pembentukan model pengenalan pola. Metode yang banyak dikenal adalah *Hidden Markov Models* (HMM). Beberapa pemanfaatan HMM diantaranya adalah dalam dunia perikanan [1], klimatologi [2], kesehatan [3], dan biologi [4]. Namun metode ini mengalami kekurangan yaitu data disajikan hanya dalam bentuk satu dimensi. Sehingga Lafferty et.al mengusulkan penggunaan *Conditional Random Fields* (CRF) [5]. Tidak hanya permasalahan satu dimensi data saja yang teratasi, namun metode ini juga dapat menghindari permasalahan bias yang sering muncul di dalam metode bertipe *probabilistic graphical models* (PGM). Tetapi sayangnya CRF hanya terbatas untuk memetakan struktur eksternal dari data yang diinputkan sehingga hanya interaksi label yang menjadi acuannya dalam pembentukan model tiap label atau kelas. Dengan demikian, *Hidden-state Conditional Random Fields* dikaji untuk mengatasi struktur internal [6,7] dan akhirnya juga dibangun model yang dapat menggabungkan antara struktur eksternal dan internal yang juga sering disebut *Latent Dynamic Conditional Random Fields* [8]. Namun dari kesemua model tersebut masih belum memetakan hubungan *input* dan *output* dari data sebagaimana yang terjadi pada metode berjenis *Artificial Neural Networks* (ANN). Hal inilah yang mendasarkan Jian Peng et.al menggabungkan antara ANN dan CRF sebagai satu kesatuan model yang utuh yaitu menjadi CNF

Oleh karena itu, di dalam makalah ini disajikan pemaparan pemanfaatan CNF di dalam pengenalan pola khususnya pengenalan fase gerak dimana fase gerak yang diolah berasal dari video. Dengan adanya penelitian yang dibahas di dalam makalah ini, diharapkan metode ini dapat memberikan akurasi yang lebih baik di dalam mengenali data berbentuk sekuensial fase gerak terutama jika dibandingkan dengan CRF di dalam mengenali fase gerak dari data yang sama.

II. PENELITIAN YANG TERKAIT

Untuk dapat mencapai kinerja yang lebih baik dari HMM, Lafferty et al. mengusulkan penggunaan *Conditional Random Field* (CRF) untuk membagi dan melabelkan data sekuens namun dapat mengatasi keterbatasan yang dimiliki HMM [5]. CRF merupakan model *graf* berarah yang telah banyak diterapkan untuk melakukan segmentasi dan melabelkan urutan data yang muncul dalam berbagai domain. CRF dapat dilihat sebagai model diskriminatif yang mewakili probabilitas bersyarat $p(y/x)$ dengan diberikan urutan label tertentu Y dan urutan pengamatan X. Akibatnya, CRF mampu menggabungkan fitur kompleks dari urutan pengamatan X yang tidak memerlukan asumsi ketidak-tergantungan antara variabel pengamatan di X. Tidak seperti HMM, CRF mengakomodasi data observasi vektor dan mampu memecahkan masalah-label bias yang ada secara inheren dalam HMM. Label bias yang didefinisikan oleh Lafferty et.al terjadi karena transisi yang meninggalkan state hanya berkompetisi satu sama lainnya bukan melawan semua transisi lainnya dalam model.

Pengaruh label-bias, dalam kasus yang ekstrim, adalah bahwa state dengan transisi tunggal yang keluar secara efektif mengabaikan faktor pengamatan. Masalah yang umum yang terjadi pada model probabilistik seperti HMM atau *Maximum Entropy Markov Model* (MEMM) yaitu state memiliki beberapa state penggantinya.

CRF mengasumsikan bahwa input X dan output Y adalah *random vectors*, $\theta = \{\theta_k\}$ menjadi *real parameter vectors*, dan $\{f_k(y, y', x_t)\}_{k=1}^K$ menjadi kumpulan fungsi fitur yang bernilai real.

$$p(y|x) = \frac{1}{Z(x)} \prod_{t=1}^T \exp \left\{ \sum_{k=1}^K \theta_k f_k(y_t, y_{t-1}, x_t) \right\} \quad (1)$$

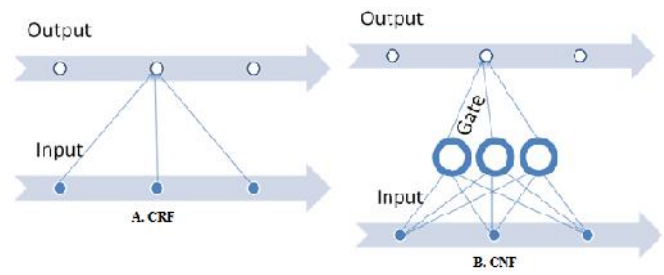
Dimana Z(x) adalah fungsi normalisasi.

$$Z(x) = \sum_y \prod_{t=1}^T \exp \left\{ \sum_{k=1}^K \theta_k f_k(y_t, y_{t-1}, x_t) \right\} \quad (2)$$

dan $f_k(y, y', x_t)$ adalah fungsi fitur dari posisi saat ini (j). θ adalah bobot dari fungsi fitur

Fungsi fitur diperoleh dengan mendaftarkan semua kemungkinan yang mungkin dari label-label yang berdekatan sedangkan bobot dari fitur tersebut didapatkan melalui proses pembelajaran dari data. Untuk mencapai konvergensi, maka umumnya CRF ini dilatih dengan menggunakan *Conjugative Gradient* ataupun *Broyden-Fletcher-Goldfarb-Shanno Algorithm*.

Namun CRF tidak mampu memetakan hubungan non-linear antara input dan output. Sehingga CNF yang menjadi kombinasi antara CRF dan ANN memetakannya melalui penambahan lapisan *gate* [9]. Lapisan ini menjadi perbedaan mendasar antara CRF dan CNF sebagaimana terlihat pada gambar 1.



Gbr. 1 A.CRF dan B. CNF

$$p(y|x) = \frac{1}{Z(x)} \prod_{t=1}^T \exp \left\{ \sum_{k=1}^K \sum_{g=1}^G \Psi(\alpha_g f_k(y_t, y_{t-1}, x_t)) \right\} \quad (3)$$

Dimana

Ψ adalah fungsi *gate*

α_g adalah bobot fungsi *gate*

III. METODE PENELITIAN

A. Dataset

Untuk melihat performansi dari metode yang digunakan maka penelitian ini diujikan dengan menggunakan dataset fase gerak yang berasal dari *School of Art, Sciences and Humanities University of Sao Paulo* [10]. Data tersebut dapat diunduh di https://archive.ics.uci.edu/ml/machine-learning-databases/00302/gesture_phase_dataset.zip. Terdapat tiga video yang digunakan untuk penelitian ini dimana setiap video tersusun oleh frame-frame dan untuk setiap *frame* telah diekstraksi menjadi sejumlah fitur. Rincian fitur ditampilkan pada tabel 1.

B. Alur Penelitian

Alur dalam penelitian ini dijelaskan di dalam gambar 2. Dari 20 fitur yang terdapat di dalam data, maka fitur *timestamp* dihapus dan fase menjadi label *output* dalam sistem pengenalan. 18 fitur sisanya diproses (*preprocessing*). Jika pada awalnya sebuah posisi memiliki tiga buah koordinat maka fitur tersebut akan diubah menjadi fitur scalar dengan menghitung jarak ketiga koordinat tersebut terhadap pusat koordinat. Perhitungan tersebut didasarkan pada rumus *Euclidean Distance* yaitu:

$$d_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \quad (4)$$

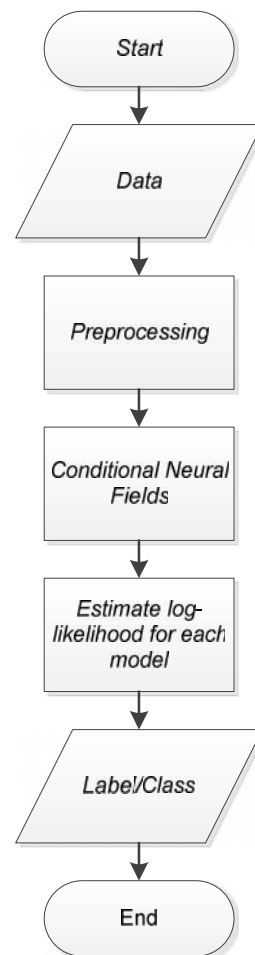
Dimana i adalah posisi

Dari *preprocessing* tersebut maka input dan output dari sistem pengenalan ini menjadi hanya enam input dan satu output. Hasil *preprocessing* disajikan pada tabel 2.

TABEL I
DAFTAR FITUR AWAL

ID	Fitur
lhx	Posisi tangan kiri (koordinat x)
lhy	Posisi tangan kiri (koordinat y)
lhz	Posisi tangan kiri (koordinat z)
rhx	Posisi tangan kanan (koordinat x)
rhy	Posisi tangan kanan (koordinat y)
rhz	Posisi tangan kanan (koordinat z)
hx	Posisi kepala (koordinat x)
hy	Posisi kepala (koordinat y)
hz	Posisi kepala (koordinat z)
sx	Posisi tulang (koordinat x)
sy	Posisi tulang (koordinat y)
sz	Posisi tulang (koordinat z)
lwx	Posisi pergelangan tangan kiri (koordinat x)
lwy	Posisi pergelangan tangan kiri (koordinat y)
lwz	Posisi pergelangan tangan kiri (koordinat z)
rxw	Posisi pergelangan tangan kanan (koordinat x)
rwy	Posisi pergelangan tangan kanan (koordinat y)
rwz	Posisi pergelangan tangan kanan (koordinat z)
Timestamp	Waktu kemunculan fase
Fase	-- Rest (1) -- Preparation (2) -- Stroke (3) -- Hold (4) -- Retraction (5)

Dengan demikian maka *input* dari sistem pengenalan terdiri atas enam fitur yaitu lh, rh, h, s, lw, dan rw. Selanjutnya urutan input atas keenam fitur digunakan untuk membangun model dengan menggunakan CNF. Hasil dari proses ini adalah model-model peluang untuk setiap fase gerak.



Gbr. 2 Flowchart

Penentuan label dari data dengan mengestimasi kemiripan / *likelihood* dari data terhadap model-model tersebut. Model yang memiliki peluang tertinggi akan menjadi label / kelas (*class*) prediksi.

TABEL II
DAFTAR FITUR SETELAH PREPROCESSING

Fitur	Keterangan
lh	Posisi tangan kiri
rh	Posisi tangan kanan
h	Posisi kepala
s	Posisi tulang
lw	Posisi pergelangan tangan kiri
rw	Posisi pergelangan tangan kanan
Fase	-- Rest (1) -- Preparation (2) -- Stroke (3) -- Hold (4) -- Retraction (5)

Di dalam pembangunan setiap model, mekanisme konvergensi dilakukan dengan menggunakan *Broyden-Fletcher-Goldfarb-Shanno Algorithm* di dalam 1000 iterasi.

C. Evaluasi

Dua dari tiga data yang terdapat dalam dataset digunakan sebagai data latih sedangkan sisanya sebagai data uji. Label prediksi yang diperoleh dari sistem pengenalan akan diuji berdasarkan performansi akurasi, banyaknya iterasi yang dibutuhkan, dan waktu eksekusi. Perhitungan akurasi berdasarkan rumus berikut:

$$Akurasi = \frac{Banyaknya\ prediksi\ yang\ benar}{Banyaknya\ data\ uji} \cdot 100\% \quad (5)$$

IV. HASIL DAN PEMBAHASAN

Performansi CNF diuji berdasarkan penggunaan jumlah *gate* yang digunakan. Analisis performansi dengan memperhatikan aspek akurasi, banyaknya iterasi yang dibutuhkan sehingga mencapai konvergensi, serta waktu eksekusi.

A. Pengaruh Gate

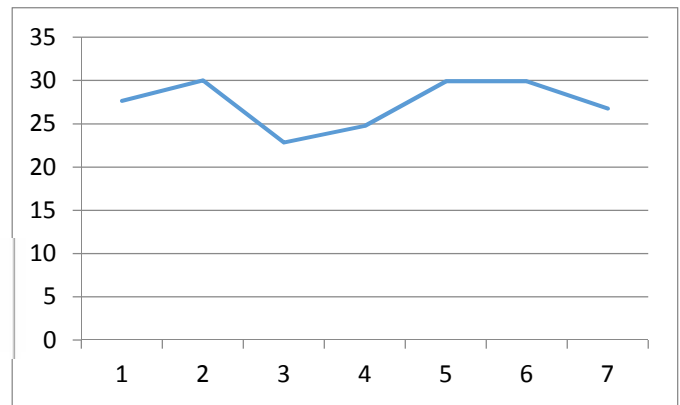
Perbedaan penting antara CRF dan CNF terletak pada *gate* yang dimiliki oleh CNF. Oleh karena itu seberapa besar pengaruh banyaknya *gate* terhadap performansi ditampilkan pada tabel 3.

Dari tabel, dapat diketahui bahwa performansi terbaik diperoleh ketika banyaknya *gate* yang digunakan adalah tiga buah dengan akurasi yang didapatkan adalah 30.05 %. Namun dari percobaan dengan menguji penggunaan satu hingga tujuh *gate* ini, diperoleh bahwa penggunaan tiga *gate* mencapai kondisi terburuk. Namun akurasi kecilnya bisa disebabkan karena banyaknya iterasi yang dijalankan telah mencapai batas iterasi maksimum oleh program yaitu sebesar 1000 iterasi sehingga belum tercapai konvergensi.

TABEL III
PERBANDINGAN PERFORMANSI TIAP GATE

n(Gate)	n(Iterasi)	Akurasi (%)	Waktu (detik)
1	361	27.65	8.29
2	428	30.05	9.48
3	1000	22.84	23.72
4	798	24.79	20.06
5	1000	29.94	27.15
6	1000	29.94	31.19
7	1000	26.79	37.06

Berdasarkan dari percobaan yang telah dilakukan, maka *trend* yang terjadi disajikan pada gambar 3.



Gbr. 3 Akurasi tiap *gate* dari CNF. Sumbu X adalah banyaknya *gate* sedangkan sumbu Y adalah persentase akurasi.

Pada gambar 3, akurasi dari satu *gate* adalah 27.65 %. Meskipun akurasi kecil namun konvergensi telah tercapai. Selain itu, penggunaan dua *gate* mampu menaikkan akurasi hingga mencapai 30.05 %. Namun penggunaan 3 atau 4 *gate* justru menurunkan akurasi. Selanjutnya konvergensi belum tercapai maksimal pada penambahan *gate* lagi dikarenakan iterasi yang digunakan terhenti pada banyaknya iterasi maksimum pada program. Hal ini menyiratkan bahwa penggunaan banyak *gate* ternyata tidak efektif untuk memetakan hubungan non-linear antara input dan output untuk proses pengenalan fase gerak.

Di sisi lain, semakin banyak *gate* yang digunakan maka CNF memerlukan lebih banyak iterasi untuk mencapai konvergensi, dengan kata lain maka waktu eksekusi yang dibutuhkan menjadi semakin banyak.

B. Perbandingan antara CRF dan CNF

Pengaruh kombinasi antara ANN dan CRF disajikan pada tabel dengan membandingkan antara hasil yang diperoleh CRF dan CNF. CNF yang digunakan sebagai pembanding adalah CNF dengan menggunakan dua *gate*.

TABEL IV
PERBANDINGAN PERFORMANSI CNF DAN CRF

Metode	n(Iterasi)	Akurasi (%)	Waktu (detik)
CRF	428	29.08	12.56
CNF	133	30.05	9.48

Berdasarkan tabel 4, diketahui bahwa akurasi CNF lebih tinggi dibandingkan CRF namun banyaknya iterasi yang dibutuhkan CNF menjadi tiga kali lebih sedikit dibandingkan CRF. Hal ini mengindikasikan bahwa penggunaan *gate* terbukti efektif di dalam proses pengenalan fase gerak. Dengan adanya *gate*, maka hubungan non-linear antara input dan output terpetakan sehingga *gate* mampu merepresentasikan sebagai fitur bentukan baru yang dapat mengefektifkan sistem untuk membentuk model-model yang memiliki karakteristik yang berbeda satu dengan lainnya.

Dengan banyaknya iterasi yang dibutuhkan untuk proses pemodelan lebih sedikit maka waktu eksekusi CNF juga lebih sedikit dibandingkan CRF.

V. KESIMPULAN

Dari percobaan yang telah dilakukan dengan menguji banyaknya *gate* dari CNF serta membandingkan hasilnya dengan CRF maka diperoleh bahwa:

1. Penggunaan *gate* pada CNF terbukti lebih efektif dibandingkan CRF dikarenakan akurasinya lebih tinggi.
2. Penggunaan terlalu banyak *gate* menjadi tidak efektif dikarenakan sistem lebih susah menemukan konvergensi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada LPDP, Machine Learning and Computer Vision Laboratory di Universitas Indonesia, dan juga Departemen Ilmu Komputer Universitas Padjadjaran yang telah memberi dukungan atas riset yang dilakukan.

DAFTAR PUSTAKA

- [1] C. Spampinato, S. Palazzo. 2012. Hidden Markov Models for Detecting Anomalous Fish Trajectories in Underwater Footage. *2012 IEEE International Workshop on Machine Learning for Signal Processing*, Santander, Spain.
- [2] Lambert, M.F., Whiting, J.P. and Metcalfe, A.V., 2003. A non-parametric hidden Markov model for climate state identification. *Hydrology and Earth System Sciences Discussions*, 7(5), pp.652-667.
- [3] Cooper, Ben., and Lipsitch, Marc., 2004. The Analysis of Hospital Infection Data Using Hidden Markov Models. *Biostatistics*, vol 5, no 2, Pp.223-237.
- [4] Yoon, Byung-Jun. 2009. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current Genomics*, vol.10 page 402-415
- [5] Lafferty, J., McCallum, A. and Pereira, F., 2001, June. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the eighteenth international conference on machine learning, ICML* (Vol. 1, pp. 282-289).
- [6] Gunawardana, A., Mahajan, M., Acero, A. and Platt, J.C., 2005, September. Hidden conditional random fields for phone classification. In *INTERSPEECH* (pp. 1117-1120).
- [7] Quattoni, A., Wang, S., Morency, L.P., Collins, M. and Darrell, T., 2007. Hidden conditional random fields. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(10), pp.1848-1852.
- [8] Zhang, S., 2012. Fuzzy-based latent-dynamic conditional random fields for continuous gesture recognition. *Optical Engineering*, 51(6), p.067202.
- [9] Peng, J., Bo, L. and Xu, J., 2009. Conditional neural fields. In *Advances in neural information processing systems* (pp. 1419-1427).
- [10] Madeo, R.C.B., Wagner, P.K. and Peres, S.M., 2013. A review of temporal aspects of hand gesture analysis applied to discourse analysis and natural conversation. *International Journal of Computer Science and Information Technology*, 5(4), pp.1-20.