

Aplikasi *Deep Learning* dalam Berbagai Domain : Review Paper

Firmansyah
Balai Pendidikan dan Pelatihan
Tambang Bawah Tanah
Sawahlunto

Yohanes Gultom
Faculty Of Computer Science
University Of Indonesia
Depok, Indonesia
Yohanes.gultom@ui.ac.id

Abstrak

Deep learning berkembang pesat beberapa tahun terakhir karena kemampuannya untuk mempelajari representasi data yang kemudian digunakan untuk deteksi atau klasifikasi. Makalah ini berisi rangkuman dari beberapa contoh aplikasi *deep learning* dari berbagai domain, yaitu *image classification*, *speech recognition*, penerjemahan bahasa, peringkasan multi dokumen, *drug-target interactions (DTI) predictions* dan *time-series data predictions*. Tujuan dari rangkuman ini adalah memberi gambaran mengenai strategi penerapan *deep learning* pada berbagai domain yang diharapkan dapat memberikan pemahaman yang lebih baik mengenai *deep learning* dan membantu penelitian pada domain serupa atau lainnya.

1 Latar Belakang

Deep learning adalah varian dari pembelajaran mesin yang berbasis jaringan syaraf tiruan dengan banyak *hidden layers* yang memiliki kemampuan untuk mempelajari representasi atau fitur data secara otomatis [1]. Sebelum *deep learning* dikenal (di akhir era tahun 90), fitur dari sebuah pembelajaran mesin dalam domain tertentu harus dianalisis dan ditentukan oleh perekayasanya dan dintegrasikan dalam algoritma pembelajaran karena penemuan fitur secara otomatis hanya dengan sedikit *prior knowledge* dianggap tidak mungkin [1].

Seperti jaringan syaraf tiruan lainnya, secara umum arsitektur *deep learning* terdiri dari *visible* dan *hidden layers* di mana bobot (*weight*) dari tiap unit *perceptron* nya dioptimalkan menggunakan algoritma *backpropagation* [1]. Tetapi dalam aplikasinya, *deep learning* memiliki banyak varian dalam hal implementasi arsitektur (jenis nilai unit, jumlah *hidden layers*, relasi antar unit dan antar *layer* .dsb) dan algoritmanya (aktivasi unit, pemutakhiran bobot .dsb). Contoh algoritma yang menerapkan *deep learning*

antara lain *Deep Convolutional Neural Networks* (DCNN) untuk *image classification* [2], *Deep Belief Network - Deep Neural Network* (DBN - DNN) untuk *speech recognition* [3], *Recurrent Neural Network* (RNN) untuk penerjemahan bahasa [4], *Query-Oriented Deep Extraction* (QODE) yang berbasis *Restricted Boltzmann Machine* (RBM) untuk peringkasan multi dokumen [5], *Conditional Restricted Boltzmann Machine* (RBM) untuk *Drug-Target Interaction (DTI) prediction* [6] dan *Deep Belief Network* (DBN) untuk *time-series data prediction* [7].

2 Aplikasi Deep Learning

2.1 Image Classification

Sampai saat ini permasalahan *image classification* diselesaikan dengan menggunakan pendekatan pembelajaran mesin (*machine learning*) di mana peningkatan kualitas klasifikasi dapat dilakukan dengan cara menambah jumlah *training dataset*, menggunakan model yang lebih mangkus dan menggunakan teknik-teknik untuk mencegah *overfitting*. Ketiga cara tersebut saling berhubungan sama lain karena ketika jumlah *training dataset* dibuat menjadi lebih besar maka model yang digunakan juga harus memiliki kapasitas yang besar dan teknik pencegahan *overfitting* yang digunakan juga perlu disesuaikan dengan kapasitas data yang dipelajari. Selain itu, seiring dengan meningkatnya ukuran *dataset* dibutuhkan kinerja perangkat keras yang lebih baik untuk menjaga durasi *training* pada batas yang masuk akal. [2]

Deep Convolutional Neural Network (DCNN) menggunakan delapan buah layers yang terdiri dari *Rectified Linear Units* (ReLU) dan teknik *dropout* yang dijalankan menggunakan dua buah *Graphical Processing Unit* (GPU) telah mampu memecahkan rekor klasifikasi citra terhadap *dataset ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2010* (1,2 juta citra resolusi tinggi dari 1000 kelas) dengan kesalahan rata-rata 37,5% (top-1) dan 17% (top-5) [2]. Di mana top-1 adalah penilaian klasifikasi di mana hasil dianggap valid ketika satu buah kelas yang dipilih tepat sama. Sedangkan top-5 adalah penilaian klasifikasi di mana hasil dianggap valid ketika salah satu dari lima buah kemungkinan kelas yang dipilih sama dengan kelas yang diharapkan. Proses training DCNN ini menghabiskan lima sampai enam hari.

Evaluasi kualitatif terhadap hasil eksperimen yang dilakukan juga menunjukkan bahwa top-5 kemungkinan kelas yang dipilih oleh DCNN memiliki kemiripan karakter (contoh: *leopard, jaguar, cheetah, snow leopard, egyptian cat*). Hal ini menunjukkan bahwa DCNN berhasil mempelajari abstraksi yang valid dari dataset yang digunakan untuk *training*. [2]

2.2 Speech Recognition

Hidden Markov Models yang dikombinasikan dengan *Gaussian Mixture Models* (HMM-GMM) adalah

model yang sangat sukses untuk pemodelan masalah *real-world speech recognition*. Kesuksesan HMM-GMM membuatnya sulit ditandingi oleh algoritma-algoritma baru untuk menyelesaikan masalah yang sama. Satu-satunya kelemahan yang cukup serius adalah HMM-GMM kurang efisien untuk memodelkan data yang terletak atau dekat dengan *non-linear manifold* pada ruang data (contoh: data berupa titik-titik yang berada dekat dengan permukaan bola (*sphere*)). [3]

Deep Belief Network - Deep Neural Network (DBN-DNN) adalah arsitektur baru yang merupakan kombinasi dari dua buah arsitektur (DBN dan DNN) di mana arsitektur ini merupakan kombinasi tiga buah *Gaussian-Bernoulli Restricted Boltzmann Machines (GRBMs)* yang koneksi antar *layer* nya dibuat *feed-forward* mengikuti karakteristik DNN. Saat dilatih dan diuji menggunakan dataset *Texas Instruments Massachusetts Institute of Technology (TIMIT)*, arsitektur *deep learning* ini ternyata mampu mengalahkan HMM-GMM bahkan dengan *margin* besar di beberapa pengujian. [3]

2.3 Penerjemahan Bahasa

Pada domain penerjemahan bahasa statistikal atau *Statistical Machine Translation (SMT)*, pemodelan bahasa menggunakan arsitektur *deep learning RNN Encoder-Decoder* diajukan sebagai pendekatan baru untuk meningkatkan kinerja pendekatan yang sudah ada yaitu *Continuous Space Language Modeling (CSLM)*. Sedangkan algoritma sebagai *baseline* perbandingan digunakan kaskas Moses dengan konfigurasi *default*. [4]

Arsitektur *RNN Encoder-Decoder* ini terdiri dari dua buah RNN yang memiliki fungsi yang berbeda: *encoder* dan *decoder*. *RNN Encoder* bertugas melakukan *encoding* simbol sekuensial pada frase yang akan diterjemahkan menjadi bentuk vektor untuk dijadikan input *deep learning*. Sedangkan *RNN Decoder* mengembalikan vektor output dari *deep learning* menjadi simbol sekuensial kembali. Selain itu arsitektur ini juga menggunakan jenis *hidden units* berbeda yaitu menggunakan *activation function* yang mirip dengan *Long-Short Term Memory (LSTM)* dan juga memiliki *update gate* dan *reset gate* yang berfungsi untuk membuang (*drop*) yang kurang relevan dalam proses pembelajaran. [4]

Evaluasi kuantitatif terhadap eksperimen yang dilakukan menunjukkan bahwa *RNN Encoder-Decoder* sendiri mencapai nilai BLEU yang lebih baik dibandingkan dengan *baseline*. Nilai yang lebih baik dicapai ketika *RNN Encoder-Decoder* digabungkan dengan CSLM dan teknik *Work Penalty*. [4]

Evaluasi kualitatif terhadap eksperimen yang dilakukan menunjukkan bahwa *RNN Encoder-Decoder* ini mampu memilih terjemahan frase yang sama dengan translasi manual/literal pada hampir semua kasus. [4]

2. 4 Peringkasan Multi Dokumen

Salah satu kunci dari peringkasan multi dokumen adalah memahami proses komposisi kata-kata menjadi kalimat-kalimat yang menyusun tiap dokumen. Dengan pemahaman terhadap proses tersebut, peringkasan multi dokumen dapat dilakukan secara ekstraktif yaitu dengan memilih sejumlah kalimat yang mewakili seluruh dokumen dengan topik yang sama. Oleh karena itu *deep learning* dapat digunakan untuk menyelesaikan masalah ini dengan kemampuannya untuk mempelajari komposisi secara otomatis. [5]

Solusi baru yang berbasis *deep learning* yang ditawarkan untuk menyelesaikan masalah peringkasan multi dokumen secara ekstraktif ini adalah *Query-Oriented Deep Extraction* (QODE). QODE merupakan arsitektur *unsupervised deep learning* yang menggunakan *stacked Restricted Boltzmann Machines (RBMs)* dikombinasikan dengan teknik menggunakan *user query* sebagai basis untuk menentukan bobot jaringan awal antara *input layer* dengan *hidden layer* pertama dan *dynamic programming* untuk memilih kalimat-kalimat yang akan dijadikan ringkasan akhir. Sedangkan bobot pada jaringan disesuaikan dengan algoritma *backpropagation* dengan *cross-entropy error*. [5]

Hasil eksperimen yang diperoleh adalah QODE mampu memberikan nilai ROUGE yang lebih baik daripada pendekatan peringkasan dokumen lainnya pada hampir semua pengujian. Tetapi pada beberapa kasus pengujian nilai yang dicapai QODE masih di bawah *existing approach* yaitu Graph Based (dataset DUC 2005 dan 2006) & Ranking SVM (dataset DUC 2007). Penulis berkesimpulan bahwa kinerja yang lebih baik dapat dicapai jika QODE diintegrasikan dengan *supervised learning*. [5]

2.5 Drug-Target Interaction Prediction

Salah satu contoh aplikasi *deep learning* pada bidang bioinformatika adalah pada prediksi efek obat atau *Drug Target Interaction* (DTI). Prediksi ini dapat digunakan untuk obat yang baru atau sudah lama tidak digunakan untuk meningkatkan efisiensi eksperimen langsung dengan obat dan protein yang notabene sulit dilakukan berulang-ulang dan menghabiskan biaya yang tidak sedikit. [6]

Arsitektur *deep learning* yang diajukan untuk prediksi DTI ini adalah *Conditional Restricted Boltzmann Machine* (RBM) dengan dua *hidden layers* yang dioptimasi bobot jaringannya dengan *Contrastive Divergence* (CD). Arsitektur ini di-training secara *supervised* menggunakan dataset MATADOR dan STITCH yang berisi informasi mengenai interaksi antara obat dan protein yang sudah diketahui dan *mode of action* dari tiap interaksi. Kontribusi penelitian ini dibanding penelitian terkait lainnya adalah mempertimbangkan juga *mode of action* dari tiap interaksi. [6]

Hasil eksperimen yang dilakukan menunjukkan bahwa algoritma ini mencapai akurasi prediksi (*precision-recall*) 89.6% yang termasuk cukup akurat dan lebih tinggi dari *simple logic approach* yang digunakan sebagai pembanding. Hasil eksperimen lain yang dilakukan juga menemukan bahwa hasil prediksi akan meningkat akurasinya ketika dimensi interaksi digabungkan (*mixed*) sebagai input dari RBM (mempertimbangkan *mode of action* juga). [6]

2.6 Time-Series Data Prediction

Jaringan syaraf tiruan telah lama menjadi solusi pilihan untuk memprediksi nilai masa depan dari sebuah *time-series data* (contoh: *stock market index, foreign exchange rates, electricity consumption rate* .dsb). Tapi ada beberapa masalah yang dialami oleh *existing approach*, yaitu *overfitting, initial weights configuration* dan *optimization learning parameters*. [7]

Sebagai salah satu opsi solusi dari masalah yang ada, penulis mengajukan arsitektur *Deep Belief Network* (DBN) yang terdiri dari 3 *layers* dengan *Particle Swarm Optimization* (PSO) untuk mencari nilai optimum dari *learning parameters* (jumlah unit pada *visible* dan *hidden layers, learning rate*). Jaringan DBN dioptimasi menggunakan *descent probabilities* dari fungsi energi aktivasi. [7]

Eksperimen dilakukan menggunakan dataset *Competition on Artificial Time Series* (CATS) untuk membandingkan akurasi prediksi *short term* dan *long term* DBN dengan algoritma *Multi Layer Perceptron* (MLP) dan *Autoregressive Moving Average* (ARIMA). Hasil yang diperoleh menunjukkan bahwa DBN memberikan hasil prediksi *short term* dan *long term* yang lebih baik daripada MLP dan ARIMA diukur menggunakan *Lorentz chaos* dan *logistic map*. Tapi DBN belum mampu memecahkan rekor prediksi CATS yang saat ini dipegang oleh algoritma *Kalman Smoother Model* dan *Ensemble Models*. [7]

3 Analisis dan Kesimpulan

Perbandingan aplikasi *deep learning* pada berbagai domain yang dikaji dapat dilihat pada Tabel 1.

No	Domain	Jenis Data	Arsitektur	Deskripsi Arsitektur	Hasil Eksperimen
2.1	<i>Image classification</i>	Foto	<i>Deep Convolutional Neural Networks (DCNN)</i>	CNN dengan 8 <i>layers</i> berisi ReLUs dan menggunakan teknik <i>dropout</i> tanpa <i>pre-training</i>	DCNN lebih baik dari pendekatan-pendekatan yang ada

2.2	<i>Speech recognition</i>	<i>Speech audio</i>	<i>Deep Belief Network - Deep Neural Network (DBN-DNN)</i>	Feedforward DBN dengan <i>pre-training</i>	DBN-DNN lebih baik pendekatan yang ada
2.3	<i>Penerjemahan bahasa</i>	Teks frase	<i>Recurrent Neural Network (RNN) Encoder - Decoder</i>	2 buah RNN (<i>encoder & decoder</i>) dengan <i>LSTM-like hidden unit</i>	RNN menunjukkan peningkatan dibanding pendekatan yang ada secara ortogonal
2.4	Peringkasan multi dokumen	Teks (kumpulan dokumen)	<i>Query Oriented Deep Extraction (QODE)</i>	<i>Unsupervised stacked Gaussian RBM</i> dengan 3 <i>hidden layers</i>	QODE lebih baik dari hampir semua algoritma kecuali Graph Based dan Ranking SVM dalam beberapa kasus
2.5	<i>Drug-Target Interaction (DTI) Prediction</i>	Relasi obat - protein	<i>Conditional Restricted Boltzmann Machines (RBMs)</i>	2 layers of <i>undirected conditional RBM</i> dengan <i>Contrastive Divergence (CD)</i>	Akurasi prediksi (<i>precision-recall</i>) dari algoritma ini mencapai 89.6% yang lebih baik dari <i>simple logic approach</i>
2.6	<i>Time-series data prediction</i>	Angka historikal	<i>Deep Belief Network (DBN)</i>	DBN dengan 3 layers dengan <i>Particle Swarm Optimization (PSO)</i> untuk optimasi <i>learning parameters</i>	Akurasi prediksi DBN lebih tinggi dari algoritma konvensional MLP dan ARIMA tetapi belum dapat mengalahkan rekor <i>CATS benchmark</i> yaitu <i>Kalman Smoother Model & Ensemble Models</i>

Tabel 1 Perbandingan aplikasi deep learning

Di seluruh domain yang dikaji dalam makalah ini, *deep learning* menunjukkan kinerja yang lebih baik daripada *existing* atau *state-of-the-art approach* untuk tiap domain. Bahkan pada domain *image classification* [2] dan *speech recognition* [3], *deep learning* mencapai nilai evaluasi dengan margin yang cukup jauh dengan pendekatan lainnya.

Hampir semua domain yang dikaji menggunakan arsitektur *deep learning* yang menggunakan *Restricted Boltzmann Machine (RBM)* sebagai fondasinya [3][4][5][6][7]. Hanya domain *Image Classification* yang tidak menggunakan *RBM* sebagai fondasi [1]. Hal ini menunjukkan bahwa *RBM* merupakan implementasi arsitektur yang paling generik dan dapat digunakan untuk berbagai domain.

Selain evaluasi kuantitatif yang umum dilakukan pada penelitian ilmu komputer, karakteristik *deep learning* di mana proses pembelajaran pada *hidden layers* dapat diamati perkembangannya, ternyata memungkinkan dilakukannya evaluasi kualitatif pada proses pembelajaran tersebut. Penelitian pada domain *image classification* [2] dan penerjemahan bahasa [4] menggunakan evaluasi kualitatif tersebut untuk membuktikan validitas proses pembelajaran yang terjadi di masing-masing arsitektur.

Arsitektur dari *DCNN* pada domain *image classification* menggunakan teknik *dropout* untuk membuang fitur representasi yang dinilai kurang relevan [2]. Sekalipun tidak menggunakan istilah dan mekanisme yang persis sama, arsitektur *RNN Encoder Decoder* pada domain penerjemahan bahasa juga menggunakan jenis *hidden units* yang membuang fitur representasi yang kurang relevan [4]. Berdasarkan kedua penelitian ini dapat disimpulkan bahwa teknik *dropout* ini dapat diterapkan untuk mempercepat pembelajaran terlepas dari domain aplikasinya.

Algoritma *unsupervised deep learning*, sama seperti *unsupervised learning* pada umumnya, memang sudah dapat memberikan hasil yang memadai untuk domain yang cocok. Tetapi berdasarkan kesimpulan penelitian mengenai *deep learning* untuk peringkasan multi dokumen, kombinasi dengan *supervised learning* akan dapat meningkatkan kinerja [5]. Hal ini didukung oleh penjelasan mengenai pencetus *deep learning* yang menjelaskan bahwa salah satu kelebihan *deep learning* adalah menemukan representasi data yang lebih akurat dan dengan efisien menggunakan *labeled data* atau disebut juga *supervised learning* [1].

4 Referensi

1. Hinton, G., Bengio, Y., LeCun, Y. (2015). Deep learning review. Nature vol 521 28 May 2015 436-444
2. Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems 25 1090–1098 (2012).
3. Hinton, G. et al. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine 29, 82–97 (2012).
4. Cho, K. et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proc. Conference on Empirical Methods in Natural Language Processing 1724–1734 (2014).
5. Zhong, S. et al. (2015). Query-oriented multi-document summarization via unsupervised deep learning. Expert Systems with Applications 42 (2015) 8146–8155.
6. Zeng, J., Wang, Y. (2013). Predicting drug-target interactions using restricted Boltzmann machines. Bioinformatics Vol. 29 ISMB/ECCB 2013, pages i126–i134.
7. Kuremoto T. et al. (2014). Time series forecasting using a deep belief network with restricted Boltzmann machines. Neurocomputing 137 (2014) 47–56