

ESTIMASI DATA HILANG PADA JARINGAN TRAFIK INTERNET MENGUNAKAN METODE SRSVD

¹Ary Indah Ivrihanita, ²Syamsul El Yumin²

Magister Teknik Elektro, Institut Sains Teknologi Nasional, ISTN Jakarta
Jalan Moh. Kahfi II, Jagakarsa, Jakarta Selatan.

¹ email: aryi891@gmail.com

² email: sel_yumin@yahoo.com

ABSTRAK

Data lalu lintas Jaringan terdiri dari Traffic Matrices (TMs), yang mewakili volume lalu lintas antara pasangan Asal dan Tujuan (OD) dalam jaringan yang merupakan parameter input kunci dari tugas-tugas teknik jaringan. Kerangka estimasi pergerakan yang besar dapat mengalami efek buruk dari kesalahan atau informasi yang hilang. Compressive sensing merupakan filosofi untuk mengelola informasi yang hilang yang menggunakan kedekatan jenis struktur tertentu dan kelebihan informasi dari berbagai kerangka asli. Untuk mengelola informasi yang hilang yang menggunakan kedekatan jenis struktur tertentu dan kelebihan informasi dari berbagai kerangka asli. Dalam penelitian sebelumnya, metode penyisipan yang diusulkan untuk secara tepat membuat ulang kualitas yang hilang di TM dalam kaitannya dengan estimasi tak lengkap. Dalam penelitian ini, meskipun banyak kemajuan terbaru dalam bidang compressive sensing, dengan mengembangkan sparsity regularized SVD (SRSVD) menggunakan ℓ_2 -optimasi, teknik norma yang menemukan peringkat rendah perkiraan dari TM yang menjelaskan sifat spasial TM nyata. Berdasarkan yang dapat digunakan untuk mencari solusi SPL adalah konsisten dan solusi terbaik untuk mendekati SPL tidak konsisten dan SRSVD dapat digunakan untuk menemukan invers pseudo dan peringkat matriks. Hasil analisis algoritma yang digunakan, penulis mampu melakukan rekonstruksi hingga 98% dengan NMSE 3×10^{-3} lebih baik dari metode lain yang biasa digunakan dalam proses interpolasi.

Kata Kunci : Lalu Lintas, Data Hilang, Estimasi, Compressive Sensing, Sparsity Regularized SVD (SRSVD).

ABSTRACT

Arrange activity information comprises of Traffic Matrix (TM), which speaks to the volumes of movement amongst Origin and Destination (OD) combines in the system. Indeed, even great movement estimation frameworks can experience the ill effects of blunders or missing information. Compressive sensing is a bland philosophy for managing missing information that use the nearness of specific sorts of structure and excess in information from numerous genuine frameworks. In past research, the proposed insertion methods to precisely remake missing qualities in TM in light of incomplete and roundabout estimations. In this exploration, in spite of much late advance in the range of compressive sensing, with creating Sparsity Regularized SVD (SRSVD) utilizing ℓ_2 -enhancement standard system, which discovers low-rank approximations of TM that record for spatial properties of genuine TM. Based that can be utilized to discover arrangements of SPL is steady and best answers for approach the SPL is conflicting and SRSVD can be utilized to locate the pseudo reverse and rank of a network. The consequences of investigation the calculations utilized, creator can do recreation to 98% with NMSE 3×10^{-3} superior to anything different strategies ordinarily utilized as a part of the interjection procedure.

Keyword—Traffic, Missing Data, Estimation, Compressive Sensing, Sparsity Regularized SVD (SRSVD).

PENDAHULUAN

Lalu lintas data dari jaringan penting untuk melakukan manajemen jaringan yang lebih baik. Ini terdiri dari data Traffic Matrix (TM), yang mewakili volume lalu lintas antara pasangan Asal dan Tujuan (OD) dalam jaringan [1]. Compressive sensing adalah metodologi umum untuk menangani jenis struktur tertentu dan redundansi dalam data dari banyak sistem dunia nyata. Dalam penelitian sebelumnya, teknik interpolasi yang diusulkan untuk secara akurat merekonstruksi data yang hilang pada pengukuran parsial dan tidak langsung.

Seminar Nasional Riset dan Teknologi, Jakarta 13 Oktober 2018

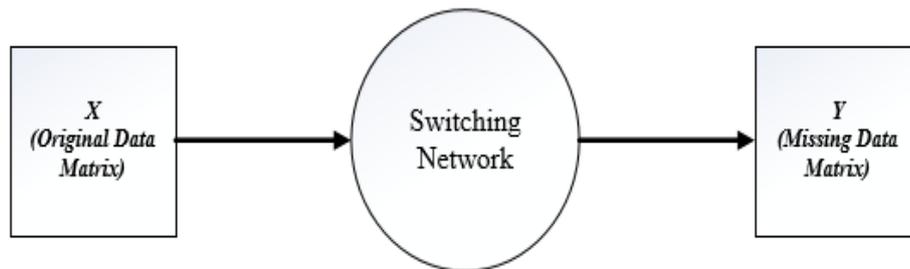
Compressive Sensing (CS) [2] mengambil keuntungan dari sparsity data untuk menyimpulkan data yang hilang, seperti kerangka untuk interpolasi lalu lintas dan pendekatan rekonstruksi compressive sensing untuk lalu lintas jaringan [3]. Sparsity Regularized Matrix Factorization (SRMF) [4] menggunakan teknik dekomposisi matriks untuk memulihkan entri yang hilang dalam matriks. Dalam tulisan ini, kontribusi utama adalah vektor yang diperkenalkan ke model serangkaian waktu dari matriks lalu lintas spasial murni, yang menjaga sifat multi-arah dari data lalu lintas jaringan dan mengekstrak struktur laten lalu lintas melalui faktorisasi vektor dan dengan mengambil keuntungan dekomposisi vektor, yang memproyeksikan kejadian ke dalam ruang berdimensi rendah, dan informasi spasial dengan model regularisasi, mengusulkan metode penyelesaian vektor spasial baru untuk memperkirakan data lalu lintas yang hilang. Analisis SVD dapat digunakan untuk mencari dasar orthonormal dari empat sub-rentang matriks, basis yang dapat digunakan untuk mencari solusi SPL. Hasil analisis algoritma yang digunakan, penulis mampu melakukan rekonstruksi hingga 98% lebih baik dari metode lain yang biasa digunakan dalam proses interpolasi. Dapat disimpulkan bahwa berdasarkan SRSVD A, setiap matriks SPL dan estimasi data yang hilang dapat dicari menggunakan algoritma SRSVD dengan mendapatkan nilai NMSE yang lebih rendah.

STUDI LITERATUR

Sistem Model Matriks Lalu Lintas Internet

Traffic Matriks (TM) adalah iktisar dari seluruh lalu lintas jaringan, yang menunjukkan volume lalu lintas antara pasangan node asal ke tujuan (OD) dalam jaringan. Ada kekurangan dalam sistem pengukuran yang baik dan sistem pengumpulan data dapat gagal karena protokol transport yang tidak dapat diandalkan dan komponen jaringan legacy. TM tidak dapat diamati secara langsung, tetapi dapat diperkirakan dengan lalu lintas tautan pengamatan tidak langsung. Untuk contoh sistem model matriks lalu lintas internet seperti Gambar 1:

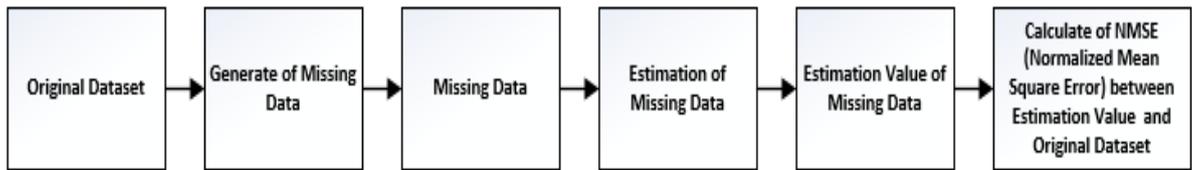
Berbagai karya telah ditambahkan ke bidang ini, misalnya, teknik pendahuluan terdekat k-Nearest Neighbors Imputation (k-NNI) dan Sparsity Regularized SVD (SRSVD) Algorithm. Strategi-strategi ini adalah kemampuan beradaptasi yang buruk, dan tidak mampu menghadapi sejumlah besar informasi yang hilang. Compressive Sensing (CS) adalah prosedur intens untuk menyelesaikan informasi yang hilang.



Gambar1 Model Sistem Matriks Lalu Lintas Internet

Sistem Model Estimasi Data Hilang

Sistem yang akan dibangun adalah sistem yang dapat melakukan optimasi pada estimasi data yang hilang pada matriks lalu lintas internet. Dalam prosesnya, penulis akan menggunakan dataset pengukuran yang biasa digunakan oleh peneliti dalam berbagai eksperimen. Dataset awal berisi nilai yang lengkap dengan matriks 12 x 12 ukuran 12 baris dan 12 kolom. Maka penulis akan menghasilkan data yang hilang pada rentang acak 0% s / d 96%. Data estimasi hilang dengan interpolasi global yang menggunakan prosedur least-square atau l_2 -optimization algoritma, dan interpolasi lokal dengan menggunakan k-Nearest Neighbor Imputasi (k-NNI). Proses terakhir adalah membandingkan nilai NMSE dari hasil rekonstruksi. Proses estimasi data yang hilang akan dilakukan dalam penelitian ini sepertipada Gambar 2 Estimasi Model Sistem Data Hilang.



Gambar 2 Sistem Model Estimasi Hilang Da

Proses Sparsity Regularized SVD (SRSVD)

Dalam penelitian ini untuk mengatasi masalah optimasi, akan dilibatkan algoritma Basis Pursuit dan fase regularisasi mengoptimalkan hasil estimasi data yang hilang untuk direkonstruksi sebagai nilai yang sebenarnya. Persamaan matematika yang digunak

$$\min \|X\| \text{subject to } AX = Y \tag{1}$$

dimana:

- A adalah matriks Gaussian acak
- X adalah perkiraan matriks data yang hilang
- Y adalah matriks data yang hilang

Proses Singular (SVD) dimulai dengan transpose A^T dan menghitung $A^T A$. Kemudian, hitung nilai Eigen (λ) dari $A^T A$. Yang ketiga, hitung nilai singular (Σ) dari λ . Buat matriks diagonal dengan matriks diagonal dengan menghitung hitungan (Σ) dan invers (Σ^{-1}). Hitung vektor (V) dan transpose menjadi (V^T) untuk menghasilkan Eigenvector Kanan. Langkah keenam, menghitung U menghasilkan nilai Kiri vector eigen dari langkah terakhir dan membuktikan dihasilkan ekstrak hasil data data set awal, dan dapat dibuktikan dengan $X = U \Sigma V^T$.

Parameter regularisasi λ memungkinkan tradeoff antara tepat sesuai dengan data yang diukur dan tujuan mencapai peringkat rendah. Penulis berasal estimasi data yang hilang dengan interpolasi global yang menggunakan bolak prosedur least-square atau l_2 optimisasi algoritma. Seperti di l_0 optimisasi kasus, masalah meminimalkan l_2 norm dirumuskan oleh,

$$\min \|X\|_2 \text{subject to } (\mathcal{A}AX = B) \tag{2}$$

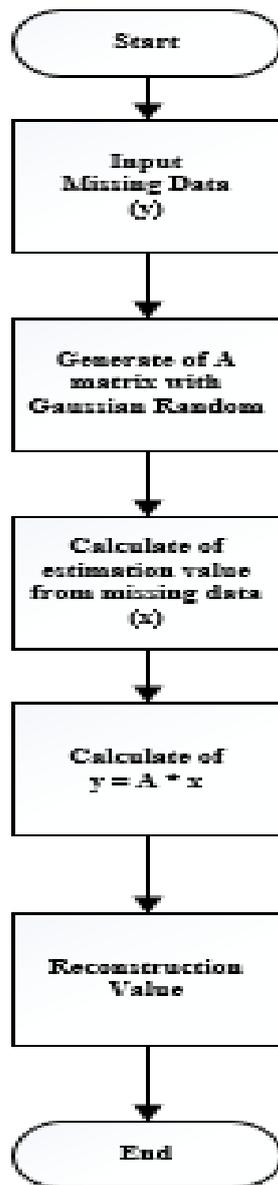
Asumsikan bahwa matriks kendala A memiliki pangkat penuh, masalah ini sekarang menjadi sistem yang kurang tertata yang memiliki solusi tak terbatas. Tujuan dalam hal ini adalah untuk menarik keluar solusi terbaik, yaitu memiliki terendah l_2 -norm, dari ini tak terhingga banyaknya solusi. Dengan menggunakan trik pengganda Lagrange, kita dapat mendefinisikan Lagrangian,

$$\mathcal{L}(x) = \|X\|_2^2 + \lambda^T (\mathcal{A}AX = B) \tag{3}$$

$$\hat{X} = \mathcal{A}A^T (\mathcal{A}A\mathcal{A}A^T)^{-1} B \tag{4}$$

Dengan menggunakan persamaan ini, dapat dihitung solusi optimal dari l_2 -optimisation masalah. Persamaan ini dikenal sebagai Moore-Penrose pseudo inverse dan masalah itu sendiri biasanya dikenal sebagai Least masalah Square, regresi Least Square, atau optimasi Least Square. Namun, meskipun solusi metode Least Square mudah untuk dihitung, tidak perlu menjadi solusi terbaik. Karena sifat halus dari l_2 -norm itu sendiri, ini sulit untuk menemukan satu solusi terbaik untuk masalah ini. Untuk lebih jelasnya dapat dilihat pada Gambar 3 Proses Sparsity Regularized SVD (SRSVD):

METODA PENELITIAN



Gambar 3 Proses SRSVD Algoritma

HASIL DAN PEMBAHASAN

Full Dataset di bawah ini adalah pengukuran yang digunakan dalam eksperimen, perkiraan data yang hilang pada matriks lalu lintas internet. Dataset menggunakan dataset jaringan TCP yang diperoleh dari hasil pengukuran yang secara langsung diketahui oleh Dataset Abilene yang digunakan dalam penelitian oleh Matthew Roughan, Profesor Universitas Adelaide di The School of Mathematical Sciences. Misalnya pengukuran dataset seperti matriks X :

$$X = \begin{bmatrix} 3.33E-06 & 6.53E-05 & 0.000205 & 4.20E-05 & 5.16E-05 & 6.12E-05 & 4.56E-05 & 0.000102 & 5.65E-05 & 9.34E-05 & 4.85E-05 & 0.000393 \\ 5.56E-05 & 0.010333 & 0.002035 & 0.000646 & 0.000491 & 0.003419 & 0.000606 & 0.002279 & 0.0016 & 0.000178 & 0.00112 & 0.006469 \\ 0.000474 & 0.001592 & 0.003593 & 0.001717 & 0.001229 & 0.003295 & 0.000817 & 0.003472 & 0.001762 & 0.000227 & 0.000562 & 0.001331 \\ 2.89E-05 & 0.000293 & 0.004815 & 0.00047 & 0.001051 & 0.000901 & 0.000494 & 0.002672 & 0.000715 & 0.001798 & 0.001326 & 0.001531 \\ 2.99E-05 & 0.00037 & 0.002111 & 0.000462 & 0.000701 & 0.001143 & 0.000891 & 0.012307 & 0.000908 & 0.000368 & 0.00028 & 0.001063 \\ 0.000596 & 0.001157 & 0.015256 & 0.002672 & 0.004147 & 0.002856 & 0.001236 & 0.003051 & 0.005077 & 0.000433 & 0.001737 & 0.005142 \\ 5.26E-05 & 0.000555 & 0.003372 & 0.000674 & 0.000685 & 0.001002 & 0.000503 & 0.001084 & 0.001605 & 0.000153 & 0.000306 & 0.001506 \\ 4.25E-05 & 0.002424 & 0.011215 & 0.00113 & 0.001129 & 0.00534 & 0.001696 & 0.001421 & 0.007646 & 0.000289 & 0.00319 & 0.006966 \\ 0.000487 & 0.005111 & 0.006709 & 0.002043 & 0.002998 & 0.010416 & 0.003096 & 0.008878 & 0.0171 & 0.001199 & 0.002742 & 0.013983 \\ 3.33E-06 & 0.00013 & 0.000631 & 0.001301 & 0.00018 & 0.000483 & 0.000262 & 0.00025 & 0.000276 & 0.001784 & 0.000412 & 0.000248 \\ 1.39E-05 & 0.001985 & 0.002814 & 0.000543 & 0.001413 & 0.000961 & 0.000283 & 0.002221 & 0.003106 & 0.000594 & 0.00013 & 0.001315 \\ 0.001402 & 0.015742 & 0.008318 & 0.004508 & 0.00193 & 0.007848 & 0.00408 & 0.011459 & 0.016708 & 0.000248 & 0.00372 & 0.023457 \end{bmatrix}$$

Gambar 4 Full Dataset

Matrix X adalah dataset yang menggambarkan proses transmisi data dari sumber ke tujuan. Dataset menggunakan 12 baris dan 12 kolom. Dataset berbentuk seperti matriks yang memiliki koefisien (i) sebagai sumber dan koefisien (j) sebagai tujuan. Hilangnya data dapat menyebabkan keterlambatan pengiriman dan penerimaan paket juga meningkat dan jika data hilang dalam jumlah yang cukup besar akan menyebabkan buruknya kualitas data yang dikirimkan, seperti jumlah paket yang hilang selama transmisi berkurang karena sedang berjalan dan sebagainya. Oleh karena itu, dalam percobaan ini, dengan menggunakan matriks asli maka akan dilakukan proses missing data pada traffic matrix internet dengan pola acak. Matriks Y adalah matriks data yang hilang digunakan untuk percobaan. Untuk detail lebih lanjut seperti matriks Y :

$$Y = \begin{bmatrix} 3.33E-06 & 6.53E-05 & 0.000205 & 4.20E-05 & 5.16E-05 & 6.12E-05 & 4.56E-05 & 0.000102 & 0 & 9.34E-05 & 4.85E-05 & 0.000393 \\ 5.56E-05 & 0.010333 & 0.002035 & 0.000646 & 0.000491 & 0.003419 & 0.000606 & 0.002279 & 0.0016 & 0.000178 & 0 & 0.006469 \\ 0.000474 & 0.001592 & 0 & 0.001717 & 0.001229 & 0.003295 & 0.000817 & 0.003472 & 0.001762 & 0.000227 & 0.000562 & 0.001331 \\ 2.89E-05 & 0.000293 & 0.004815 & 0.00047 & 0 & 0.000901 & 0.000494 & 0.002672 & 0 & 0.001798 & 0 & 0.001531 \\ 2.99E-05 & 0 & 0 & 0.000462 & 0.000701 & 0.001143 & 0.000891 & 0.012307 & 0.000908 & 0.000368 & 0.00028 & 0.001063 \\ 0.000596 & 0 & 0.015256 & 0.002672 & 0.004147 & 0.002856 & 0 & 0.003051 & 0 & 0 & 0.001737 & 0.005142 \\ 5.26E-05 & 0 & 0.003372 & 0.000674 & 0 & 0.001002 & 0.000503 & 0.001084 & 0.001605 & 0.000153 & 0.000306 & 0.001506 \\ 4.25E-05 & 0.002424 & 0 & 0.00113 & 0.001129 & 0.00534 & 0.001696 & 0 & 0.007646 & 0 & 0 & 0.006966 \\ 0 & 0.005111 & 0.006709 & 0.002043 & 0.002998 & 0.010416 & 0.003096 & 0 & 0.0171 & 0.001199 & 0.002742 & 0.013983 \\ 3.33E-06 & 0.00013 & 0.000631 & 0.001301 & 0.00018 & 0.000483 & 0.000262 & 0.00025 & 0.000276 & 0.001784 & 0.000412 & 0.000248 \\ 1.39E-05 & 0.001985 & 0.002814 & 0 & 0.001413 & 0.000961 & 0.000283 & 0.002221 & 0.003106 & 0.000594 & 0.00013 & 0.001315 \\ 0.001402 & 0.015742 & 0 & 0 & 0.00193 & 0.007848 & 0.00408 & 0.011459 & 0 & 0 & 0.00372 & 0.023457 \end{bmatrix}$$

Gambar 5 Hasil Matriks 12X12 Data Hilang

A. Hasil Estimasi Data Hilang Menggunakan Sparsity Regularized SVD (SRSVD)

Berdasarkan hasil eksperimen pada estimasi data yang hilang menggunakan Sparsity Regularized SVD (SRSVD) Algoritma, algoritma ini dapat dianalisis semakin besar proporsi data yang hilang maka hal ini juga merupakan NMSE semakin besar. Berdasarkan hasil eksperimen terhadap estimasi data yang hilang menggunakan algoritma SRSVD dapat dianalisis dari NMSE yang merupakan rata-rata kesalahan dari perbandingan dua buah data yang memiliki nilai berbeda. Dengan menghitung hasil estimasi NMSE hilang data dan dibandingkan dengan nilai sebenarnya.

Menurut tujuan dari penelitian ini adalah untuk melakukan estimasi data yang hilang maka algoritma SRSVD mampu melakukan estimasi data yang hilang hingga 98% dari nilai yang hilang. Estimasi dapat dilakukan dengan NMSE sekecil mungkin seperti Tabel 1 menunjukkan nilai semakin besar NMSE bersama dengan banyak nilainya hilang. Ini berpengaruh pada proses estimasi dengan mencari persamaan antara nilai yang terkandung dalam matriks. Algoritma SRSVD dimulai dengan melakukan dekomposisi matriks menjadi 3 matriks, nilai U, Σ, dan V. Hasil estimasi data yang hilang menggunakan SRSVD dengan 20% data yang hilang seperti matriks X'_{SRSVD} :

$$\hat{X}_{SRSVD} = \begin{bmatrix} 3.16E-06 & 5.39E-06 & 3.22E-06 & 3.07E-06 & 3.65E-06 & 4.16E-06 & 4.28E-06 & 4.06E-06 & 2.18E-06 & 3.14E-06 & 2.51E-06 & 0.000393 \\ 3.22E-06 & 2.08E-06 & 2.61E-06 & 0.000646 & 4.98E-06 & 2.53E-06 & 3.56E-06 & 2.58E-06 & 3.47E-06 & 2.83E-06 & 5.74E-06 & 5.12E-06 \\ 5.54E-06 & 3.41E-06 & 2.75E-06 & 4.67E-06 & 3.52E-06 & 4.10E-06 & 3.13E-06 & 2.88E-06 & 3.46E-06 & 4.03E-06 & 2.78E-06 & 3.47E-06 \\ 4.26E-06 & 2.30E-06 & 2.88E-06 & 4.91E-06 & 3.22E-06 & 3.94E-06 & 4.08E-06 & 5.26E-06 & 6.15E-06 & 0.001798 & 2.89E-06 & 3.19E-06 \\ 3.42E-06 & 2.17E-06 & 3.75E-06 & 5.23E-06 & 3.97E-06 & 3.56E-06 & 2.89E-06 & 3.40E-06 & 4.17E-06 & 3.50E-06 & 4.76E-06 & 2.30E-06 \\ 3.15E-06 & 5.65E-06 & 3.45E-06 & 3.33E-06 & 5.24E-06 & 2.95E-06 & 3.08E-06 & 3.27E-06 & 2.66E-06 & 3.36E-06 & 3.29E-06 & 2.81E-06 \\ 4.50E-06 & 4.39E-06 & 2.93E-06 & 3.31E-06 & 6.88E-06 & 2.33E-06 & 3.63E-06 & 3.00E-06 & 3.54E-06 & 3.26E-06 & 2.91E-06 & 3.19E-06 \\ 3.52E-06 & 2.53E-06 & 3.49E-06 & 3.01E-06 & 4.10E-06 & 4.40E-06 & 3.18E-06 & 3.14E-06 & 3.83E-06 & 3.04E-06 & 6.07E-06 & 3.68E-06 \\ 3.49E-06 & 2.73E-06 & 3.91E-06 & 4.17E-06 & 3.25E-06 & 2.60E-06 & 3.30E-06 & 2.86E-06 & 2.67E-06 & 3.24E-06 & 2.17E-06 & 5.17E-06 \\ 1.93E-06 & 3.07E-06 & 3.30E-06 & 4.18E-06 & 3.91E-06 & 5.27E-06 & 4.25E-06 & 2.22E-06 & 2.63E-06 & 2.59E-06 & 2.45E-06 & 1.96E-06 \\ 3.39E-06 & 3.00E-06 & 2.78E-06 & 4.35E-06 & 0.001413 & 3.60E-06 & 3.19E-06 & 3.12E-06 & 3.55E-06 & 4.98E-06 & 2.71E-06 & 5.79E-06 \\ 3.81E-06 & 0.015742 & 3.38E-06 & 5.18E-06 & 4.21E-06 & 4.32E-06 & 4.21E-06 & 1.42E-06 & 3.19E-06 & 2.80E-06 & 3.67E-06 & 4.14E-06 \end{bmatrix}$$

Gambar 6 Matrix Hasil Estimasi Data Hilang Dengan SRSVD

Faktor dekomposisi nilai singular pada nilai matriks, dari kemiripan itu adalah data karena data diolah oleh informasi dari matriks sparse acak dan iterasi dilakukan hingga 1000 kali nilai yang diperoleh akan dioptimalkan untuk meminimalkan persamaan yang diberikan. Hasil estimasi data yang hilang menunjukkan kemampuan metode yang mampu melakukan rekonstruksi hingga 98% dengan NMSE terkecil 0,003198664.

B. Hasil Estimasi Data yang Hilang Menggunakan metoda k-NNI

Algoritma lain untuk memperkirakan data yang hilang dapat menggunakan interpolasi lokal, yang disebut k-Nearest Neighbors Imputation (k-NNI).

Hasil estimasi data yang hilang menggunakan k-NNI dapat dilihat pada matriks \hat{X}_{k-NNI} . Namun, metode k-NNI hanya mampu melakukan estimasi data yang hilang hingga 20% karena algoritma ini bekerja dengan melakukan perhitungan data yang hilang berdasarkan jumlah tetangga terdekat.

$$\hat{X}_{k-NNI} = \begin{bmatrix} 3.33E-06 & 6.53E-05 & 0.000205 & 4.20E-05 & 5.16E-05 & 6.12E-05 & 4.56E-05 & 0.000102 & 4.56E-05 & 9.34E-05 & 4.85E-05 & 0.000393 \\ 5.56E-05 & 0.010333 & 0.002035 & 0.000646 & 0.000491 & 0.003419 & 0.000606 & 0.002279 & 0.0016 & 0.000178 & 0.003419 & 0.002279 \\ 0.000474 & 0.001592 & 0.003295 & 0.001717 & 0.001229 & 0.003295 & 0.000817 & 0.003472 & 0.001762 & 0.000227 & 0.000562 & 0.001331 \\ 2.89E-05 & 0.000293 & 0.004815 & 0.00047 & 0.000293 & 0.000901 & 0.000494 & 0.002672 & 0.000494 & 0.001798 & 0.000901 & 0.002672 \\ 2.99E-05 & 0.000701 & 0.001143 & 0.000462 & 0.000701 & 0.001143 & 0.000891 & 0.012307 & 0.000908 & 0.000368 & 0.00028 & 0.012307 \\ 0.000596 & 0.004147 & 0.015256 & 0.002672 & 0.004147 & 0.002856 & 0.003051 & 0.003051 & 0.003051 & 0.002672 & 0.001737 & 0.005142 \\ 5.26E-05 & 0.001506 & 0.003372 & 0.000674 & 0.001506 & 0.001002 & 0.000503 & 0.001084 & 0.001605 & 0.000153 & 0.000306 & 0.001506 \\ 4.25E-05 & 0.002424 & 0.00534 & 0.00113 & 0.001129 & 0.00534 & 0.001696 & 0.006966 & 0.007646 & 0.00113 & 0.00534 & 0.006966 \\ 0.005111 & 0.005111 & 0.006709 & 0.002043 & 0.002998 & 0.010416 & 0.003096 & 0.003096 & 0.0171 & 0.001199 & 0.002742 & 0.003096 \\ 3.33E-06 & 0.00013 & 0.000631 & 0.001301 & 0.00018 & 0.000483 & 0.000262 & 0.00025 & 0.000276 & 0.001784 & 0.000412 & 0.000248 \\ 1.39E-05 & 0.001985 & 0.002814 & 0.000594 & 0.001413 & 0.000961 & 0.000283 & 0.002221 & 0.003106 & 0.000594 & 0.00013 & 0.001315 \\ 0.001402 & 0.015742 & 0.007848 & 0.007848 & 0.00193 & 0.007848 & 0.00408 & 0.011459 & 0.00408 & 0.007848 & 0.00372 & 0.023457 \end{bmatrix}$$

Gambar 7. Matrix Hasil Estimasi Data Hilang Dengan k-NNI

Jika semua tetangga mengalami data yang hilang maka algoritma ini tidak dapat melakukan perhitungan jumlah tetangganya. Oleh karena itu, kesimpulan yang dapat diambil dari kedua algoritma ini memiliki kelebihan dan kekurangan masing-masing ketika melakukan estimasi data yang hilang. Dengan menghitung jarak terdekat antara variabel yang mengalami data yang hilang dengan variabelnya, maka lakukan pembobotan pada masing-masing variabel tersebut di mana jarak terdekat akan mendapatkan bobot dan imputasi terbesar pada data yang hilang berdasarkan jarak terkecil. Dimana dalam metode ini adalah nilai yang diperhitungkan pada data yang hilang diperoleh berdasarkan estimasi nilai bobot variabel dalam kumpulan data yang digunakan sehingga menghasilkan nilai NMAE yang lebih kecil dari estimasi koefisien korelasi dengan menggunakan pembobotan variabel yang ada dalam satu dataset. Dan dapat dilihat pada variabel di baris ke-7 dan kolom ke-2, estimasi data yang hilang diperoleh 0,001506 dan dataset awalnya bernilai 0,000555, menemukan perbedaan sekitar 0,000951. Namun, jika ada data yang hilang di semua baris dari data input maka algoritma ini tidak akan dapat memproses imputasi pada data yang hilang karena tidak ada perbandingan nilai dari tetangga terdekatnya jika keduanya mengandung data yang hilang. Ubah nilai yang tidak terlalu jauh sehingga hasil

rekonstruksi dari pendekatan ke nilai yang benar, serta untuk variabel data yang hilang. Perbandingan rata-rata NMAE di atas, dapat disimpulkan bahwa rata-rata metode k-NNI memiliki kinerja yang hebat dengan hasil NMAE terkecil bila dibandingkan dengan metode perbandingan. Namun, dalam kasus-kasus tertentu, k-NNI tidak dapat melakukan estimasi data yang hilang jika semua tetangganya kehilangan data karena melakukan perhitungan berdasarkan jumlah tetangga terdekat.

C. Perbandingan Hasil Estimasi Data Hilang Dengan SRSVD dan k-NNI

Berdasarkan hasil eksperimen pada estimasi data yang hilang menggunakan Sparsity Regularized SVD (SRSVD) Algorithms dan metode kNNI, dapat dianalisis bahwa semakin besar proporsi data yang hilang maka NMSE juga semakin besar, seperti terlihat pada Tabel 1. Disini perbedaan nilai error dinyatakan dengan NMSE yang diperoleh dengan menggunakan SRSVD, dan k-NNI, dengan data yang hilang sebesar 0% sampai dengan 20% dari jumlah data, dan dilakukan percobaan untuk estimasi data yang hilang.

Tabel 1. Perbedaan Normalized Mean Absolute Error (NMAE) Menggunakan SRSVD, k-NNI dan PCA

Proporsi Missing Data (%)	Normalized Mean Square Error (NMAE)	
	SRSVD	k-NNI
0	0	0
4	4.1032E-02	6.0878E-03
8	4.2491E-02	8.9730E-02
12	9.9253E-02	7.4907E-02
16	1.8562E-01	2.3119E-02
20	2.5690E-01	9.0555E-03

Pada proporsi data yang hilang adalah 0% hingga 20% menggunakan SRSVD diperoleh NMAE semakin besar dan dalam 0,25. Namun, saat menggunakan k-NNI memperoleh NMAE yang tidak stabil khususnya kondisi NMAE ada yang turun dan ada yang naik. Pada proporsi data yang hilang adalah 0% hingga 8%, NMAE semakin besar, tetapi data pada proporsi kehilangan NMAE menurun 12% memiliki nilai 0,074. Kemudian dalam proporsi data yang hilang adalah 16% hingga 20% dari NMAE-nya semakin kecil lagi..

KESIMPULAN

Evaluasi menggunakan dataset TM dari Dataset Abilene Network menunjukkan efektivitas pendekatan perwujudan yang dijelaskan di sini. Apakah diterapkan pada estimasi TM, metode perwujudan yang dijelaskan di sini secara konsisten keluar melakukan metode lain yang umum digunakan dan melakukannya di seluruh rentang skenario data yang hilang dari yang murni acak ke yang sangat terstruktur di mana seluruh kolom dan atau baris TM hilang dan dari data yang hilang tingkat sangat rendah menjadi sangat tinggi (misalnya, 90% dan lebih). Algoritma menjelaskan di sini dalam skala secara linear dengan ukuran data sehingga dataset yang lebih besar dapat dianalisis.

Masalah data yang hilang dapat dianalogikan sebagai sistem sparse persamaan linear yang banyak mengandung 0 pada matriks. Nilai NMSE menunjukkan kemampuan algoritma melakukan estimasi data yang hilang sehingga kondisi dapat dipenuhi dalam sistem linier yang belum ditentukan sehingga elemen dari matriks tidak ada yang bernilai 0. Jika dalam kondisi lalu lintas internet yang sebenarnya, proses estimasi data yang hilang sangat membantu pengguna pada saat pertukaran data sedang berlangsung, karena jika setiap node memiliki data yang lengkap, semua garis dapat dilewati tanpa kehilangan data terlalu banyak. Terutama jika data yang hilang dapat diperkirakan sama dengan data asli sehingga informasinya diterima penuh oleh penerima.

Sehingga analisis algoritma SRSVD dapat digunakan untuk menemukan basis orthonormal dari sub-range matriks, basis yang dapat digunakan untuk mencari solusi SPL konsisten dan mencari solusi pendekatan terbaik SPL tidak konsisten dan SRSVD dapat digunakan untuk menemukan invers pseudo dan peringkat matriks. Hasil analisis algoritma yang digunakan, penulis mampu melakukan rekonstruksi hingga 98% dengan NMSE 3×10^{-3} lebih baik dari metode lain yang

biasa digunakan dalam proses interpolasi. Dapat disimpulkan bahwa berdasarkan SRSVD matriks A, jadi setiap SPL dan estimasi data yang hilang dapat dicari menggunakan algoritma SRSVD dengan mendapatkan nilai NMSE yang lebih kecil.

DAFTAR PUKTAKA

- [1]. Green, P. "An introduction to network architectures and protocols." IEEE Transactions on Communications 28.4 (1980): 413-424.
- [2]. Roughan, Matthew, Mikkel Thorup, and Yin Zhang. "Traffic engineering with estimated traffic matrices." Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement. ACM, 2003.
- [3]. Y.Zhang, M.Roughan, N.Duffield, and A.Greenberg. Fast accurate computation of large-scale IP traffic matrices from link loads. In Proc.of ACM SIGMETRICS, 2003.
- [4]. A.Feldmann, A.Greenberg, C.Lund, N.Reingold, J.Rexford, and F.True. Deriving traffic demands for operational IP networks: Methodology and experience. IEEE/ACM Transactions on Networking, pages 265–279, 2001.