

Tugas Pertemuan 9

~ Clustering ~

Paper 1

Judul = Correction to: Community detection and unveiling of hierarchy in networks: a density-based clustering approach

Penulis = Zineb Felfli, Roy George, Khalil Shujaee, Mohamed Kerwat

Tahun = 2020

Review =

Paper ini menyajikan pendekatan mekanika statistik yang disebut Correlation Density Rank, yang digunakan untuk menentukan peringkat simpul dalam sebuah jaringan berdasarkan tingkat kepentingan relatifnya dan mengidentifikasi pemimpin langsung setiap anggota, termasuk munculnya pemimpin kelompok. Pendekatan ini menggunakan fungsi Gaussian yang dinormalisasi untuk menggambarkan dampak sebuah simpul dalam lingkungannya dan menghasilkan peringkat kepadatan simpul dengan mempertimbangkan jarak antar simpul sebagai hukuman. Teknik ini menggunakan prosedur hill-climbing untuk menentukan penarik kepadatan dan mengidentifikasi orang tua unik (pemimpin) dari setiap anggota serta pemimpin kelompok.

Pendekatan ini didasarkan pada karya sebelumnya yang mengevaluasi jaringan sintetik kecil, dan dalam artikel ini, para penulis menggunakan jaringan yang lebih besar yang dihasilkan oleh benchmark LFR dalam upaya untuk memvalidasi dan menguji kegunaan pendekatan ini serta mengidentifikasi batasan-batasannya. Metode ini diuji secara menyeluruh menggunakan jaringan sintetik yang dihasilkan oleh algoritma benchmarking LFR untuk ukuran jaringan antara 500 hingga 30000 simpul dan parameter pencampuran antara 0,1 hingga 0,9.

Paper ini menyimpulkan dengan mengevaluasi kinerja algoritme dan membahas batasan-batasnya. Bagian 4 artikel memberikan ringkasan artikel dan membahas batasan pendekatan ini. Para penulis mencatat bahwa pendekatan ini mengasumsikan distribusi kepadatan Gaussian, yang mungkin tidak cocok untuk semua jenis jaringan. Mereka juga menyadari bahwa pendekatan ini mungkin tidak cocok untuk jaringan yang sangat besar karena keterbatasan komputasi.

Paper 2

Judul = Density-based clustering of static and dynamic functional MRI connectivity features obtained from subjects with cognitive impairment

Penulis = D. Rangaprakash, Toluwanimi Odemuyiwa⁴, D. Narayana Dutts, Gopikrishna Deshpande, Alzheimer's Disease Neuroimaging Initiative
Tahun = 2020

Review =

Paper ini menyajikan diskusi terperinci tentang pengelompokan berbasis kepadatan fitur konektivitas fungsional MRI pada individu dengan gangguan kognitif. Studi ini membandingkan kinerja dua metode pengelompokan populer, yaitu DBSCAN dan OPTICS, dalam mengidentifikasi secara akurat individu dengan gangguan kognitif, termasuk penyakit Alzheimer. Paper ini menyoroti keunggulan penggunaan pengelompokan berbasis kepadatan dibandingkan dengan pengklasifikasi berbimbing dalam analisis data neuroimaging, termasuk ketidakpekaan terhadap pencilan dan kemampuan untuk bekerja dengan data tanpa label.

Studi ini menggunakan fitur konektivitas fungsional statis dan dinamis untuk pengelompokan, yang menggambarkan kekuatan dan variasi temporal konektivitas otak, masing-masing. Paper ini menjelaskan perbedaan antara fitur konektivitas statis dan dinamis serta bagaimana fitur-fitur tersebut dapat digunakan untuk mengidentifikasi individu dengan gangguan kognitif.

Untuk menilai ketahanan pengelompokan terhadap noise/pencilan, studi ini mengusulkan metode baru yang disebut recursive-clustering using additive-noise (R-CLAN). Hasilnya menunjukkan bahwa kedua algoritma pengelompokan efektif, meskipun OPTICS dengan fitur konektivitas dinamis lebih unggul dalam hal kemurnian kluster (95,46%) dan ketahanan terhadap noise/pencilan.

Paper ini membahas implikasi studi ini dalam diagnosis dan pengobatan gangguan kognitif, termasuk penyakit Alzheimer. Studi ini menunjukkan bahwa pengelompokan berbasis kepadatan dapat mengidentifikasi secara akurat dan kuat kelas diagnostik dengan cara yang tidak terbimbing menggunakan konektivitas otak. Hal ini memiliki implikasi penting bagi diagnosis dini dan pengobatan gangguan kognitif, yang dapat meningkatkan hasil dan kualitas hidup pasien.

Paper 3

Judul = Characterizing communities of hashtag usage on twitter during the 2020 COVID-19 pandemic by multi-view clustering

Penulis = Lain J. Cruickshank, Kathleen M. Carley

Tahun = 2020

Review =

Paper ini menyajikan sebuah studi tentang karakterisasi komunitas penggunaan hashtag di Twitter selama pandemi COVID-19 tahun 2020 menggunakan metode pengelompokan multi-view. Studi ini mengusulkan analisis topik diskusi di Twitter dengan mengelompokkan hashtag dan memperkenalkan teknik pengelompokan multi-view yang baru yang menggabungkan beberapa jenis data yang berbeda yang dapat digunakan untuk menggambarkan interaksi pengguna dengan

hashtag. Studi ini memberikan wawasan unik tentang topik diskusi dan bagaimana topik-topik tersebut berkembang seiring berjalannya pandemi.

Paper ini berisi informasi tentang metrik graf untuk grafik pandangan harian, pengelompokan ensemble untuk setiap periode pengelompokan hashtag harian, dan analisis data. Pengelompokan ensemble, meta-cluster pada setiap periode waktu menghasilkan struktur pengelompokan yang agak berbeda untuk setiap periode waktu tersebut. Periode ketiga memiliki sedikit lebih banyak pengelompokan dan pengelompokan yang lebih kecil dan berukuran lebih teratur dibandingkan periode waktu lainnya. Periode waktu ketiga juga memiliki kemiripan pengelompokan yang lebih tinggi, dalam hal rata-rata pasangan ARI antara pengelompokan harian yang terdiri darinya. Untuk periode waktu pertama dan kedua, seringkali terdapat satu pengelompokan besar yang memiliki sekitar 4.400 hashtag, sedangkan pengelompokan terbesar pada periode waktu ketiga memiliki 3.025 hashtag. Oleh karena itu, periode waktu ketiga memiliki struktur pengelompokan yang lebih seimbang, sepanjang seluruh periode waktu, dibandingkan dua periode waktu lainnya.

Menganalisis hashtag yang ada dalam pengelompokan dari meta-pengelompokan ensemble yang berbeda mengungkapkan kelompok topik yang persisten dan kelompok yang berubah seiring waktu. Pada setiap periode waktu, selalu ada pengelompokan yang memiliki hashtag untuk berita terbaru atau sumber berita, pengelompokan yang memiliki hashtag terkait bisnis dan perdagangan, pengelompokan yang memiliki hashtag terkait politik Amerika Serikat, dan pengelompokan bahasa asing, terutama Italia, Jerman, dan Spanyol.

Wawasan dari studi ini dapat digunakan untuk menginformasikan pesan kesehatan masyarakat selama pandemi. Dengan memahami topik dan tren dalam diskusi Twitter tentang COVID-19, pejabat kesehatan masyarakat dapat menyusun pesan mereka untuk lebih mencapai dan melibatkan masyarakat. Secara keseluruhan, paper ini menyediakan informasi berharga tentang penggunaan pengelompokan multi-view untuk menganalisis penggunaan hashtag di Twitter selama pandemi dan wawasan yang diperoleh dari analisis tersebut.