

Building Bridges: Aligning Embeddings for Biomedical Applications



Biomedical data is multimodal

Different data formats ...

Detecting Protein-Protein Interaction Based on Protein Fragment Complementation Assay

Tianwen Wang¹, Ningning Yang¹, Chen Liang¹, Hongju Xu¹, Yafei An¹, Sha Xiao¹, Mengyuan Zheng¹, Lu Liu¹, Gaozhan Wang¹, Lei Nie¹

Affiliations + expand

PMID: 32053071 DOI: 10.2174/1389203721666200213102829

Abstract

Proteins are the most critical executive molecules by genetic materials in any form of life. More frequently, that interacts with other protein(s), which is more evic in the real context of a cell. Identifying the interaction: for the biochemistry investigation of an individual pro picture for the interacting members at the scale of prc mapping). Here, we introduced the currently available the interaction between candidate protein pairs base particular proteins. Emphasis was put on the principle systems are dihydrofolate reductase (DHFR), β -lactam luciferase, β -galactosidase, GAL4, horseradish perox fluorescent protein (GFP), and ubiquitin.

Keywords: Protein-protein interaction; enzyme; fluor protein complementation assay; ubiquitin.

Bi-dimensional principal gene feature selection from big gene expression data

Xiaoqian Hou¹, Jingyu Hou¹, Guangyan Huang¹

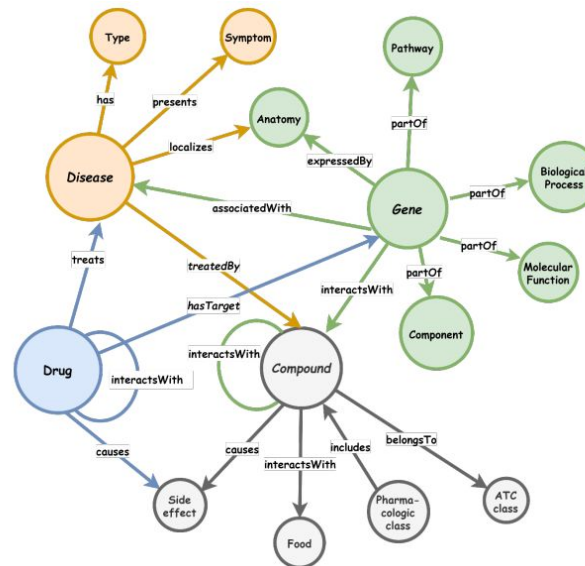
Affiliations + expand

PMID: 36477666 PMID: PMC9728919 DOI: 10.1371/journal.pone.0278583

Abstract

Gene expression sample data, which usually contains massive expression profiles of genes, is commonly used for disease related gene analysis. The selection of relevant genes from huge amount of genes is always a fundamental process in applications of gene expression data. As more and more genes have been detected, the size of gene expression data becomes larger and larger; this challenges the computing efficiency for extracting the relevant and important genes from gene expression data. In this paper, we provide a novel Bi-dimensional Principal Feature Selection (BPFS) method for efficiently extracting critical genes from big gene expression data. It applies the principal component analysis (PCA) method on sample and gene domains successively, aiming at extracting the relevant gene features and reducing redundancies while losing less information. The experimental results on four real-world cancer gene expression datasets show that the proposed BPFS method greatly reduces the data size and achieves a nearly double processing speed compared to the counterpart methods, while maintaining better accuracy and effectiveness.

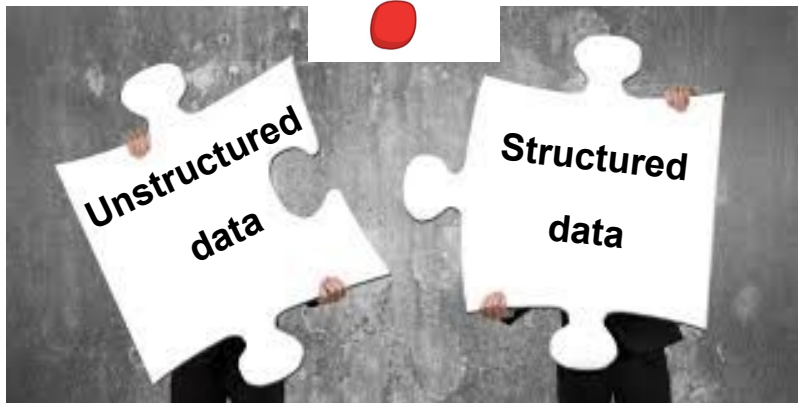
Copyright: © 2022 Hou et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



... require different processing methods

Biomedical data is multimodal

How to integrate and leverage both types of information ...



... to derive new biomedical insights?

Team Project Variants

Variant 1

**Biomedical Concept
Alignment**

Variant 2

**Mapping Biomedical
Text to Knowledge
Graph (KG)**

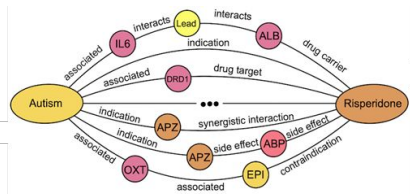


Team Project Variants

Variant 1

Biomedical Concept Alignment

i.e., improve drug discovery through text and graph integration



Can risperidone be used for traumatic brain injury?

Review > Chem Biol Interact. 2023 Jan 5;369:110296. doi: 10.1016/j.cbi.2022.110296. Epub 2022 Dec 7.

Neuropharmacological effect of risperidone: From chemistry to medicine

Abstract

As the second-oldest atypical antipsychotic, risperidone has a long history of off-label usage for treating behavioural and psychological signs and symptoms of dementia (BPSD), such as agitation, aggressiveness, and psychosis. Risperidone has been shown in several trials to have a statistically significant benefits when used in a therapeutic context. Several lines of evidence suggest a possible role of risperidone via the antagonistic effect of Dopamine D2 and 5HT-1A receptor in different neurological diseases like cognitive dysfunction of schizophrenia, neuroinflammation, Huntington's disease, and sleep cycle management. Therefore, the pharmacological interactions of risperidone in all these diseases were investigated. Some reports on the use of risperidone in the treatment of dopaminergic psychosis have been slightly conflicting. However, more research is needed to evaluate the role of risperidone in the treatment of these neurological diseases.

Keywords: Dopaminergic psychosis; Neuroinflammation; Neurological diseases; Neuropharmacological treatment; Risperidone.

Variant 2

Mapping Biomedical Text to KG

i.e., mapping gene-disease relations from text to links in the graph

BRCA1 Gene in Breast Cancer

ELIOT M. ROSEN,^{1,2,3,5*} SAIJUN FAN,^{1,2} RICHARD G. PESTELL,^{2,3,4} AND ITZHAK D. GOLDBERG^{1,2}

¹Department of Radiation Oncology, Long Island Jewish Medical Center, New York, New York

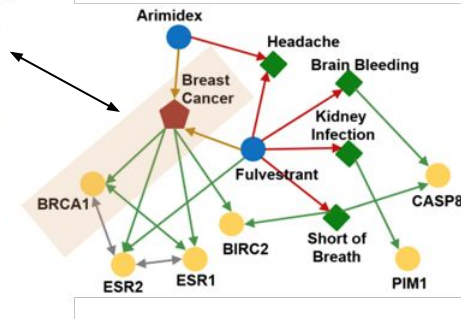
²Department of Radiation Oncology, Albert Einstein College of Medicine, Bronx, New York

³Department of Developmental and Molecular Biology, Albert Einstein College of Medicine, Bronx, New York

⁴Department of Medicine, Albert Einstein College of Medicine, Bronx, New York

⁵Division of Hormone-Dependent Tumor Biology, Albert Einstein College of Medicine, Bronx, New York

The *BRCA1* gene was identified and cloned in 1994 based on its linkage to early onset breast cancer and breast-ovarian cancer syndromes in women. While inherited mutations of *BRCA1* are responsible for about 40–45% of hereditary breast cancers, these mutations account for only 2–3% of all breast cancers, since the *BRCA1* gene is rarely mutated in sporadic breast cancers. However, *BRCA1* expression is frequently reduced or absent in sporadic cancers, suggesting a much wider role in mammary carcinogenesis. Since *BRCA1* was cloned in 1994, its molecular function has been the subject of intense investigation. These studies have revealed multiple functions of the *BRCA1* that may contribute to its tumor suppressor activity, including roles in: cell cycle progression, several highly specialized DNA repair processes, DNA damage-responsive cell cycle checkpoints, regulation of a set of specific transcriptional pathways, and apoptosis. Many of these functions are linked to protein-protein interactions involving different portions of the 1,863 amino acid (aa) *BRCA1* protein. *BRCA1* functions in cell cycle progression and the DNA damage response appear to be regulated by distinct and specific phosphorylation events, but the molecular pathways activated by these phosphorylations are only beginning to be unraveled. In addition, the reason that *BRCA1* mutation carriers develop specific tumor types (breast and ovarian cancers in women and possibly prostate cancers in men) is not clearly understood. Elucidation of the precise molecular functions of the *BRCA1* gene product will greatly enhance our understanding of the pathogenesis of hereditary as well as sporadic mammary carcinogenesis. *J. Cell. Physiol.* 196: 19–41, 2003. © 2003 Wiley-Liss, Inc.



V1: Biomedical Concept Alignment

Does combining different types of data improve biomedical tasks?

Tasks

- Collect data
- Create entity representations
 - ◆ Generate representations of the same concept (e.g., a disease) from unstructured text (e.g., w/ pre-trained language models) & structured data (e.g., w/ KG embeddings)
- Align representations
 - ◆ Combine different representations of the same concept
- Evaluate performance
 - ◆ Test how well the combined representations improve biomedical tasks (e.g., disease diagnosis, treatment recommendations)



V2: Mapping Biomedical Text to KG

How to map relationships between biomedical concepts in literature to KG?

Tasks

- Collect data
- Create entity representations
 - ◆ Generate representations of the same concept (e.g., a disease) from unstructured text (e.g., w/ pre-trained language models) & structured data (e.g., w/ KG embeddings)
- Design a mapping function
 - ◆ How to map the relationships between concepts in literature to KG triples?
- Evaluate performance
 - ◆ i.e., on biomedical databases



Requirements

- Duration: 6 months
- Number of participants: min 2, max 8
 - ◆ **Applying as a group** is possible!

- Prerequisites:
 - ◆ One of Machine Learning, Knowledge Graphs, Data Mining AND
 - ◆ One of Text Analytics, Advanced Methods in Text Analytics AND
 - ◆ Knowledge in Python

- Interest in knowledge representation learning & NLP
- Biomedical knowledge not required

- Applicable: Business Informatics and MMDS
- Instructors: Andreea Iana, Rita Torres de Sousa