Multi-modality Helps in Solving Biomedical Problems: Theory and Applications

Dr. Sriparna Saha and Mr. Pratik Dutta,
AI-NLP-ML Group, Department of Computer Science and Engineering, Indian Institute of Technology Patna

Overview

- Introduction
- Basics of machine learning and Deep Learning
- Concepts of Multi-modal approaches and its application
- Introduction of Computational Biology Problems
- Application of ML and DL in Computational Biology Problems
- Deep learning-based multi-modal approach in Computational Problem
- Conclusion and future Work

Introduction

- Technological advances in genomics and imaging have led to an explosion of molecular and cellular profiling data.
- The rise of this biological datasets maintains a exponential growth
- This rapid increase in biological data dimension and acquisition rate is challenging conventional analysis strategies.
- Machine learning advancements promise to understand the hidden information within the large datasets, and make accurate predictions than traditional approaches

Introduction

- Deep learning has inherent power to extract informative features from the hidden abstract representation of the data
- The main challenge of converting the abstract representation into machine understandable format
- Deep learning performs better than traditional machine learning approaches for converting the representation
- This power leads deep learning shows promising results in various research fields

Introduction

- Also, it has been noticed, different representations of the data makes more intelligent model than a single representation
  - Single modality(representation) does not capture all the aspects of the data representation
  - Various representation help to analysis the underlying information of the data
- This leads the researchers to intrigue how the deep multi-modal approaches solves various research problems

Basics of Machine Learning and Deep Learning

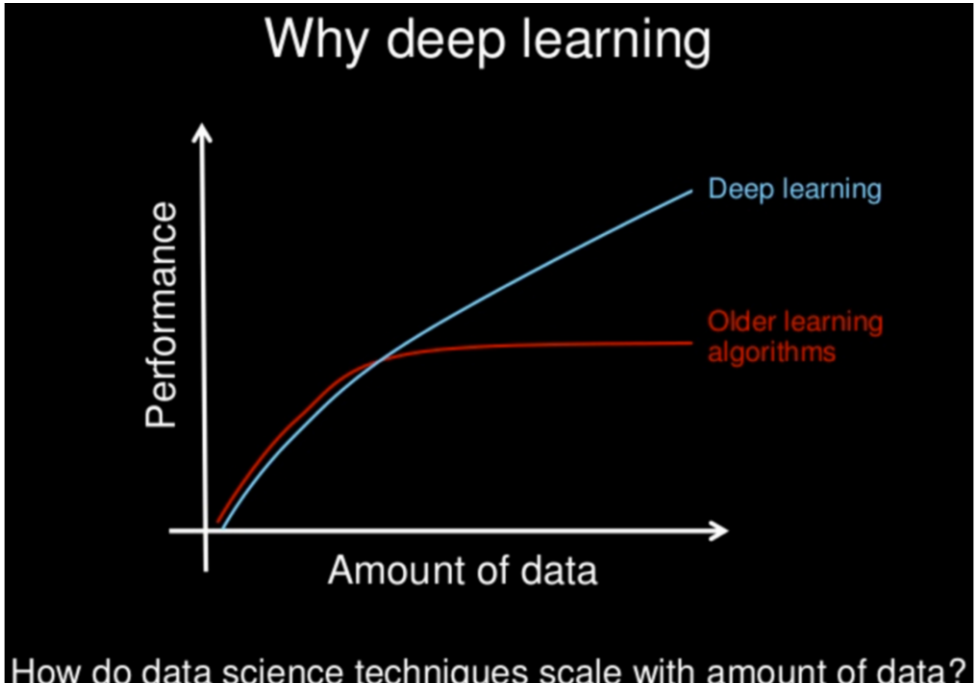
- The two main concepts of AI
  - Machine Learning:
    - *“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E ”- Tom Mitchell*
    - *Parse data, learn from that data, and then apply what they’ve learned to make informed decisions*
  - Deep Learning:
    - *A particular kind of machine learning*
    - *Achieves great power and flexibility by learning to represent the world as nested hierarchy of concepts*
    - *Each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones*

Basics of Machine Learning and Deep Learning

- Differences between machine learning and deep learning
  - For small data, traditional machine learning performs well
  - With the increase with data, deep learning performs better than traditional machine learning algorithms
  - Deep learning algorithms try to learn high-level features from data, whereas ML algorithms need hand-coded or defined features
  - Deep learning algorithms heavily depend on high-end machines, whereas traditional ML algorithms can work on low-end machines

Basics of Machine Learning and Deep Learning

- Differences between machine learning and deep learning



Basics of Machine Learning and Deep Learning

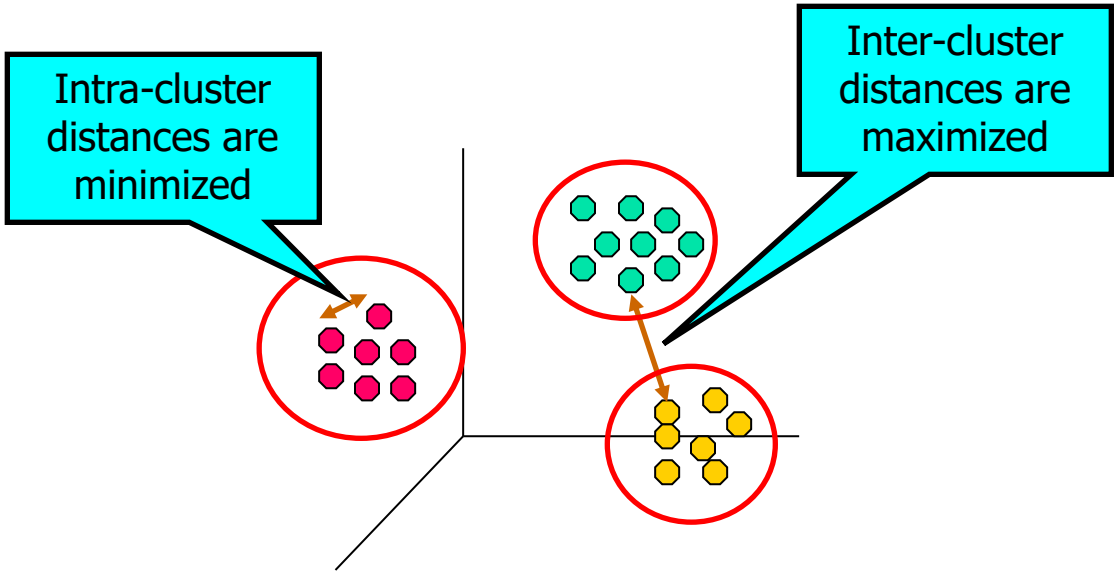
- Machine Learning
  - Semi-supervised Classification
    - Application in Satellite Image Segmentation
    - MR Brain Image Segmentation
    - Gene expression data clustering
  - Multiobjective Multiview Learning
    - Search result clustering
    - Cyberbully detection
- Multiobjective based Clustering Techniques
  - Bioinformatics
  - Information Retrieval
  - Evidence Based Medicine
  - Entity Disambiguation
  - IOT Data
  - SOM
- Information Extraction of Biomedical Texts
  - Named entity recognition
    - Feature selection techniques
  - Patient de-identification
  - Case similarity
  - Sentiment expressed in medical blogs
  - Chronological ordering of events

Basics of Machine Learning

- Machine Learning
  - Unsupervised Technique
  - Supervised Technique
  - Semi-supervised Technique
  - Reinforcement Learning

Unsupervised Technique: Clustering

- Finding groups of objects such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups



What is Cluster Analysis?

- Cluster analysis (or *clustering*, *data segmentation, …*)
  - Finding similarities between data according to the characteristics found in the data and grouping similar data objects into clusters

- Unsupervised learning: no predefined classes (i.e., *learning by observations* vs. learning by examples: supervised)

- Typical applications
  - As a stand-alone tool to get insight into data distribution
  - As a preprocessing step for other algorithms

Clustering as a Preprocessing Tool (Utility)

- Summarization:
  - Preprocessing for regression, PCA, classification, and association analysis
- Compression:
  - Image processing: vector quantization
- Finding K-nearest Neighbors
  - Localizing search to one or a small number of clusters
- Outlier detection
  - Outliers are often viewed as those “far away” from any cluster

Quality: What Is Good Clustering?

- A good clustering method will produce high quality clusters
  - high intra-class similarity: cohesive within clusters
  - low inter-class similarity: distinctive between clusters
- Quality of a clustering method depends on
  - the similarity measure used by the method
  - its implementation, and
  - its ability to discover some or all of the hidden patterns

Measure the Quality of Clustering

- Dissimilarity/Similarity metric
  - Similarity is expressed in terms of a distance function, typically metric: $d(i, j)$
  - Definitions of distance functions are usually rather different for interval-scaled, boolean, categorical, ordinal ratio, and vector variables
  - Weights should be associated with different variables based on applications and data semantics
- Quality of clustering:
  - There is usually a separate “quality” function that measures the “goodness” of a cluster
  - It is hard to define “similar enough” or “good enough”
    - Answer is typically highly subjective

Considerations for Cluster Analysis

- Partitioning criteria
  - Single level vs. hierarchical partitioning (often, multi-level hierarchical partitioning is desirable)
- Separation of clusters
  - Exclusive (e.g., one customer belongs to only one region) vs. non-exclusive (e.g., one document may belong to more than one class)
- Similarity measure
  - Distance-based (e.g., Euclidean, road network, vector) vs. connectivity-based (e.g., density or contiguity)
- Clustering space
  - Full space (often when low dimensional) vs. subspaces (often in high-dimensional clustering)

Requirements and Challenges

- Scalability
  - Clustering all the data instead of only on samples
- Ability to deal with different types of attributes
  - Numerical, binary, categorical, ordinal, linked, and mixture of these
- Constraint-based clustering
  - User may give constraints
  - Use domain knowledge to determine input parameters
- Interpretability and usability
- Others
  - Discovery of clusters with arbitrary shape
  - Ability to deal with noisy data
  - Incremental clustering and insensitivity to input order
  - High dimensionality

Major Clustering Approaches (I)

- Partitioning approach:
  - Construct various partitions and then evaluate them by some criterion, e.g., minimizing the sum of square errors
  - Typical methods: k-means, k-medoids, CLARANS
- Hierarchical approach:
  - Create a hierarchical decomposition of the set of data (or objects) using some criterion
  - Typical methods: Diana, Agnes, BIRCH, CAMELEON
- Density-based approach:
  - Based on connectivity and density functions
  - Typical methods: DBSACN, OPTICS, DenClue
- Grid-based approach:
  - Based on a multiple-level granularity structure
  - Typical methods: STING, WaveCluster, CLIQUE

Major Clustering Approaches (II)

- Model-based:
  - A model is hypothesized for each of the clusters and tries to find the best fit of that model to each other
  - Typical methods: EM, COBWEB
- Frequent pattern-based:
  - Based on the analysis of frequent patterns
  - Typical methods: p-Cluster
- User-guided or constraint-based:
  - Clustering by considering user-specified or application-specific constraints
  - Typical methods: COD (obstacles), constrained clustering
- Link-based clustering:
  - Objects are often linked together in various ways
  - Massive links can be used to cluster objects: SimRank, LinkClus

Supervised Technique: Classification

- Given a collection of records (*training set*)
  - Each record contains a set of *attributes*, one of the attributes is the *class*
- Find a *model* for class attribute as a function of the values of other attributes
- Goal: previously unseen records should be assigned a class as accurately as possible
  - A *test set* is used to determine the accuracy of the model. Usually, the given data set is divided into training and test sets, with training set used to build the model and test set used to validate it

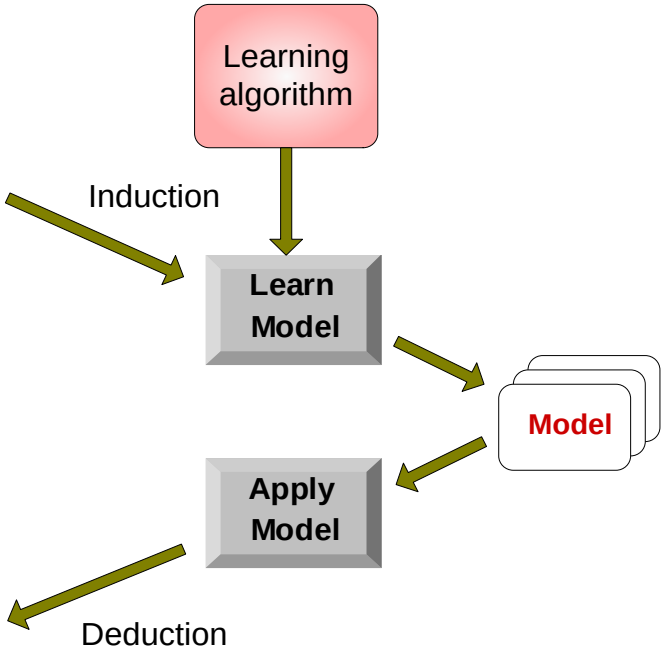
Illustrating Classification Task

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

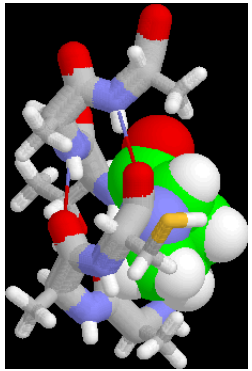
Test Set



Examples of Classification Task

- Predicting tumor cells as benign or malignant
- Classifying credit card transactions as legitimate or fraudulent

- Classifying secondary structures of protein as alpha-helix, beta-sheet, or random coil



- Categorizing news stories as finance, weather, entertainment, sports, etc

Classification Techniques

- Decision Tree based Methods
- Rule-based Methods
- Memory based reasoning
- Neural Networks
- Naïve Bayes and Bayesian Belief Networks
- Support Vector Machines

Multiobjective Optimization: Example of purchasing a car

- Optimizing criteria
  - minimizing the cost, insurance premium and weight and
  - maximizing the feel good factor while in the car.
- Constraints
  - car should have good stereo system, seats for 6 adults and a mileage of 20 kmpl.
- Decision variables
  - the available cars
- In many real world problems we have to simultaneously optimize two or more different objectives which are often competitive in nature
  - finding a single solution in these cases is very difficult.
  - optimizing each criterion separately may lead to good value of one objective while some unacceptably low value of the other objective(s).

Formal Definition of Multiobjective Optimization

- **The multiobjective optimization can be formally stated as:**
  - **Find the vector of decision variables**

$$x=[x_1, x_2 \ldots\ldots\ x_n]^T$$

which will satisfy the m inequality constraints:

$$g_i(x) >= 0,\ i=1,2,\ldots m,$$

And the p equality constraints

$$h_i(x)=0\ ,\ i=1,2,\ldots p.$$

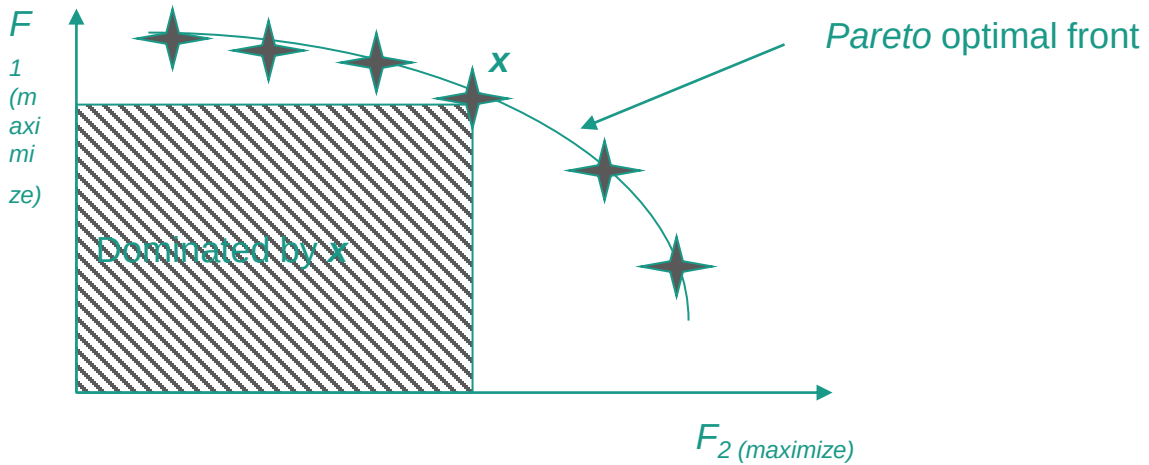
And simultaneously optimizes M objective functions

$$f_1(x), f_2(x) \ldots. f_M(x).$$

  - **No concept of global optimum**
  - **Produce a set of trade-off solutions**
    - **Pareto optimal set**

Multi-Objective Problems: Dominance

- we say *x* dominates *y* if it is at least as good on all criteria and ***better*** on at least one



Domination Relation and Pareto Optimality

- Let us consider two solutions ***a*** *and* ***b*** .Then ***a*** *is said to dominate* ***b*** iff

  $\forall k \in 1,2,\ldots M, f_k(\mathbf{a}) \geq f_k(\mathbf{b})$ *and* $\exists k \in 1,2,\ldots M$

  such that $f_k(\mathbf{a}) > f_k(\mathbf{b})$

  i.e., for all functions f_i, *a* has a higher or equal value than that of *b* and also there exists at least one function f_j for which *a*'s value is strictly greater than that of *b*.

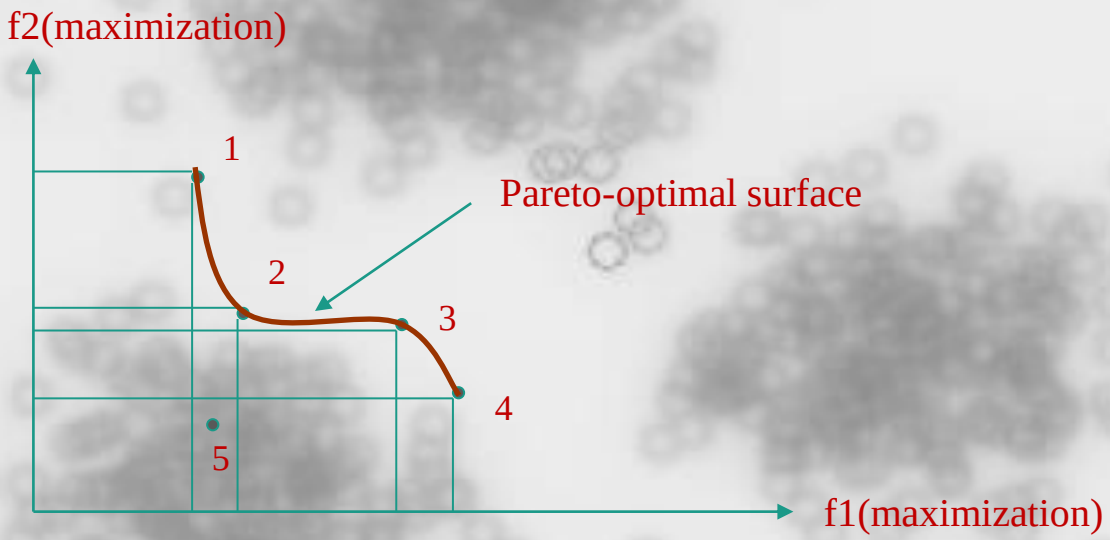
- Non-dominated set

  Among a set of solutions P, the non dominated set of solutions P' are those that are not dominated by any member of the set P. A solution $\rightarrow_X$ is said to be non-dominating w.r.t all the solutions if there exist no solution $\rightarrow_X^*$ which dominates $\rightarrow_X$.

- Pareto-optimal Set:

  The non dominated set of the entire search space S is the globally Pareto optimal set.

Example of Dominance and Pareto-Optimality



- Here solutions 1, 2, 3 and 4 are non-dominating to each other.
- 5 is dominated by 2, 3 and 4, not by 1.

Existing Evolutionary MOO Strategies

- A multi-objective optimization algorithm must achieve:
  1. ***Guide the search towards the global Pareto-Optimal front.***
  2. ***Maintain solution diversity in the Pareto-Optimal front.***

- EAs
  - **Search and optimization tools**
  - **Provide near-optimal solutions for complex, hard, multimodal problems**.

- Multiobjective EAs are more popular primarily because of their *population based nature*.

Simulated Annealing and Multiobjective Optimization Problem

- **Simulated Annealing (SA)** is another popular search algorithm
  - utilizes the principles of statistical mechanics, regarding the behavior of a large number of atoms at low temperature, for finding minimal cost solutions to large optimization problems by minimizing the associated energy.

- Only a few attempts at using SA for MOOP
  - SA usually finds **one solution instead of a set of solutions**
  - Difficulty in **computing the acceptance probability**
  - Generally in the SA based MOOP algorithms, the set of

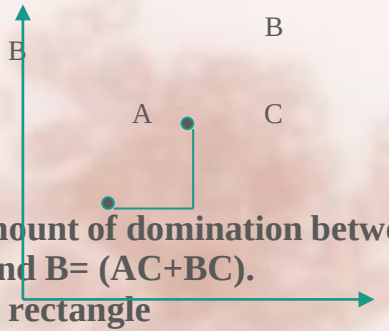
Archived Multiobjective Simulated Annealing Algorithm (AMOSA)

- **AMOSA**
  - **Based on Simulated Annealing.**
  - **Incorporates the concept of an archive where the non-dominated solutions seen so far are stored**
  - **Uses clustering to restrict the size of the archive and to ensure diversity**
  - **Uses *amount of domination* for computing the acceptance probability *depending on domination status of the new solution, current solution and archive***

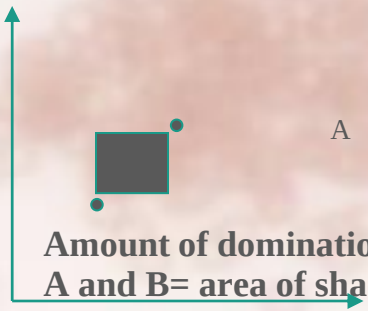
- **Two limits kept on the size of the archive: Hard-limit and Soft-limit.**
- **During the process the non-dominated solutions stored in the archive as and when they are generated until the size of the archive increases to Soft-limit.**
- **There after if more non-dominated solutions are generated, the size of the archive is first reduced to Hard-limit by applying clustering .**

Amount of Domination

Given two solutions a and b, let the amount of domination be Δdom. This Δdom value is used in AMOSA while computing the probability of acceptance of a newly generated solution.



Amount of domination between A and B= (AC+BC). rectangle



Amount of domination between A and B= area of shaded

The AMOSA Algorithm

```
Set Tmax, Tmin, Hard-limit, Soft-limit, iter, alpha.
Set temp=Tmax.
Initialize the archive.
Current-pt=random(archive).
While (temp> Tmin){
  for(i=0; i<iter; i++){
        New-pt=Perturb(Current-pt)
        Decision about Current-pt based on domination status
        (depending on the position of current-pt, new-pt and points of the
     archive)
  }
  temp=alpha*temp
}
If Archive-size > Soft-limit
    Archive=cluster (Archive, Hard-limit)
Output Archive
```

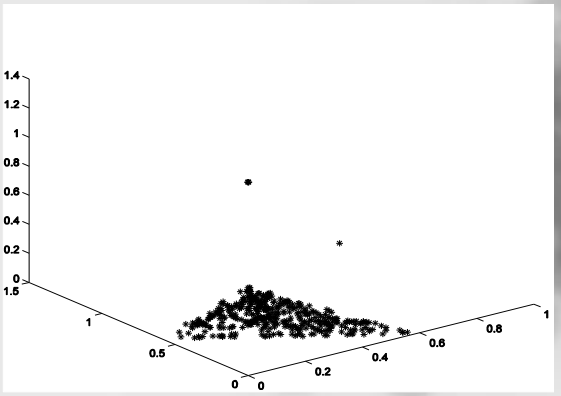
A Simulated Annealing Based Multi-objective Optimization Algorithm: AMOSA , S Bandyopadhyay, S Saha, U Maulik, K Deb, IEEE Transaction on Evolutionary Computation, Volume 12, No. 3, June 2008, Pages 269-283 (citations: 423).

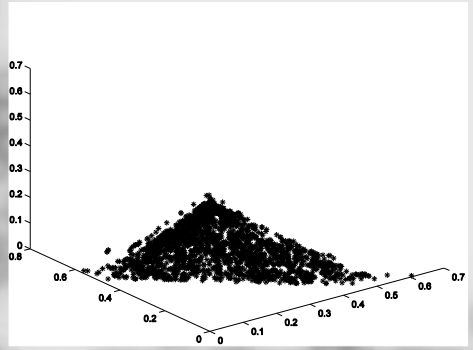
Characteristics and Advantages of AMOSA

- Concept of ***amount of domination*** is used to determine the acceptance probability
- ***Clustering*** used to enforce diversity of solutions.
- In AMOSA a new solution worse than the current solution may be selected.
  - In contrast to most other MOEA's where if a choice needs to be made between two solutions x and y and if x dominates y then x is always selected.
  - leads to reduced possibility of getting stuck at suboptimal regions.
    - Characteristic of single objective EAs or SAs
- All MOEAs are so designed that this characteristics is lost.
- The AMOSA algorithm provides a way of incorporating this feature.
  - good performance for problems where other algorithms got stuck at local optima.

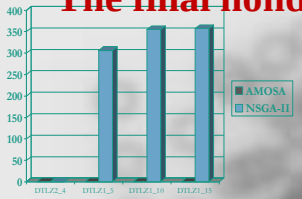
A Simulated Annealing Based Multi-objective Optimization Algorithm: AMOSA , S Bandyopadhyay, S Saha, U Maulik, K Deb, IEEE Transaction on Evolutionary Computation, Volume 12, No. 3, June 2008, Pages 269-283 (citations: 423).

Results with Real Coded AMOSA





The final nondominated front obtained by (a) MOSA and (b) AMOSA for DTLZ1

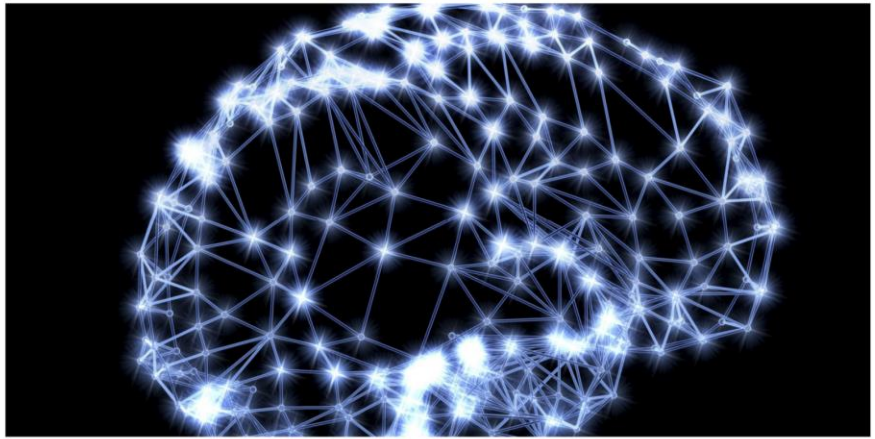


Plots of Convergence measures by AMOSA and NSGA-II for many objective test problems

Basics of Deep Learning

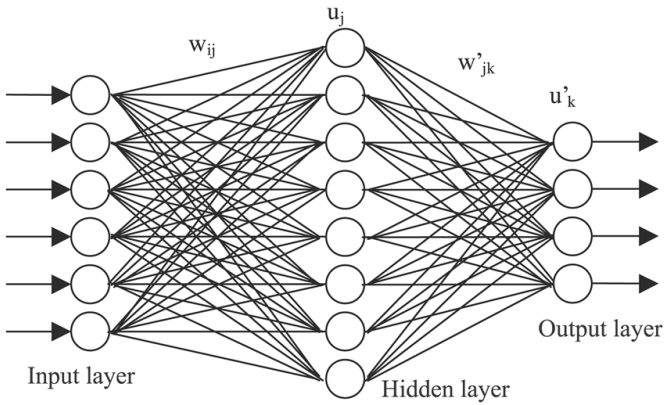
Neural Network

- *Mimics the functionality of a brain.*
- *A neural network is a graph with neurons (nodes, units etc.) connected by links.*



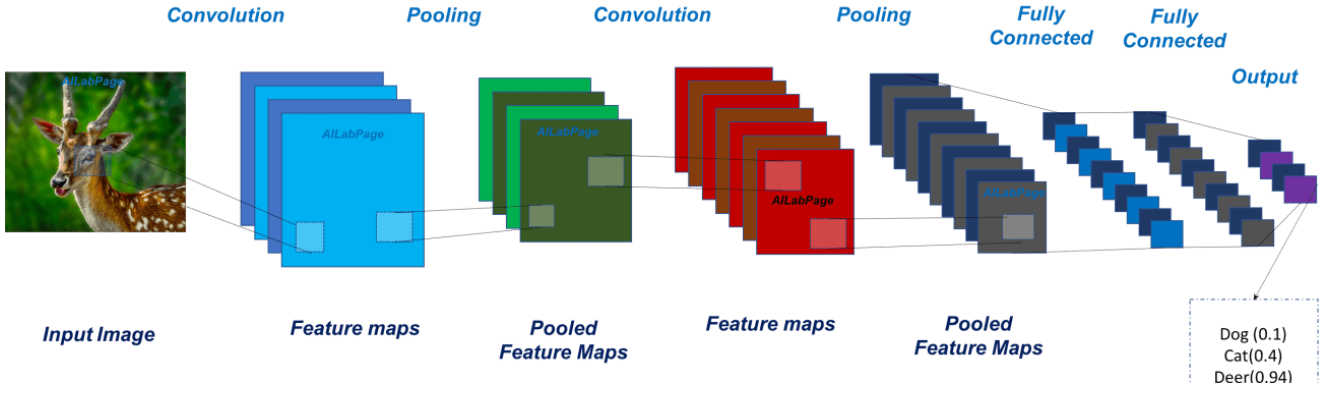
Neural Architecture

- **Feedforward Fully-Connected Neural Network(Also known as MLP)**
  - A fully-connected network (FCN) consists of multiple layers of neurons,
  - Each neuron is connected to every neuron in the previous layer,
  - Each connection has its own weight.
  - It is usually used in supervised learning when labels are provided.



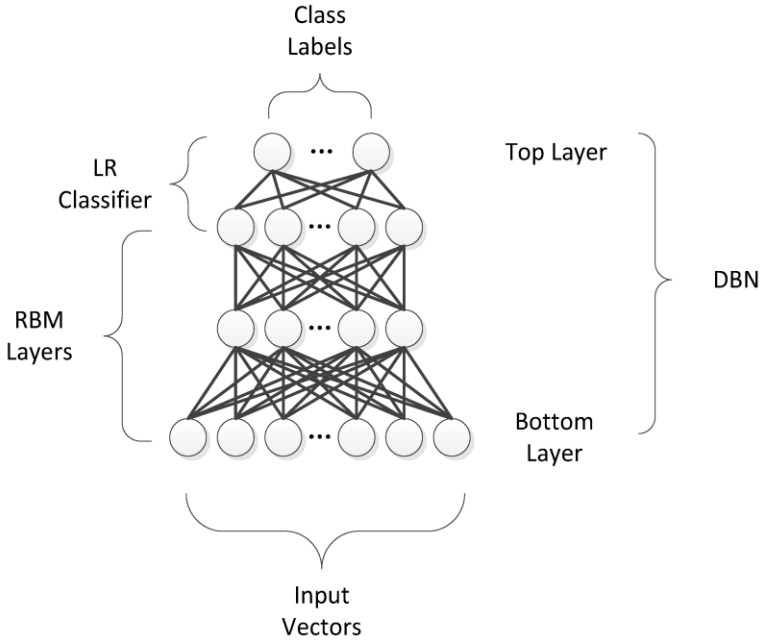
Neural Architecture

- **Convolutional Neural Network(CNN)**
  - Convolutional neural networks (CNNs) were inspired by biological process, in which the connectivity pattern between neurons is inspired by the organization of the animal visual cortex.
  - A ConvNet is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters



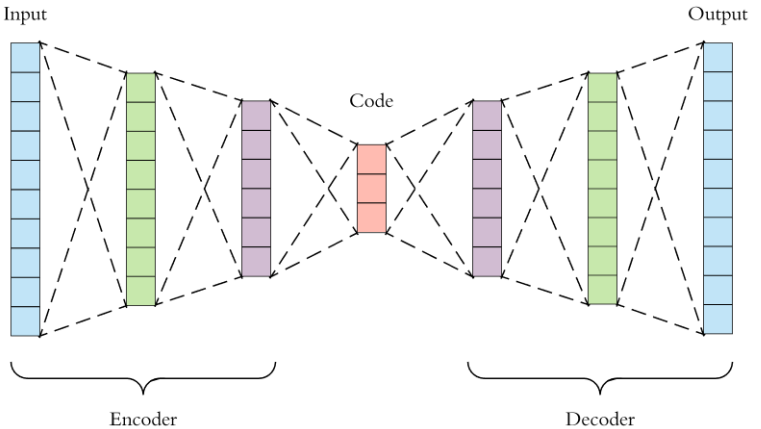
Neural Architecture

- **Deep Belief Network(DBN)**
  - Deep Belief Networks (DBNs) are generative graphical models which learn to extract a deep hierarchical representation of the input data.
  - It is composed of several shallow networks such as restricted Boltzmann machines, such that the hidden layer of each sub-network serves as the visible layer of the next sub-network.
  - The greedy layer-wise unsupervised training is applied to DBNs with RBMs as the building blocks for each layer.



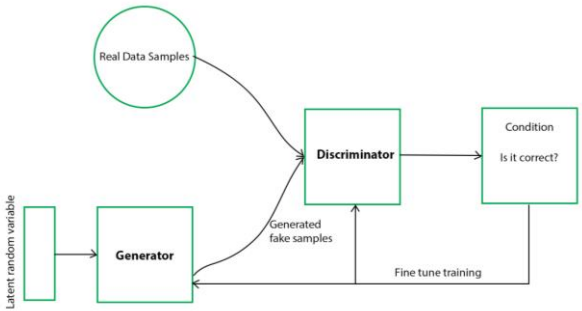
Neural Architecture

- **Autoencoder**
  - Autoencoder (AE) is one of the most significant algorithms in unsupervised representation learning.
  - It is a powerful method to train a mapping function, which ensures the minimum reconstruction error between coder layer and data layer.
  - Since the hidden layer usually has smaller dimensionality than the data layer, it can help find the most salient features of data.

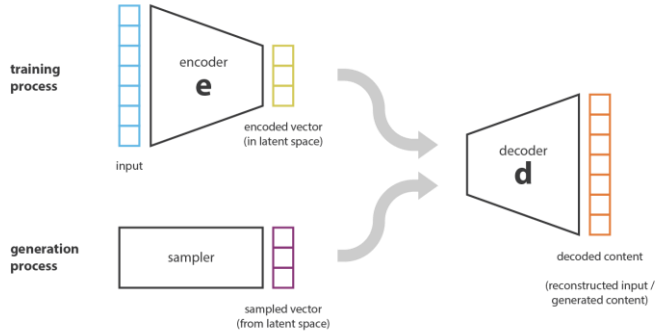


Neural Architecture

- **GAN & VAE**
  - Generative Adversarial Network (GAN) and Variational Autoencoder (VAE) are the most powerful frameworks for deep generative learning.
  - GAN aims to achieve an equilibrium between a generator and a discriminator, while VAE attempts to maximizing a lower bound of the data log-likelihood.
  - A series of model extensions have been developed for both GAN and VAE. Moreover, they have also been applied to handle clustering tasks.



Generative Adversarial Network



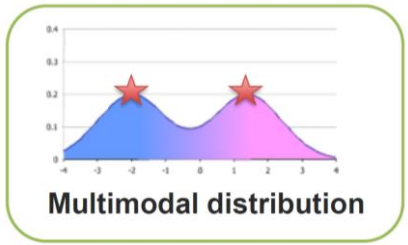
Variational Autoencoder

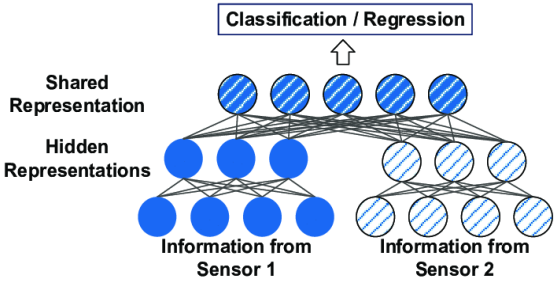
Concepts of Multi-modal Approaches

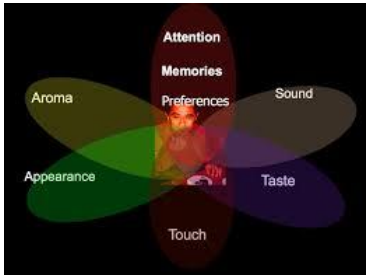
- **Introduction**
  - What is Multimodal?
    - Definiation, multimodal vs multimedia
  - Why multimodal
    - Multimodal applications: image captioning, video description, Audio Video Speech Recognition(AVSR)
  - Core technical challenges
    - Representation learning, translation, alignment, fusion and co-learning

Introduction: Multi-modal Approaches

- Introduction
  - Involves more than modalities and representations







  - Multiple modes, i.e., distinct “peaks” (local maxima) in the probability density function

Introduction: Multi-modal Approaches

- **Modality**
  - The way in which something happens or is experienced
    - Modality refers to a certain type of information and/or the representation format in which information is stored.
    - Sensory modality: one of the primary forms of sensation, as vision or touch; channel of communication.
    - Learning from multimodal sources offers the possibility of capturing correspondences between modalities and gaining an in-depth understanding of natural phenomena

Introduction: Multi-modal Approaches

- **Examples of Multi-modal instances**
  - Natural language processing (both spoken or written)
  - Visual (from images or videos)
  - Auditory (including voice, sounds and music)
  - Protein Function Prediction(Protein Sequence and Structure)
  - Biomedical Text Mining(Text, Genomic Sequence and Structure)
  - Visual(Body Language, Eye Movement, Head Gestures)

Multi-modal Approaches

- **Technical Challenges of Multi-modal instances**
  - Representation:
    - *Learning how to represent and summarize multimodal data in a way that exploits the complementarity and redundancy of multiple modalities*
  - Translation
    - *How to translate (map) data from one modality to another.*
    - *Not only is the data heterogeneous, but the relationship between modalities is often open-ended or subjective*

Multi-modal Approaches

- **Technical Challenges of Multi-modal instances**
  - Alignment:
    - *Identify the direct relations between (sub)elements from two or more different modalities.*
  - Fusion
    - *How to integrate information from two or more modalities to perform a prediction*
  - Co-learning
    - *Transfer knowledge between modalities, their representation, and their predictive models.*

Technical Challenges

	Challenges				
Applications	Representation	Translation	Alignment	Fusion	Co-learning
Speech recognition and synthesis					
Audio-visual speech recognition	✓		✓	✓	✓
(Visual) speech synthesis	✓	✓			
Event detection					
Action classification	✓			✓	✓
Multimedia event detection	✓			✓	✓
Emotion and affect					
Recognition	✓		✓	✓	✓
Synthesis	✓	✓			
Media description					
Image description	✓	✓	✓		✓
Video description	✓	✓	✓	✓	✓
Visual question-answering	✓		✓	✓	✓
Media summarization	✓	✓		✓	
Multimedia retrieval					
Cross modal retrieval	✓	✓	✓		✓
Cross modal hashing	✓				✓

Introduction to Computational Biology

- Covers both biology and computational aspects
- This research field act as a bridge between the gap of biological science and computer science
- Recently advancement of the AI helps to better understanding of the computational biology
- Various machine learning and deep learning techniques are utilized to analysis the underlying biological datasets

Examples Computational Biology

- Gene clustering
  - Grouping the genes on the basis of the gene expression profile
  - In this regard, along with the traditional clustering technique, advance clustering technique(MOO-based, ensemble-based) is used
  - While clustering, various fitness functions are optimized. Fitness function can be
    - Statistical aspects: Euclidean distance, Manhattan distance
    - Biological aspects: Biological Similarity, Biological relevance

Examples Computational Biology

- Identifying Protein-protein Interactions
  - Protein-protein interaction (PPI) plays essential roles in cellular functions.
  - Most computational methods are designed to predict whether two proteins interact but not their interacting
  - As protein interactions generally occur via domains instead of the whole molecules, predicting domain-domain interaction (DDI) is an important step toward PPI prediction
  - For predicting the interactions, researchers have utilized the protein sequence and the structures

Examples Computational Biology

- Protein Function Prediction
  - Crucial for understanding the cellular mechanisms, identifying disease-causing functional changes in genes/proteins, and understanding for disease prevention, diagnosis, and treatment
  - Computational approaches are developed that can efficiently predicts gene/protein function prediction
  - Various biological datasets are used
    - Protein genomic sequence
    - Protein Structure
    - Gene Ontology

Examples Computational Biology

- Protein Structure Prediction
  - Protein structure prediction is a longstanding challenge in computational biology
  - The prediction of inter-residue contacts and distances from coevolutionary data using deep learning has considerably advanced protein structure prediction
  - Recently, computational biologists trying to prediction efficiently the protein structure from the protein sequence

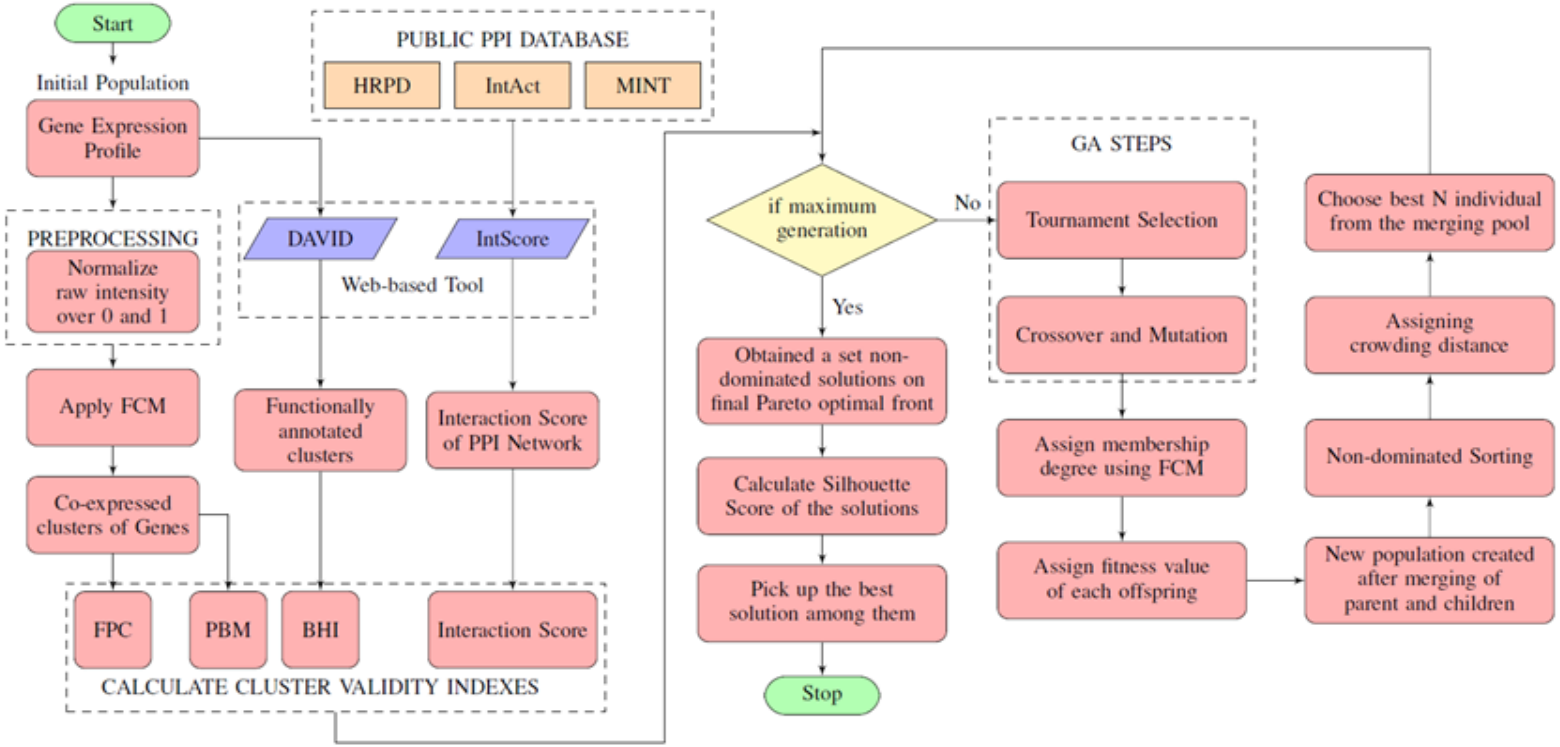
Examples Computational Biology

- Biomedical Text Mining
    - Several databases have been manually curated to cache protein interaction information such as MINT, BIND, and SwissProt in structured and standard formats
    - The rapid growth of biomedical literature has shown a significant gap between the availability of biomedical article and its automatic curation
    - In the last 20 years, the overall size of biomedical corpus has increased at a exponential compounded annual growth rate
    - This has lead to a surge in the interest of Biomedical Natural Language Processing (BioNLP) community for automatic detection and extraction of PPI information.

Application of ML and DL in Computational Biology

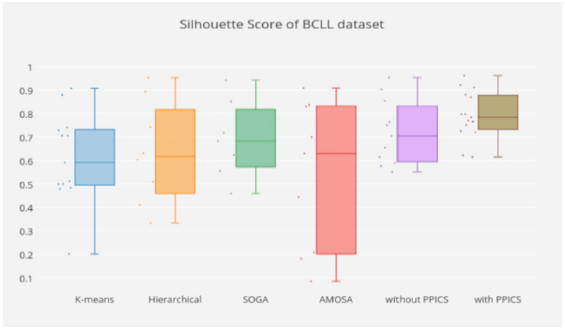
- With the advancement of the machine learning and deep learning, researchers are utilizing those for solving various computational biology problems
- Few applications are
    - Multi-objective optimization-based Gene clustering technique
    - Ensemble –based gene clustering technique
    - Deep learning-based for Genome Sequence Analysis
    - Deep Learning-based Protein Structure Prediction
    - Deep Learning for Biological Image Analysis
    - ML/DL-based Biological Text mining

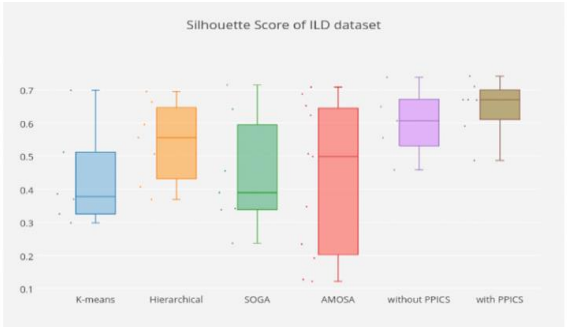
MOO-based Gene Clustering



MOO-based Gene Clustering

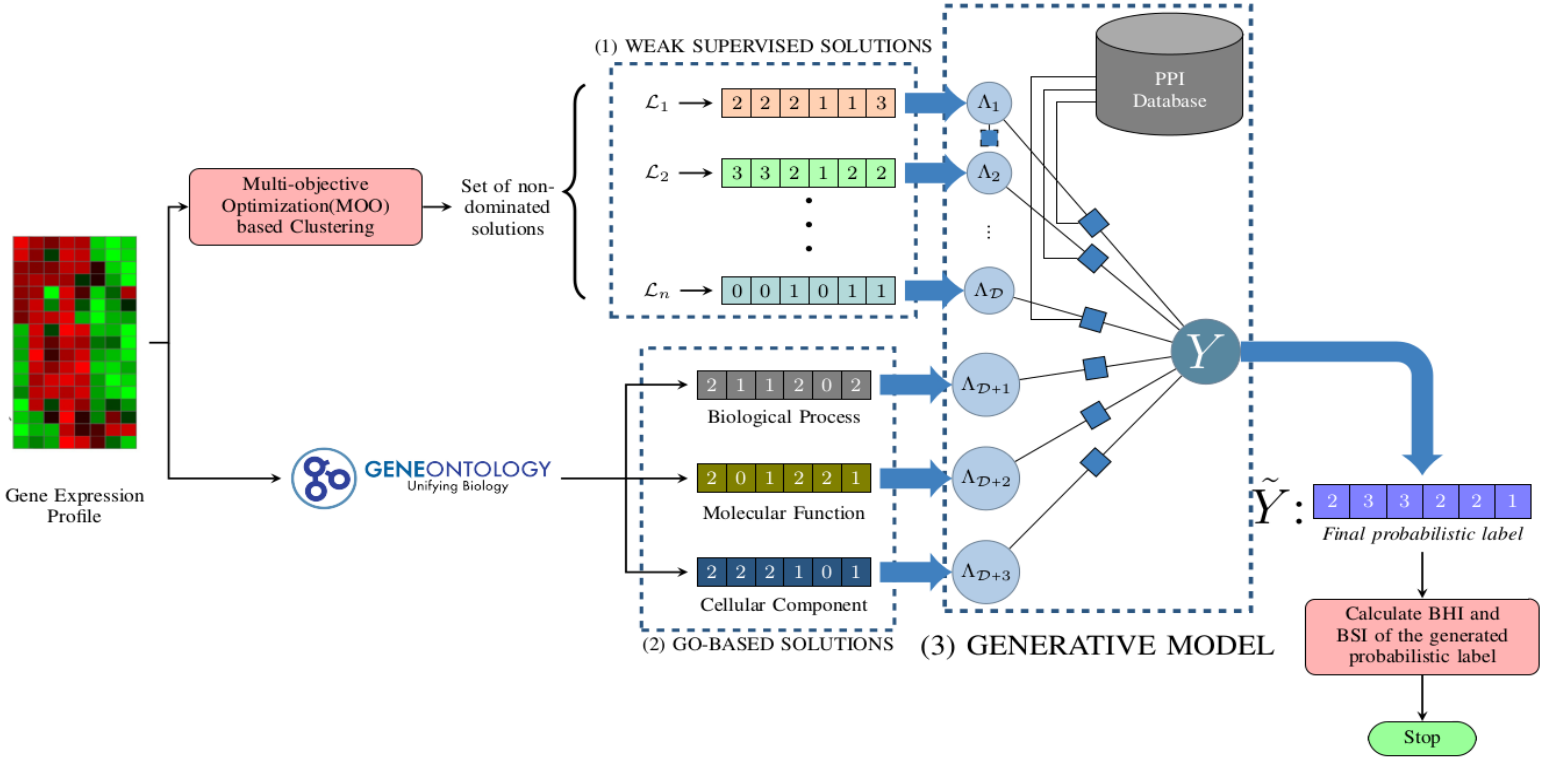
Datasets	K-means		Hierarchical		SOGA		AMOSA		DBSCAN		Prosposed Method			
											without PPICS		with PPICS	
	K	s(C)	K	s(C)	K	s(C)	K	s(C)	K	s(C)	K	s(C)	K	s(C)
B-CLL chronic lymphocytic leukemia	2	0.8792	2	0.9519	2	0.8970	6	0.9093	3	0.4041	2	0.9526	**2**	**0.9617**
ILD Interstitial lung disease	2	0.6987	2	0.6951	2	0.7142	3	0.7082	2	0.3361	2	0.7375	**2**	**0.7412**
Waldenströms macroglobulinemia	2	0.4808	2	0.4579	2	0.5147	5	0.4106	5	0.4094	2	0.5319	**2**	**0.5618**





P. Dutta and S. Saha. *Fusion of expression values and protein interaction information using multi-objective optimization for improving gene clustering*. Computers in Biology and Medicine, volume 89, pages 31–43. Elsevier, 2017.

Ensemble-based Gene Clustering



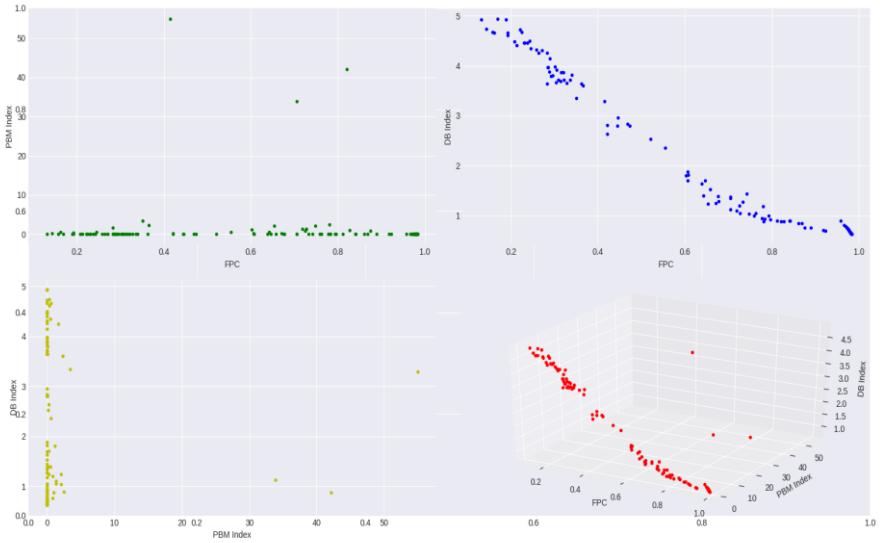
Ensemble-based Gene Clustering

Comparative Table of BHI

	B-CLL	ILD	Prostrate
K-means	0.163	0.395	0.379
DBSCAN	0.193	0.417	0.396
MODE	0.236	0.421	0.406
Best MOO-based solution	0.236	0.428	0.410
Ensemble Technique(MOO-based solution)	0.240	0.427	0.410
Ensemble Technique(MO+GM)	0.331	0.453	0.445
Ensemble Technique(MO+PPI+GM)	0.345	0.460	0.448
Ensemble Technique(MO+PPI+GO+GM)	**0.361**	**0.475**	**0.451**

Comparative Table of BSI

	B-CLL	ILD	Prostrate
K-means	0.934	0.860	0.884
DBSCAN	0.986	0.839	0.879
MODE	0.987	0.905	0.892
Best MOO-based solution	0.989	0.908	0.902
Ensemble Technique(MOO-based solution)	0.989	0.926	0.935
Ensemble Technique(MO+GM)	0.989	0.936	0.941
Ensemble Technique(MO+PPI+GM)	0.992	0.938	0.944
Ensemble Technique(MO+PPI+GO+GM)	**0.994**	**0.941**	**0.945**



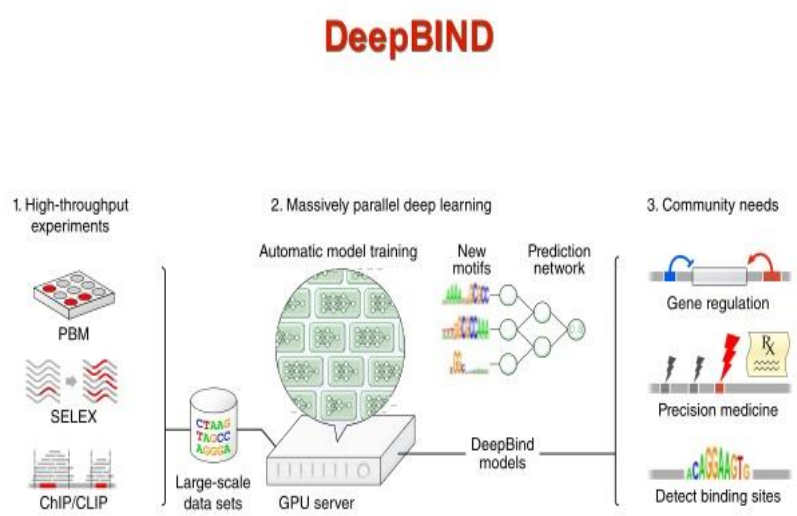
P. Dutta, S. Saha, S. Pai, and A. Kumar. *A protein inter-action information-based generative model for enhancing gene clustering*. **Scientific Reports**, **Nature Publishing Group**, volume 10, pages1–12, 2020.

P. Dutta and S. Saha. *A weak supervision technique with a generative model for improved gene clustering*. In **IEEE Congress on Evolutionary Computation (CEC)**, pages 2521–2528, 2019.

Deep learning-based for Genome Sequence Analysis

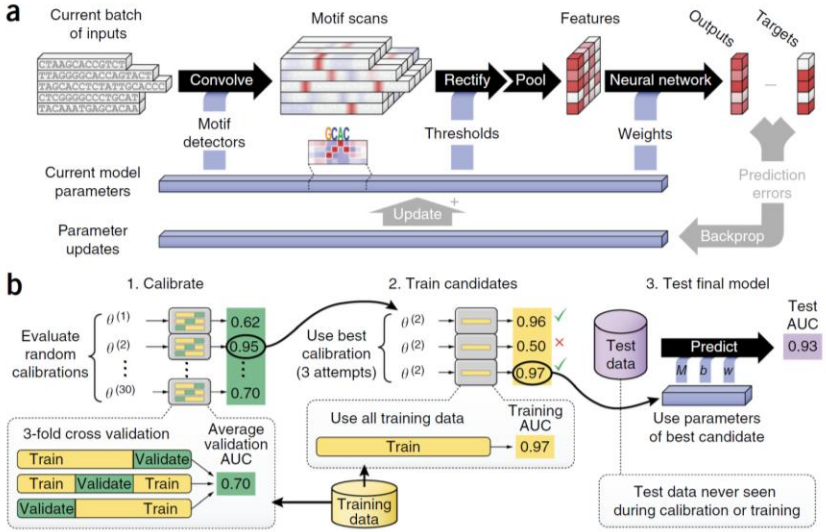
- Advances in sequencing technology have led to a large and rapidly increasing amount of genetic and protein sequences,
- The sequence can nucleotide sequence(A,T G and C) or amino acid sequence which contain crucial information for phylogenetics and evolutionary biology
- Genome sequence is widely utilized by researchers for solving popular biomedical problems

Predicting Specificities of DNA- and RNA-binding Proteins



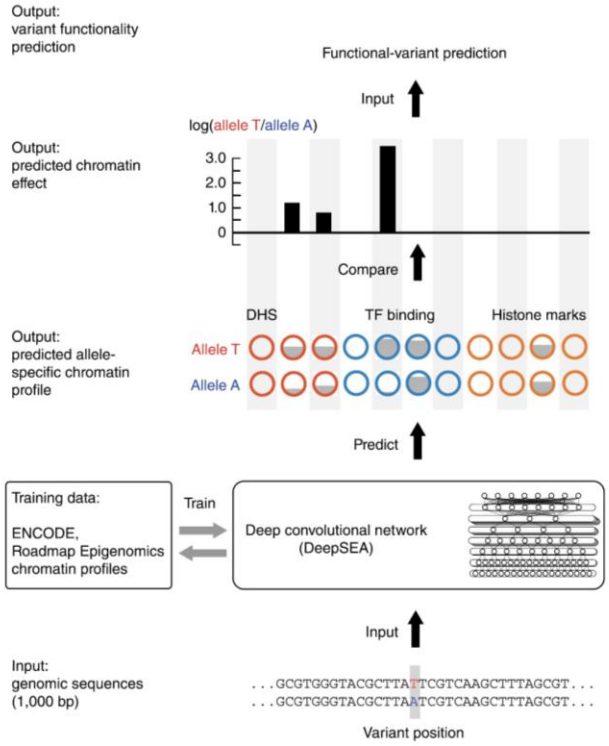
Alipanahi et al. *Nat Biotech* 2015

(Other methods: DeepSEA — Zhou & Troyanskaya, *Nat Methods* 2015; DanQ — Quang & Xie, *Nucleic Acids Res* 2016)



Predicting Effects of Noncoding Variants

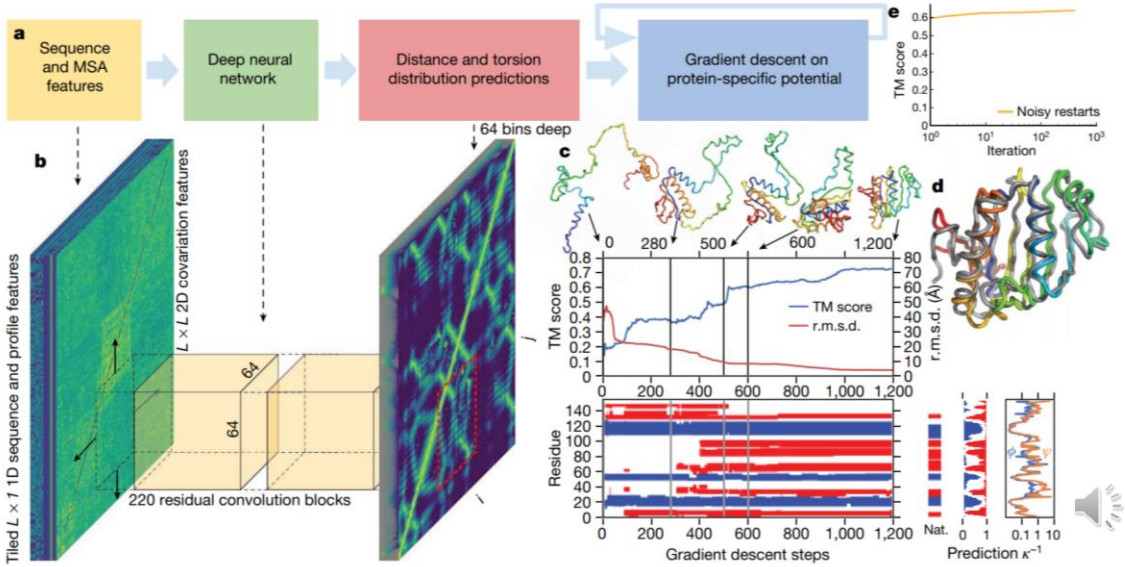
- Identifying functional effects of noncoding variants is a major challenge in human genetics.
- DeepSEA (deep learning–based sequence analyzer) use convolutional neural network
- For training the model from genomic sequence by simultaneously predict large-scale chromatin-profiling data, including TF binding, DNase I sensitivity and histone-mark profiles
- Three major features
  - *integrating sequence information from a wide sequence context*
  - *learning sequence code at multiple spatial scales with a hierarchical architecture*
  - *multitask joint learning of diverse chromatin factors sharing predictive features*



Zhou et. al. ***Nature Methods(2015)***

AlphaFold: Protein Structure Prediction

- Protein structure prediction can be used to determine the three-dimensional shape of a protein from its amino acid sequence
- Train a neural network to make accurate predictions of the distances between pairs of residues
- Predict the structure itself accurately by minimizing the potential by gradient descent
- The neural network predictions include backbone torsion angles and pairwise distances between residues
- Both the torsion angle and the distance provide more specific information for protein structure



Senior et. al. ***Nature (2019)***

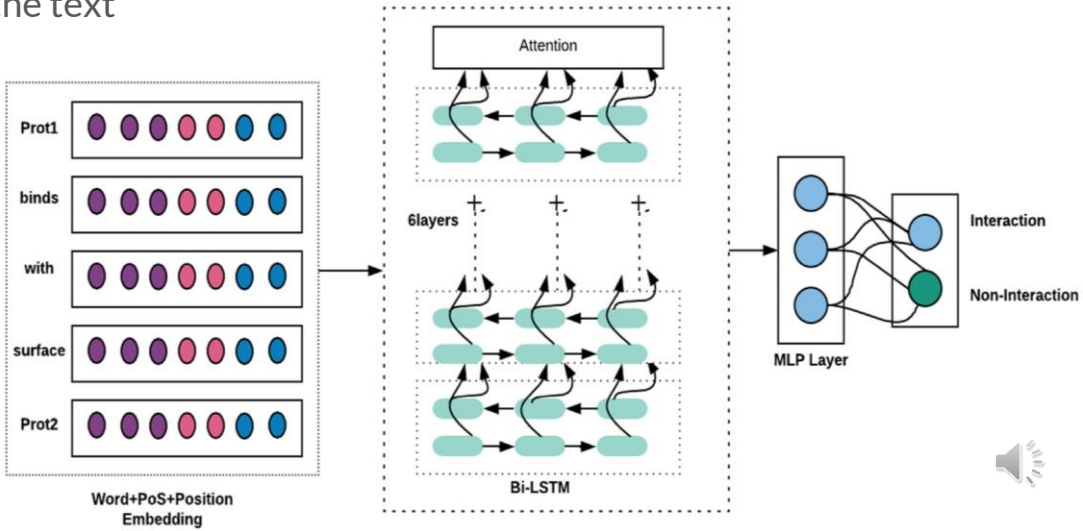
Deep Learning Fundus Image Analysis for Diabetic Retinopathy

- Diabetes is a globally prevalent disease that can cause diabetic retinopathy and macular edema in the human eye retina
- To extract features from the retinal image, convolutional neural network is used
    - Inception-v3 architecture that was pretrained on ImageNet dataset
- Provide novel results for five different screening and clinical grading systems for diabetic retinopathy and macular edema classification

Sahlsten et. al(2019), *Deep Learning Fundus Image Analysis for Diabetic Retinopathy and Macular Edema Grading*, **Scientific Reports**

Deep Learning-based Biomedical Text Mining

- The exponential growth of the biomedical corpus has lead to a surge in the interest of Biomedical Natural Language Processing (BioNLP) community for automatic detection and extraction of PPI information
- A novel method based on attentive deep RNN, which combines multiple levels of representations exploiting word sequences and dependency path related information to identify protein–protein interaction (PPI) information from the text
- Leverages joint modeling of proteins and relations in a single unified framework, which is named as the 'Attentive Shortest Dependency Path LSTM'
- five popular benchmark PPI datasets, namely AiMed, BioInfer, HPRD50, IEPA, and LLL



Yadav et. al(2019), *Feature assisted stacked attentive shortest dependency path based Bi-LSTM model for protein–protein interaction,* **Knowledge-Based Systems**

Deep Multi-modal Approach for Computational Biology

- Multi-omics data is popular due its insightful information
- Analyzing the relation between the modalities helps to get more comprehensive overview of the physiological information
- Proper deep learning models are required for extracting the information
- Amalgamation of all the extracted features

Deep Multi-modal Approach for Computational Biology

- Various modalities

GENE EXPRESSION PROFILE

ID_REF	IDENTIFIER	GSM45021	GSM45022	GSM45023	GSM45024	GSM45025	GSM45066	GSM45067	GSM45068	GSM45069	GSM45070
1000_at	MAPK3	357.7	371.2	404.6	330.7	333.2	241.7	284.6	292.4	268.3	324.2
1001_at	TIE1	20.9	18.5	10.8	13.5	20.5	10.7	3.7	12.8	25.4	23.1
1002_f_at	CYP2C19	4.7	18.9	17.6	2.6	5.5	3.3	3.2	1.8	9.8	22.1
1003_s_at	CXCR5	174.1	231.7	214.1	201.6	314.5	8.4	7.2	204.6	22.3	109.3
1004_at	CXCR5	257	194.3	221.5	185.4	376.4	355.9	176.8	244.2	186.2	743
1005_at	DUSP1	732.9	499.8	1092.3	923.6	287.5	137.9	1033.4	384.2	183.1	184
1006_at	MMP10	1.1	2.2	11.7	2.6	15.1	1.1	1.1	2.8	3.3	8.6
1007_s_at	DDR1	109.7	174.6	185.9	109.1	215.8	161.9	196.5	162	156.7	140.2
1008_f_at	EIF2AK2	887.2	551.8	2038.1	653.5	2216.7	1581.3	677.5	850.3	3090.5	1054.8
1009_at	HINT1	1145.5	1014.9	1103.1	850.4	886.6	859.3	1228.6	1231.9	1118.3	762.1
100_g_at	RABGGTA	169	200.6	196.3	151.4	167.8	111.3	164.9	240.6	155.9	215.2
1010_at	MAPK11	8.5	10.3	6.6	3.4	11.6	8	7.5	5.7	13.2	25.9
1011_s_at	YWHAE	25.9	31.7	43.3	40.9	40.8	26.7	45.1	62	60.4	63.4
1012_at	KAT2B	28	34.4	13.2	28.3	15.7	56.5	35.2	25	17.1	40.8
1013_at	SMAD5	17.2	3	6	9.9	10	16.7	1.5	10	2.9	9.3
1014_at	POLG	225.4	316.5	305.7	333.1	257.8	268.2	254.7	365.6	263	407.4
1015_s_at	LIMK1	13.9	23	7.8	34.4	18.6	10.1	10	21.7	44	7
1016_s_at	IL13RA2	2.2	10.1	2	2.3	3.1	2.8	11.1	8	3.8	7
1017_at	MSH6	37.9	31.2	26.4	33.4	24.2	21.4	22	17	19	13.4
1018_at	WNT10B	20.4	50.2	47.4	33	39.7	20.4	27.2	27.2	23.3	13.9

(http://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS1388)

PROTEIN-PROTEIN INTERACTION NETWORK

A	B	C
Interactor 1	**Interactor 2**	**Confidence Score**
ALDH1A1	ALDH1A1	NA
ITGA7	CHRNA1	0.362686541
PPP1R9A	ACTG1	0.375744512
SRGN	CD44	0.170231959
GRB7	ERBB2	0.999999942
PAK1	ERBB2	0.750312001
DLG4	ERBB2	1
PIK3R2	ERBB2	1
PTPN18	ERBB2	0.999955255
ERBB2IP	ERBB2	0.357821626
SMURF2	ARHGAP5	0.354840552
NF2	ERBB2	0.999927885
CD82	ERBB2	1
ERRFI1	ERBB2	0.872512159
MMP7	CD44	0.097916472
TOB1	ERBB2	0.037307944
MUC4	ERBB2	0.950066383
PICK1	ERBB2	0.670312255
SMURF2	TXNIP	0.354840552

(http://hprd.org/download)

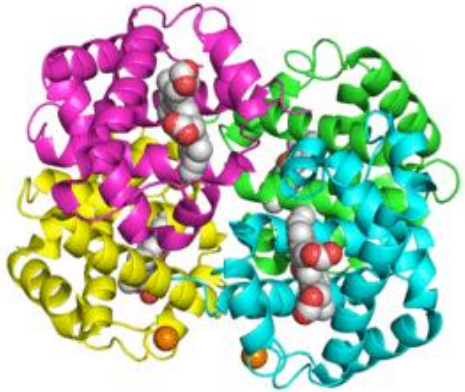
Deep Multi-modal Approach for Computational Biology

- Various modalities

PROTEIN SEQUENCE

Ensembl(https://asia.ensembl.org/index.html)

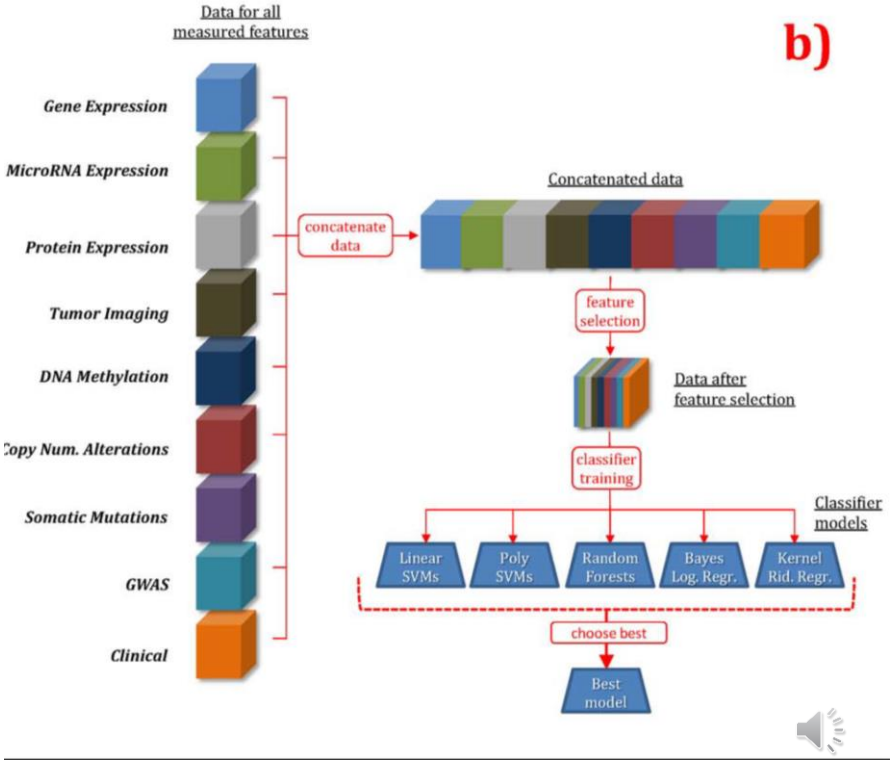
PROTEIN 3D STRUCTURE



RCSB PDB Database(https://www.rcsb.org/)

Information Content and Analysis Methods for Multi-omics data

- Analysis various modalities of omics data using computational methods
- The extracted information is finally concatenated to perform the final prediction
- 47 datasets/predictive tasks that in total span over 9 data modalities and executed analytic experiments for predicting various clinical phenotypes and outcomes
- The integrated information is more crucial than any single modality

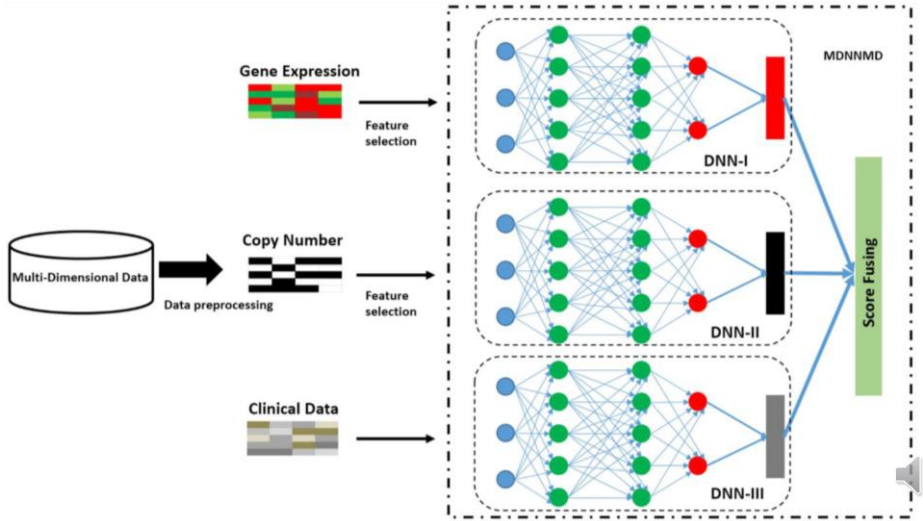


Ray et. al(2020), *Information content and analysis methods for Multi-Modal High-Throughput Biomedical Data* **Scientific Reports**

Deep Multi-modal for Breast Cancer Prognosis

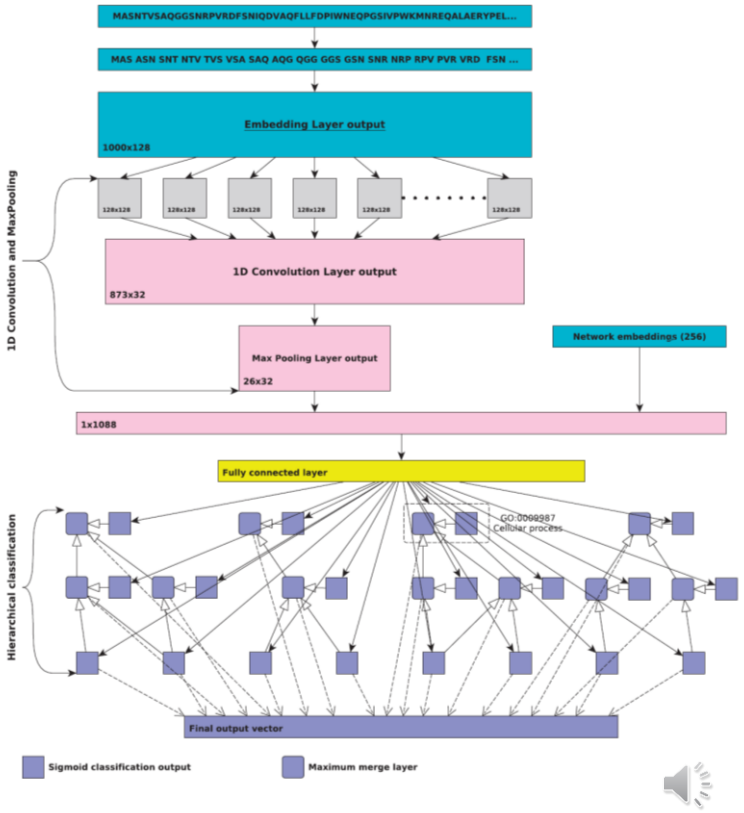
- Accurate prognosis prediction of breast cancer can spare a significant number of patients from receiving unnecessary treatment
- A Multi modal Deep Neural Network by integrating Multi-dimensional Data (MDNNMD) for the prognosis prediction of breast cancer.
- Gene expression profile , Copy number alteration (CNA) profile and clinical data are considered as various modalities
- Extracted features are integrated using score level fusion for the final prediction model

Sun et. al(2018), *A multi modal deep neural network for human breast cancer prognosis prediction by integrating multi multi-dimensional data*, **IEEE/ACM TCBB**



DeepGO: Predicting Protein Functions from Protein Sequence and Interactions

- Deep learning models are utilized to learn features from protein sequences as well as a cross-species protein–protein interaction network.
- Outputs information in the structure of the GO and utilizes the dependencies between GO classes as background information to construct a deep learning model.
- Evaluate the method using the standards established by the Computational Assessment of Function Annotation (CAFA)
- Demonstrate a significant improvement over baseline methods such as BLAST, in particular for predicting cellular locations.

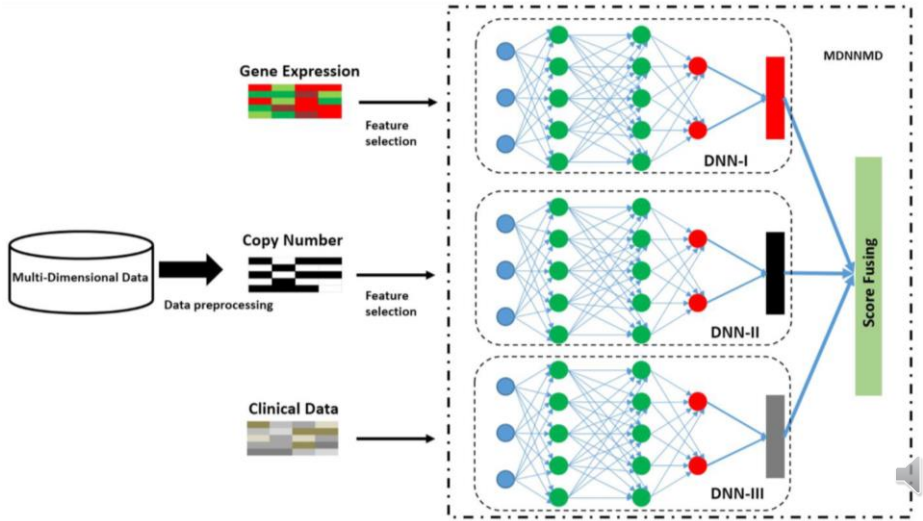


Kulmanov et. al(2017), *DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier*, **Bioinformatics**

Deep Multi-modal for Breast Cancer Prognosis

- Accurate prognosis prediction of breast cancer can spare a significant number of patients from receiving unnecessary treatment
- A Multi modal Deep Neural Network by integrating Multi-dimensional Data (MDNNMD) for the prognosis prediction of breast cancer.
- Gene expression profile , Copy number alteration (CNA) profile and clinical data are considered as various modalities
- Extracted features are integrated using score level fusion for the final prediction model

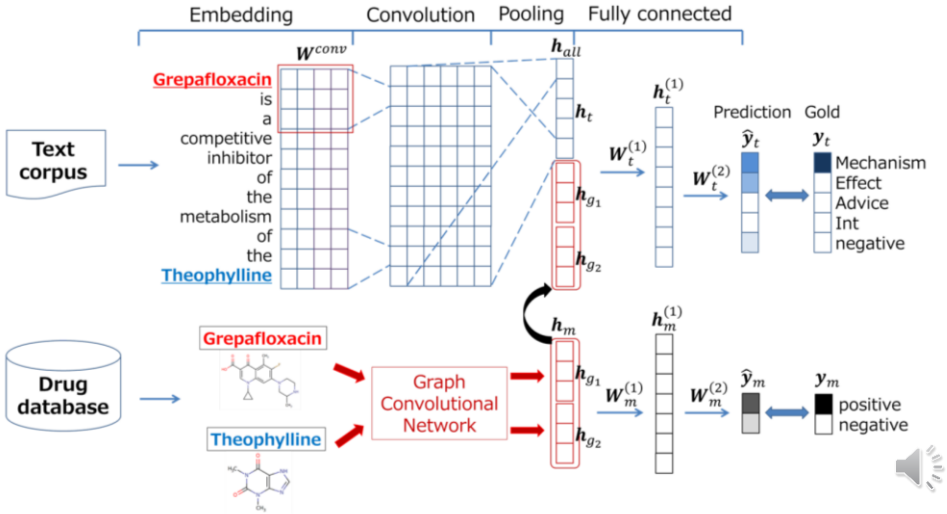
Sun et. al(2018), *A multi modal deep neural network for human breast cancer prognosis prediction by integrating multi multi-dimensional data*, **IEEE/ACM TCBB**



Enhancing Drug-drug Interaction from text and Molecular Structure

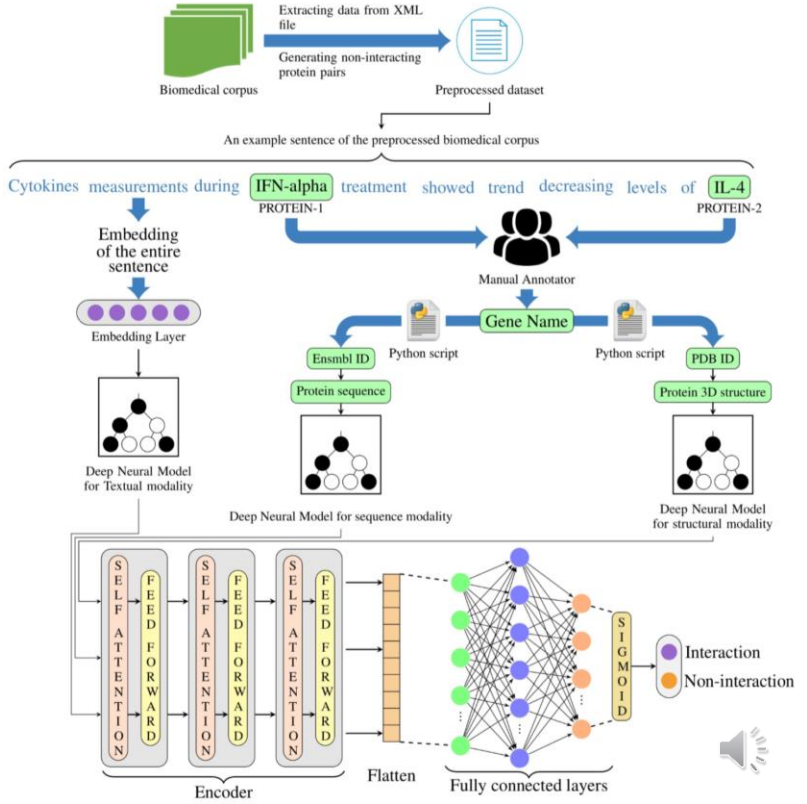
- Propose a novel neural method to extract drug-drug interactions (DDIs) from texts using external drug molecular structure information
- Encode textual drug pairs with convolutional neural networks
- Molecular pairs with graph convolutional networks (GCNs)
- concatenate the outputs of these two networks
- In the experiments, GCNs can predict DDIs from the molecular structures of drugs in high accuracy
- The molecular information can enhance text-based DDI extraction

Asada et. al(2018), *Enhancing Drug-Drug Interaction Extraction from Texts by Molecular Structure Information* **ACL**



Identifying Protein Interactions by Amalgamating Protein Sequence, Structure and Text

- Exemplified two popular benchmark PPI corpora(BioInfer and HRPD50) in multi-modal scenario
- Besides existing textual modalities, 3D protein structure and underlying genomic sequence of each proteins are added to each instance
- Utilized graph convolutional neural network for capturing the atomic representation of the protein structure
- Finally, self attention-based multi-modal architecture is designed to predict protein interactions



Dutta et. al(2020), *Amalgamation of protein sequence, structure and textual information for improving protein-protein interaction identification,* **ACL**

Future Work and Conclusion

- More efficient models for analyzing biological modalities
  - Protein structure is 3D and unsymmetrical
  - Time efficient model for understanding long protein sequences (>5000 nucleotides)
- Improved integration technique for multi-modal approach
  - attention mechanism
- Addressing incomplete multi-modal instances
  - GAN, shape Boltzmann machine

Acknowledgement

- Pratik Dutta acknowledges Visvesvaraya PhD Scheme for Electronics and IT, an initiative of Ministry of Electronics and Information Technology (MeitY), Government of India for fellowship support.
- Dr. Sriparna Saha gratefully acknowledges the Young Faculty Research Fellowship (YFRF) Award, supported by Visvesvaraya PhD scheme for Electronics and IT, Ministry of Electronics and Information Technology (MeitY), Government of India, being implemented by Digital India Corporation (formerly Media Lab Asia) for carrying out this research.

REFERENCES

- A. Ghosh, B.C. Dhara, R.K. De, Selection of genes mediating certain cancers, using a neuro-fuzzy approach, Neurocomputing 133 (2014) 122–140.
- P.G. Sun, The human drug–disease–gene network, Inf. Sci. 306 (2015) 70–80.
- Louis Philippe, Morency Tadas Baltrusaitis , "Multimodal Machine Learning: Integrating Language, Vision and Speech", ACL, 2017.
- S. Paul, P. Maji, Gene expression and protein–protein interaction data for identification of colon cancer related genes using f-information measures, Nat. Comput. 15 (2016) 449–463.
- D. Santoni, A. Swiercz, A. Z_mienko, M. Kasprzak, M. Blazewicz, P. Bertolazzi, G. Felici, An integrated approach (cluster analysis integration method) to combine expression data and protein–protein interaction networks in agrigenomics: application on arabidopsis thaliana, Omics A J. Integr. Biol. 18 (2014) 155–165.
- P. Jia, S. Zheng, J. Long, W. Zheng, Z. Zhao, Dmgwas: dense module searching for genome-wide association studies in protein–protein interaction networks, Bioinformatics 27 (2010) 95–102.
- Y. Li, J. Li, Disease gene identification by random walk on multigraphs merging heterogeneous genomic and phenotype data, BMC Genomics 13 (2012) S27.
- P. Maji, E. Shah, S. Paul, Relsim: an integrated method to identify disease genes using gene expression profiles and ppin based similarity measure, Inf. Sci. 384 (2017) 110–125.
- S. Datta, S. Datta, Methods for evaluating clustering algorithms for gene expression data using a reference set of functional classes, BMC Bioinforma. 7 (2006) 397.
- H. Jeong, S. P. Mason, A.-L. Barabasi, and Z. N. Oltvai, "Lethality and centrality in protein networks," arXiv preprint cond-mat/0105306, 2001.
- S.-H. Yook, Z. N. Oltvai, and A.-L. Barab´asi, "Functional and topological characterization of protein interaction networks," Proteomics, vol. 4, no. 4, pp. 928–942, 2004.
- R. R. Vallabhajosyula, D. Chakravarti, S. Lutfeali, A. Ray, and A. Raval, "Identifying hubs in protein interaction networks," PloS one, vol. 4, no. 4, p. e5344, 2009.

REFERENCES

- C. Tsourakakis, F. Bonchi, A. Gionis, F. Gullo, and M. Tsiarli, "Denser than the densest subgraph: extracting optimal quasi-cliques with quality guarantees," in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013, pp. 104–112.
- X. Ma and L. Gao, "Predicting protein complexes in protein interaction networks using a core-attachment algorithm based on graph communicability," Information Sciences, vol. 189, pp. 233–254, 2012.
- H. Hu, X. Yan, Y. Huang, J. Han, and X. J. Zhou, "Mining coherent dense subgraphs across massive biological networks for functional discovery," Bioinformatics, vol. 21, no. suppl 1, pp. i213–i221, 2005.
- A. V. Goldberg, Finding a maximum density subgraph. University of California Berkeley, CA, 1984
- K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: Nsga-ii, IEEE Trans. Evol. Comput. 6 (2002) 182–197.
- B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of dna-and rna-binding proteins by deep learning," Nature biotechnology, vol. 33, no. 8, p. 831, 2015.
- S. Yadav, A. Ekbal, S. Saha, A. Kumar, and P. Bhattacharyya, "Feature assisted stacked attentive shortest dependency path based bi-lstm model for protein–protein interaction," Knowledge-Based Systems, vol. 166, pp. 18–29, 2019.\
- Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, p. 436, 2015.
- Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in Advances in neural information processing systems, 2007, pp. 153–160.
- G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural computation, vol. 18, no. 7, pp. 1527–1554, 2006.
- P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, and L. Zhang, "Bottom-up and top-down attention for image captioning and visual question answering," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6077–6086.
- K. Sohn, W. Shang, and H. Lee, "Improved multimodal deep learning with variation of information," in Advances in Neural Information Processing Systems, 2014, pp. 2141–2149

REFERENCES

- Kulmanov, M. A. Khan, and R. Hoehndorf, “Deepgo: predicting protein functions from sequence and interactions using a deep ontology aware classifier,” Bioinformatics, vol. 34, no. 4, pp. 660–668, 2017.
- A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani, “Global protein function prediction from protein-protein interaction networks,” Nature biotechnology, vol. 21, no. 6, p. 697, 2003.
- M. Deng, K. Zhang, S. Mehta, T. Chen, and F. Sun, “Prediction of protein function using protein–protein interaction data,” Journal of computational biology, vol. 10, no. 6, pp. 947–960, 2003.
- G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” Neural computation, vol. 18, no. 7, pp. 1527–1554, 2006.
- G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm fordeep belief nets,” Neural computation, vol. 18, no. 7, pp. 1527–1554, 2006.
- K. Cho, B. Van Merri¨enboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using rnn encoder-decoder for statistical machine translation,” arXiv preprint arXiv:1406.1078, 2014.
- D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen et al., “Deep speech 2: End-to-end speech recognition in english and mandarin,” in International conference on machine learning, 2016, pp. 173–182.
- L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, “Sca-cnn: Spatial and channel-wise attention in convolutional
- networks for image captioning,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 5659–5667.
- Q.-T. Ho, Y.-Y. Ou et al., “Classifying the molecular functions of rab gtpases in membrane trafficking using deep convolutional neural networks,” Analytical biochemistry, vol. 555, pp. 33–41, 2018.
- N.-Q.-K. Le, Q.-T. Ho, and Y.-Y. Ou, “Incorporating deep learning with convolutional neural networks and position specific scoring matrices for identifying electron transport proteins,” Journal of Computational Chemistry, vol. 38, no. 23, pp. 2000–2006, 2017.
- N. Q. K. Le, E. K. Y. Yapp, N. Nagasundaram, M. C. H. Chua, and H.-Y. Yeh, “Computational identification of vesicular transport proteins from sequences using deep gated recurrent units architecture,” Computational and Structural Biotechnology Journal, vol. 17, pp. 1245–1254, 2019.
- R. Cao, C. Freitas, L. Chan, M. Sun, H. Jiang, and Z. Chen, “Prolango: protein function prediction using neural machine translation based on a recurrent neural network,” Molecules, vol. 22, no. 10, p. 1732, 2017.
- J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, “Multimodal deep learning,” in Proceedings of the 28th international conference on machine learning (ICML-11), 2011, pp. 689–696.
- S. E. Kahou, X. Bouthillier, P. Lamblin, C. Gulcehre, V. Michalski,K. Konda, S. Jean, P. Froumenty, Y. Dauphin, N. Boulanger-