Learning from multi-view data: clustering algorithm and text mining application

Xinhai Liu

Jury:
Prof. C. Vandecasteele (Chairman)
Prof. B. De Moor (Promotor)
Prof. J. Vandewalle
Prof. Y. Moreau
Prof. J. Suykens
Prof. M. Moens
Prof. W. Daelemans

Department of Electrical Engineering, ESAT-SCD, K.U.Leuven

15th, September, 2011

Outline

1. Introduction

2. Multi-view clustering

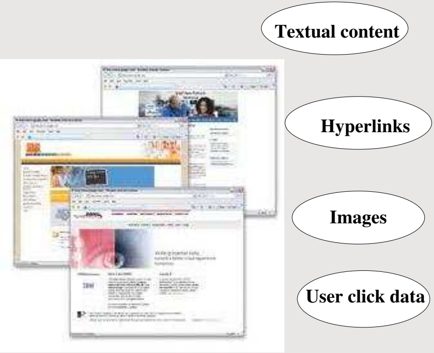
3. Multi-view text mining

4. Conclusion and outlook

Introduction

Multi-view data

The same class of entities can be observed or modeled from various perspectives, thus leading to multi-view representations.



Textual content

Hyperlinks

Images

User click data

Figure: WebPages with multi-view data

Introduction

Multi-view learning

Effectively exploring and exploiting the information from multi-view data for the purpose of improving the learning performance.

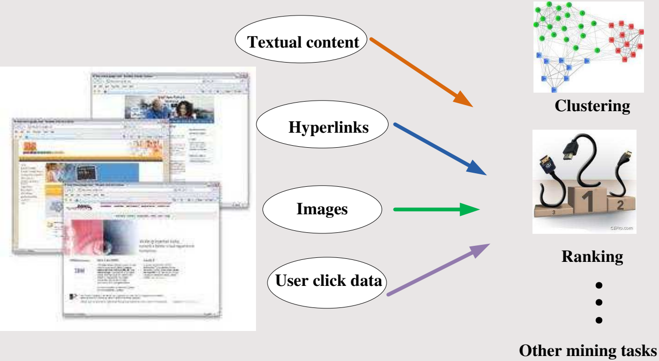


Figure: Web mining with multi-view learning

Introduction

Benefits of multi-view learning

Benefit 1: Recovering a complete pattern (Example: Scene reconstruction)

Five single-view data

View 2

View 4

View 1

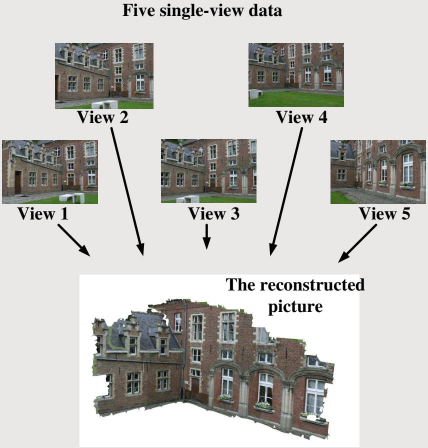
View 3

View 5

Introduction

Benefits of multi-view learning

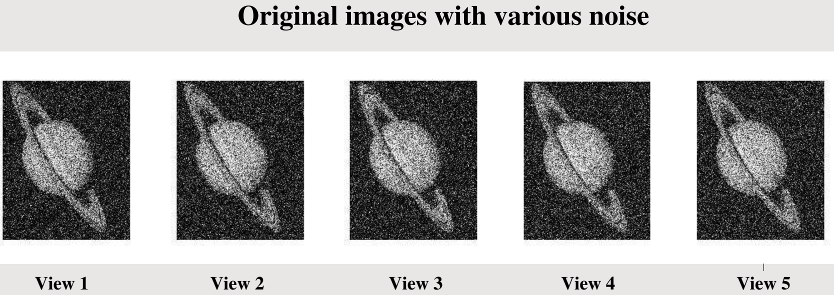
Benefit 1: Recovering a complete pattern (Example: Scene reconstruction)



Introduction

Benefits of multi-view learning

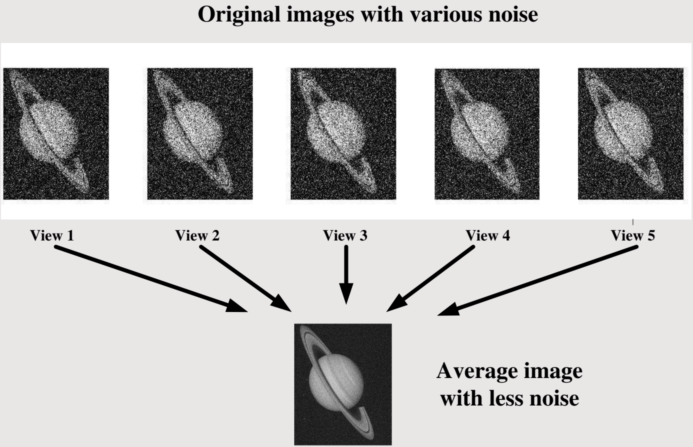
Benefit 2: Recovering a robust pattern (Example: Image denoising)



Introduction

Benefits of multi-view learning

Benefit 2: Recovering a robust pattern (Example: Image denoising)

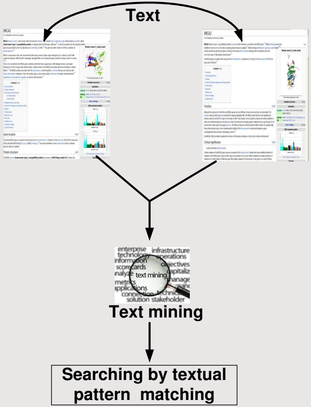


Introduction

Benefits of multi-view learning

Benefit 3: Facilitating learning tasks (Webpage retrieval: Search + Ranking)

Search: textual pattern matching



Text

Text mining

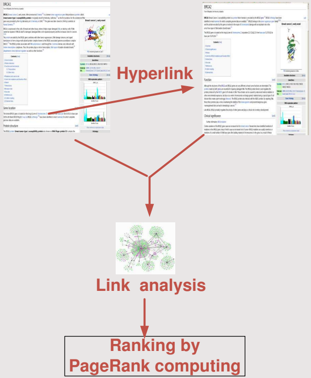
Searching by textual pattern matching

Introduction

Benefits of multi-view learning

Benefit 3: Facilitating learning tasks (Webpage retrieval: Search + Ranking)

Ranking: hyperlinks based PageRanking



Hyperlink

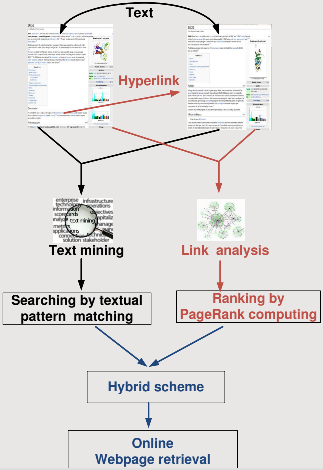
Link analysis

Ranking by PageRank computing

Introduction

Benefits of multi-view learning

Benefit 3: Facilitating learning tasks (Webpage retrieval: Search + Ranking)



Challenges of multi-view data analysis

- Jointly modeling heterogeneous data sources: A unified model for integration and analysis
- Intensive computation: Efficient algorithms for large-scale data
- The utilization of multi-view analysis in the increasing number applications (a diverse set of information sources (views))

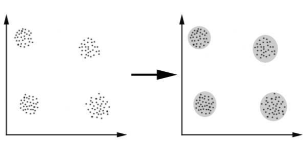
Introduction

Contributions

- Several multi-view clustering algorithms
  - From a multilinear perspective
  - Based on mutual information
  - Based on heterogeneous graph coupling
- Two real multi-view text mining applications
  - Scientific mapping of Web of Science (WoS) journal database
  - Text prior for clinical diagnosis

Clustering analysis

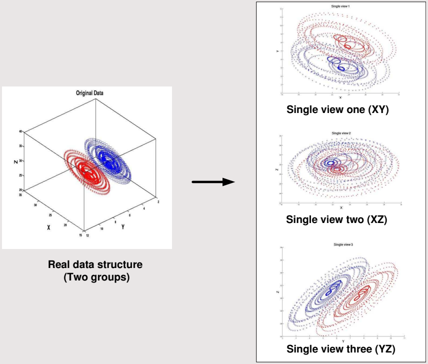
- Clustering analysis: assigning a set of objects into groups so that the objects in the same cluster are more similar to each other than to those in other clusters.



- Nonunique result and unsupervised learning: An exploratory tool and suggesting hypotheses for further analysis
- Application: Computer vision (image segmentation, image retrieval); Marketing research (grouping of shopping items, recommending systems); Biomedicine (sequence analysis); Social network analysis; Educational research . . .

Motivation of Multi-view clustering

Synthetic multi-view data



Real data structure
(Two groups)

Single view one (XY)

Single view two (XZ)

Single view three (YZ)

Real observations

Figure: Real observations from two group of data points

Motivation of Multi-view clustering

Single-view clustering

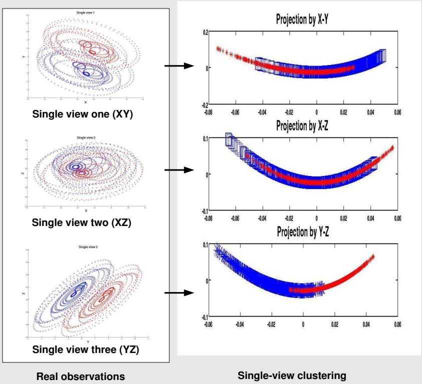
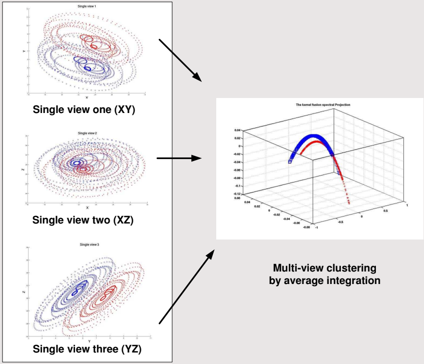


Figure: Spectral clustering on each single-view data

Motivation of Multi-view clustering

Simple multi-view clustering



Single view one (XY)

Single view two (XZ)

Single view three (YZ)

Multi-view clustering
by average integration

Real observations

Figure: Multi-view clustering by average integration

Motivation of Multi-view clustering

Tensor based multi-view clustering

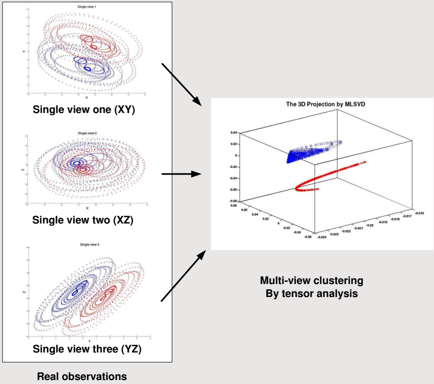


Figure: Multi-view clustering by tensor analysis

Multi-view clustering

Related work

- Multi-view clustering (Bickel & Scheffer, 2004): two views with independent assumption
- Hybrid clustering (Janssens *et al*, 2007; 2008; 2009): vector space
- Clustering ensemble (Strehl & Ghosh, 2002): integration on the partitioning level
- Multiple kernel learning (Yu *et al*, 2009; 2011): convex optimization

Multi-view clustering

From linear algebra to multilinear algebra

Our clustering work from a multilinear perspective

Single-view analysis

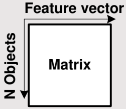
$\Downarrow$

Linear algebra

$\Downarrow$

Vector space model

$\Downarrow$



Multi-view clustering

From linear algebra to multilinear algebra

Our clustering work from a multilinear perspective

Single-view analysis

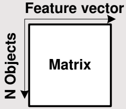
⇓

Linear algebra

⇓

Vector space model

⇓



Multi-view analysis

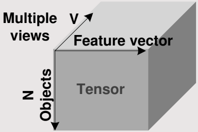
⇓

Multilinear algebra

⇓

Tensor space model

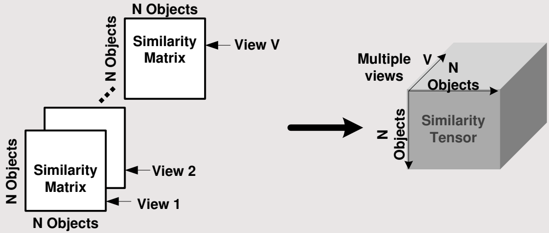
⇓



Multi-view clustering

Modeling multi-view data by a tensor

- Tensor model: integrating multiple views while keeping each view independent
- Integrating similarity matrices by combining heterogeneous data (feature spaces with various dimensionalities)



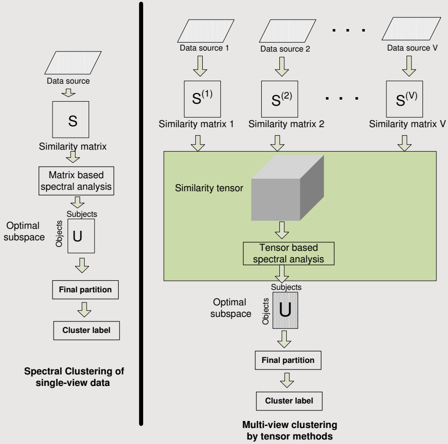
Multi-view clustering by tensor methods

Multi-view clustering by tensor methods

- Scheme 1: Obtaining a joint optimal subspace of multi-view data
- Scheme 2: Leveraging the multilinear relationship of multi-view data
- Scheme 3: Joint dimension reduction of multi-view data

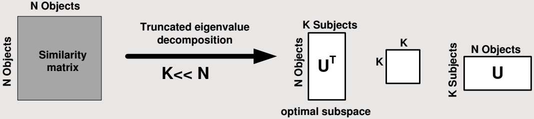
Scheme 1: obtaining a joint optimal subspace

Conceptual overview

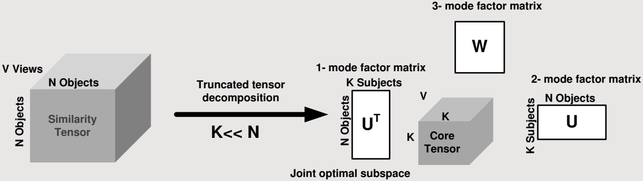


Scheme 1: obtaining a joint optimal subspace

Illustration and comparison



Spectral clustering by truncated matrix decomposition



Multi-view clustering by tensor decomposition

Scheme 1: Obtaining a joint optimal subspace

Objective function and solution

$$\max_{\mathbf{U},\mathbf{W}} \|\mathcal{A} \times_1 \mathbf{U}^T \times_2 \mathbf{U}^T \times_3 \mathbf{W}^T\|_F^2, \quad \text{s.t. } \mathbf{U}^T\mathbf{U} = \mathbf{I} \text{ and } \mathbf{W} = \mathbf{I}. \tag{1}$$

where $\mathcal{A}$ is the original similarity tensor and the columns of $\mathbf{U}$ form the joint optimal subspace.

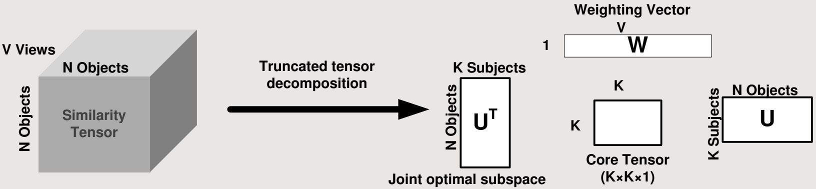
Algorithms:

- An approximate solution: Multi-view clustering by optimization integration by multilinear singular value decomposition (MC-OI-MLSVD)
- An optimal solution: Multi-view clustering by optimization integration by higher order orthogonal iteration (MC-OI-HOOI)

Scheme 2: Leveraging the multilinear relationship of multi-view data

Illustration

Principal component analysis (PCA) of the view space



W: the weighting factors of multi-view data, that is, the linear coefficients of each view to form the top principal component of the optimal view space

Scheme 2: Leveraging the multilinear relationship of multi-view data

Objective function and solution

$$\max_{\mathbf{U},W} \|\mathcal{A} \times_1 \mathbf{U}^T \times_2 \mathbf{U}^T \times_3 W^T\|_F^2, \; W = \begin{pmatrix} w_1 \\ \vdots \\ w_V \end{pmatrix} \quad \text{s.t. } \mathbf{U}^T\mathbf{U} = \mathbf{I}, \; \|W\|_F^2 = 1. \tag{2}$$

Algorithms

- Multi-view clustering of matrix integration by HOOI
- Multi-view clustering by simulatenous trace maximization (Extension algorithm based on alternating least square (ALS))

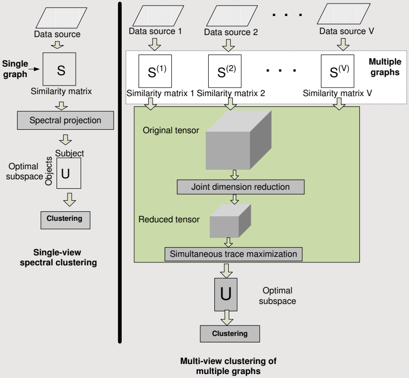
Scheme 3: joint dimension reduction of multi-view data

Motivation

- Multi-view data: high dimensional but a large amount of redundancy
- Dimension reduction by tensor methods on signal processing and computer vision (De Lathauwer, *et al* 2003; Lu, *et al*, 2009)
- The structure and correlation in the original data are preserved

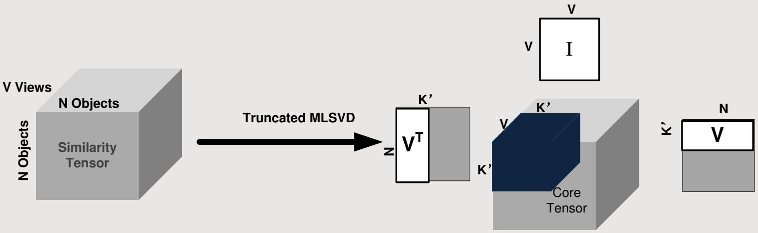
Scheme 3: joint dimension reduction of multi-view data

Conceptual overview



Scheme 3: joint dimension reduction of multi-view data

Illustration



Algorithms: Multi-view clustering by simultaneous trace maximization and MLSVD

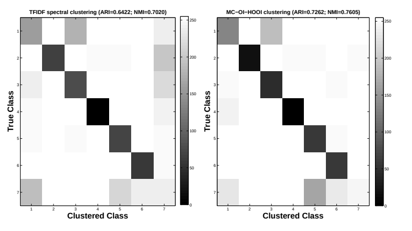
Multi-view clustering by tensor methods

Experiment: clustering on journal sets

Clustering 1424 journals into 7 categories

Multi-view data: text and citation

The reference journal categories is Essential Scientific Indicator (ESI)



Confusion matrices of two clustering strategies (best single-view clustering and MC-OI-HOOI on multi-view data). In each row, the diagonal element represents the fraction of correctly clustered journals and the off-diagonal non-zero element represents the fraction of mis-clustered journals.

Multi-view clustering by tensor methods

Experiment: clustering on synthetic data

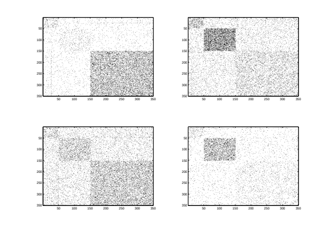


Figure: Visualization of the adjacency matrices of a synthetic multi-view data (Three clusters among 350 data points).

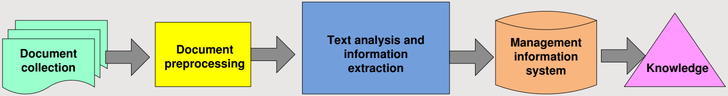
Table: Weighting analysis of MC-MI-HOOI

A1: 0.4725 (3)	A2: 0.5288 (2)
A3: 0.5643 (1)	A4: 0.4433 (4)

Multi-view text mining

Text mining

- Literature is the best knowledge
- Text mining: the process of deriving high-quality information (pattern, relationship, trend and so on) from text


Document collection
Document preprocessing
Text analysis and information extraction
Management information system
Knowledge

- Applications: Biomedicine, Marketing (customer relationship management), Online media, Security, Sentiment analysis, Academic applications (publication) . . .

Application 1: Scientific mapping of Web of Science journal database

Introduction

Objectives:

- Partitioning journals into different categories
- Analyzing the relationship of various categories and finding new trends

Database of WoS

- All abstracts and titles of more than 8,000 SCI indexed journals from 2002 to 2006
- Aggregating the text and citation from paper level to journal level

Application 1: Scientific mapping

Multi-view data

- Text data: TFIDF, TF, IDF, Binary-Text
- Link data: cross-citation, bibliographic coupling, co-citation, binary cross-citation
- Latent semantic indexing (LSI)

Application 1: Scientific mapping

Multi-view data

- Text data: TFIDF, TF, IDF, Binary-Text
- Link data: cross-citation, bibliographic coupling, co-citation, binary cross-citation
- Latent semantic indexing (LSI)

Hybrid clustering strategies

- Vector space model: based on mutual information of multi-view data
- Graph space model (for large scale data): graph coupling of citation based link structure and text based link strength

Application 1: Scientific mapping

Network of journal clusters

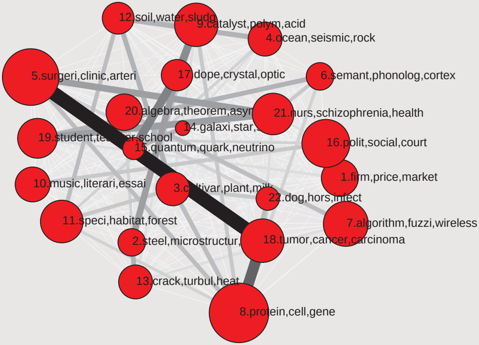
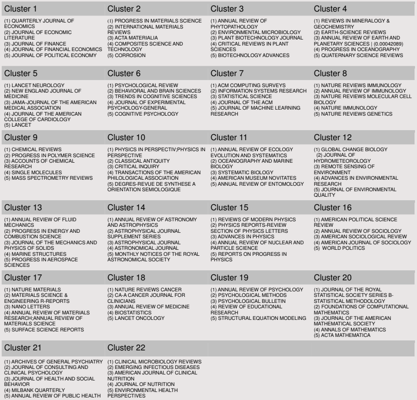


Figure: Visualization of 22 clusters on the WoS journal database by graph based hybrid clustering. (**the node**: the journal clusters where the circle size is proportional to its scale; **the edge**: cross-citation between two journal clusters; **the annotated terms**: the top three text terms within each journal clusters)

Application 1: Scientific mapping

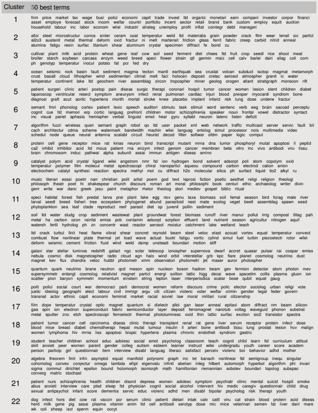
The five most important journals of each cluster ranked by modified PageRank algorithm



Cluster 1	Cluster 2	Cluster 3	Cluster 4
(1) QUARTERLY JOURNAL OF ECONOMICS (2) JOURNAL OF ECONOMIC LITERATURE (3) JOURNAL OF FINANCE (4) JOURNAL OF FINANCIAL ECONOMICS (5) JOURNAL OF POLITICAL ECONOMY	(1) PROGRESS IN MATERIALS SCIENCE (2) INTERNATIONAL MATERIALS REVIEWS (3) ACTA MATERIALIA (4) COMPOSITES SCIENCE AND TECHNOLOGY (5) CORROSION	(1) ANNUAL REVIEW OF PHYTOPATHOLOGY (2) ENVIRONMENTAL MICROBIOLOGY (3) PLANT BIOTECHNOLOGY JOURNAL (4) CRITICAL REVIEWS IN PLANT SCIENCES (5) BIOTECHNOLOGY ADVANCES	(1) REVIEWS IN MINERALOGY & GEOCHEMISTRY (2) EARTH-SCIENCE REVIEWS (3) ANNUAL REVIEW OF EARTH AND PLANETARY SCIENCES | (0.00042089) (4) PROGRESS IN OCEANOGRAPHY (5) QUATERNARY SCIENCE REVIEWS
Cluster 5	Cluster 6	Cluster 7	Cluster 8
(1) LANCET NEUROLOGY (2) NEW ENGLAND JOURNAL OF MEDICINE (3) JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION (4) JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY (5) LANCET	(1) PSYCHOLOGICAL REVIEW (2) BEHAVIORAL AND BRAIN SCIENCES (3) TRENDS IN COGNITIVE SCIENCES (4) JOURNAL OF EXPERIMENTAL PSYCHOLOGY-GENERAL (5) COGNITIVE PSYCHOLOGY	(1) ACM COMPUTING SURVEYS (2) INFORMATION SYSTEMS RESEARCH (3) STATISTICAL SCIENCE (4) JOURNAL OF THE ACM (5) JOURNAL OF MACHINE LEARNING RESEARCH	(1) NATURE REVIEWS IMMUNOLOGY (2) ANNUAL REVIEW OF IMMUNOLOGY (3) NATURE REVIEWS MOLECULAR CELL BIOLOGY (4) NATURE IMMUNOLOGY (5) NATURE REVIEWS GENETICS
Cluster 9	Cluster 10	Cluster 11	Cluster 12
(1) CHEMICAL REVIEWS (2) PROGRESS IN POLYMER SCIENCE (3) ACCOUNTS OF CHEMICAL RESEARCH (4) SINGLE MOLECULES (5) MASS SPECTROMETRY REVIEWS	(1) PHYSICS IN PERSPECTIV;PHYSICS IN PERSPECTIVE (2) CLASSICAL ANTIQUITY (3) CRITICAL INQUIRY (4) TRANSACTIONS OF THE AMERICAN PHILOLOGICAL ASSOCIATION (5) DEGRES-REVUE DE SYNTHESE A ORIENTATION SEMIOLOGIQUE	(1) ANNUAL REVIEW OF ECOLOGY EVOLUTION AND SYSTEMATICS (2) OCEANOGRAPHY AND MARINE BIOLOGY (3) SYSTEMATIC BIOLOGY (4) AMERICAN MUSEUM NOVITATES (5) ANNUAL REVIEW OF ENTOMOLOGY	(1) GLOBAL CHANGE BIOLOGY (2) JOURNAL OF HYDROMETEOROLOGY (3) REMOTE SENSING OF ENVIRONMENT (4) ADVANCES IN ENVIRONMENTAL RESEARCH (5) JOURNAL OF ENVIRONMENTAL QUALITY
Cluster 13	Cluster 14	Cluster 15	Cluster 16
(1) ANNUAL REVIEW OF FLUID MECHANICS (2) PROGRESS IN ENERGY AND COMBUSTION SCIENCE (3) JOURNAL OF THE MECHANICS AND PHYSICS OF SOLIDS (4) MARINE STRUCTURES (5) PROGRESS IN AEROSPACE SCIENCES	(1) ANNUAL REVIEW OF ASTRONOMY AND ASTROPHYSICS (2) ASTROPHYSICAL JOURNAL SUPPLEMENT SERIES (3) ASTROPHYSICAL JOURNAL (4) ASTRONOMICAL JOURNAL (5) MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY	(1) REVIEWS OF MODERN PHYSICS (2) PHYSICS REPORTS-REVIEW SECTION OF PHYSICS LETTERS (3) ADVANCES IN PHYSICS (4) ANNUAL REVIEW OF NUCLEAR AND PARTICLE SCIENCE (5) REPORTS ON PROGRESS IN PHYSICS	(1) AMERICAN POLITICAL SCIENCE REVIEW (2) ANNUAL REVIEW OF SOCIOLOGY (3) AMERICAN SOCIOLOGICAL REVIEW (4) AMERICAN JOURNAL OF SOCIOLOGY (5) WORLD POLITICS
Cluster 17	Cluster 18	Cluster 19	Cluster 20
(1) NATURE MATERIALS (2) MATERIALS SCIENCE & ENGINEERING R-REPORTS (3) NANO LETTERS (4) ANNUAL REVIEW OF MATERIALS RESEARCH;ANNUAL REVIEW OF MATERIALS SCIENCE (5) SURFACE SCIENCE REPORTS	(1) NATURE REVIEWS CANCER (2) CA-A CANCER JOURNAL FOR CLINICIANS (3) ANNUAL REVIEW OF MEDICINE (4) BIOSTATISTICS (5) LANCET ONCOLOGY	(1) ANNUAL REVIEW OF PSYCHOLOGY (2) PSYCHOLOGICAL METHODS (3) PSYCHOLOGICAL BULLETIN (4) REVIEW OF EDUCATIONAL RESEARCH (5) STRUCTURAL EQUATION MODELING	(1) JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES B-STATISTICAL METHODOLOGY (2) FOUNDATIONS OF COMPUTATIONAL MATHEMATICS (3) JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY (4) ANNALS OF MATHEMATICS (5) ACTA MATHEMATICA
Cluster 21	Cluster 22		
(1) ARCHIVES OF GENERAL PSYCHIATRY (2) JOURNAL OF CONSULTING AND CLINICAL PSYCHOLOGY (3) JOURNAL OF HEALTH AND SOCIAL BEHAVIOR (4) MILBANK QUARTERLY (5) ANNUAL REVIEW OF PUBLIC HEALTH	(1) CLINICAL MICROBIOLOGY REVIEWS (2) EMERGING INFECTIOUS DISEASES (3) AMERICAN JOURNAL OF CLINICAL NUTRITION (4) JOURNAL OF NUTRITION (5) ENVIRONMENTAL HEALTH PERSPECTIVES		

Application 1: Scientific mapping

The textual labels of the journal clusters



Cluster	50 best terms
1	firm price market tax wage busi polici economi capit trade invest ltd organiz monetari earn compani investor corpor financi asset employe forecast stock incom welfar countri portfolio incent sector retail brand bank custom employ equiti auction household labour inc labor econom wilei industri strateg unemploy profit inflat cointegr debt manageri
2	alloi steel microstructur corros sinter ceram coat temperatur weld ltd materialia grain powder crack film wear tensil sic particl al2o3 austenit metal thermal deform oxid fractur ni melt martensit friction glass ferrit fabric creep carbid nitrid anneal alumina fatigu resin surfac titanium shear aluminum crystal specimen diffract fe bond cu
3	cultivar plant milk acid protein wheat gene leaf cow soil seed ferment diet chees ltd fruit crop seedl rice shoot meat broiler starch soybean carcass enzym weed breed speci flower strain qtl germin maiz cell calv barlei dairi silag coli corn ph genotyp temperatur inocul potato fat pcr fed dry
4	ocean seismic rock basin fault sediment magma tecton mantl earthquak sea crustal volcan subduct isotop magmat metamorph crust basalt cloud lithospher wind sedimentari climat melt faci holocen deposit cretac aerosol atmospher granit ic water temperatur continent lake ltd geochem river assemblag zircon tropospher miner geolog orogen atlant stratigraph monsoon rift
5	patient surgeri clinic arteri postop pain diseas surgic therapi coronari hospit tumor cancer women lesion stent children diabet laparoscop ventricular resect symptom aneurysm infect renal pulmonari cardiac injuri blood preoper myocardi syndrom bone diagnosi graft acut aortic hypertens month mortal stroke knee placebo implant infarct risk lung dose underw fractur
6	semant fmri phonolog cortex patient lexic speech auditori stimulu task stimuli word sentenc verb eeg brain saccad perceptu cognit cue ltd memori erp cortic languag prefront children motor schizophrenia speaker noun frontal vowel distractor syntact inc visual pariet aphasia hemispher verbal linguist emot hear gyru syllabl neuron latenc listen deficit
7	algorithm fuzzi wireless queri semant graph robot qo ltd user packet xml web network traffic multicast server servic fault bit cach architectur cdma scheme watermark bandwidth machin wilei languag ontolog simul processor nois multimedia video schedul node queue neural antenna scalabl circuit heurist decod filter softwar ofdm paper logic comput
8	protein cell gene receptor mice rat kinas neuron bind transcript mutant mrna dna tumor phosphoryl mutat apoptosi il peptid ca2 inhibit inhibitor acid ltd mous patient rna enzym infect genom cancer membran beta vitro inc vivo antibodi viru tissu brain chromosom induc pathwai alpha subunit assai immun antigen diseas amino
9	catalyst polym acid crystal ligand wilei angstrom nmr ltd ion hydrogen bond solvent adsorpt poli atom copolym oxid temperatur polymer film molecul metal spectroscopi chiral nanoparticl aqueou compound carbon electrod cation anion electrochem catalyt synthesi reaction spectra methyl mol cu diffract h2o molecular silica ph surfact liquid tio2 alkyl ru
10	music literari essai poetri narr christian polit artist poem god text leprosi fiction poetic aesthet religi religion theologi philosoph theatr poet hi shakespear church discours roman art moral philosophi book centuri ethic archaeolog writer divin genr write war danc greek jesu paint metaphor rhetor theolog stori mediev gospel biblic ritual
11	speci habitat forest fish predat larva prei plant lake egg nov genu taxa biomass soil femal season bird forag male river larval seedl breed fisheri tree ecosystem phylogenet abund parasitoid nest mate ecolog veget beetl assemblag spawn seed phytoplankton sea leaf clade reproduct reef parasit diet sp juvenil pollin sediment
12	soil ltd water sludg crop sediment wastewat plant groundwat forest biomass runoff river manur pollut irrig compost tillag pah metal ha carbon ozon rainfal emiss pcb contamin adsorpt sorption effluent land nutrient season agricultur nitrogen aquif watersh fertil hydrolog ph zn concentr wast reactor aerosol moistur catchment lake wetland leach
13	ltd crack turbul finit heat flame vibrat shear concret reynold beam steel veloc elast acoust vortex equat temperatur convect combust flow nonlinear plate thermal jet load wave actuat buckl fatigu vortic cylind simul fuel turbin piezoelectr rotor wilei deform seismic cement friction fluid wind weld damp unsteadi boundari motion stiff
14	galaxi star stellar luminos redshift galact ngc solar telescop ionospher supernova dwarf accret quasar pulsar rai cospar emiss nebula cosmic disk magnetospher radio cloud agn halo wind orbit interstellar grb kpc flare planet cosmolog neutrino dust magnet kev flux chandra veloc hubbl photometr xmm observatori photometri jet maser auror photospher
15	quantum quark neutrino brane neutron qcd meson spin nucleon boson hadron beam gev fermion detector atom photon mev supersymmetr entangl cosmolog relativist magnet particl energi soliton lattic higg decai wave spacetim collis plasma gluon ion scatter pion baryon symmetri momentum einstein string lepton laser interperiodica scalar bose qubit equat gaug
16	polit polici social court war democraci parti democrat women reform discours crime polic elector sociolog urban religi vote justic ideolog geographi elect labour civil immigr argu citi citizen violenc voter welfar crimin gender legal feder govern transnat actor ethnic capit economi feminist market racial soviet law moral militari rural citizenship
17	film dope temperatur crystal optic magnet quantum si dielectr alloi gan laser anneal epitaxi atom diffract nm beam silicon gaa spin ion electron superconduct fabric semiconductor layer deposit ferromagnet nanotub voltag waveguid phonon substrat metal sputter zno etch spectroscopi ferroelectr thermal photoluminesc oxid thin lattic surfac exciton sio2 transistor spectra
18	patient tumor cancer cell carcinoma diseas clinic therapi transplant renal rat gene liver serum receptor protein infect dose blood mice breast diabet chemotherapi hepat mutat tumour insulin il arteri bone antibodi tissu lung prostat lesion hcv malign women lymphoma hiv mrna liss apoptosi biopsi hypertens plasma chronic endotheli syndrom gastric
19	student teacher children school educ adolesc social emot psycholog classroom teach cognit child learn ltd curriculum attitud skill anxieti peer women parent gender colleg autism esteem learner instruct wilei undergradu youth career score academ person particip girl questionnair item interview disabl languag literaci satisfact perceiv violenc boi behavior adhd mother
20	algebra theorem finit infin asymptot equat manifold polynomi graph inc let banach nonlinear ltd semigroup inequ singular cohomolog convex conjectur omega lambda ellipt eigenvalu infinit abelian integ hilbert automorph hyperbol algorithm phi invari sigma commut dirichlet epsilon bound holomorph isomorph math hamiltonian riemannian sobolev boundari topolog subspac converg matric stochast
21	patient nurs schizophrenia health children disord depress women adolesc symptom psychiatr clinic mental suicid hospit smoke abus anxieti interview care ptsd sleep ltd physician cognit social alcohol intervent hiv medic caregiv questionnair child drug sexual antipsychot infant score dementia servic educ violenc adhd men disabl bipolar psycholog risk therapi youth
22	dog infect hors diet cow rat vaccin pcr serum clinic patient dietari intak calv cattl viru cat strain blood protein acid diseas herd milk gene pig assai plasma vitamin anim ltd cell antibodi serotyp dose mic mice veterinari semen fat liver dairi mare wk coli sheep isol sperm equin oocyt

Application 2: Text Prior project

Objective

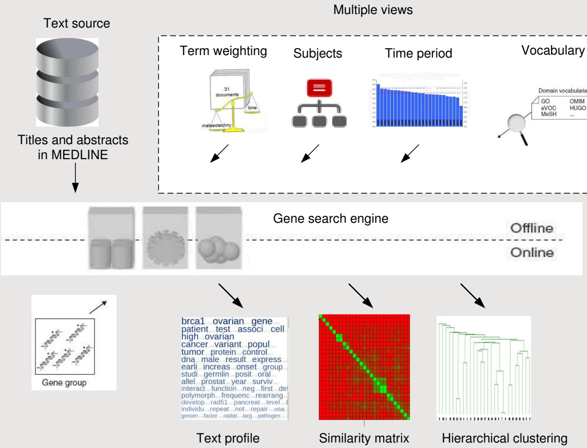
- Finding the relationship among genes to aid the cancer diagnosis & Providing prior information for typical clinical decisions support algorithms

Strategies

- Data fusion by integrating multi-view text mining data
- Vertical observation from a specific view

Application 2: Text Prior Project

Conceptual overview



Multiple views

Text source

Titles and abstracts in MEDLINE

Term weighting

Subjects

Time period

Vocabulary

Gene search engine

Offline

Online

Gene group

Text profile

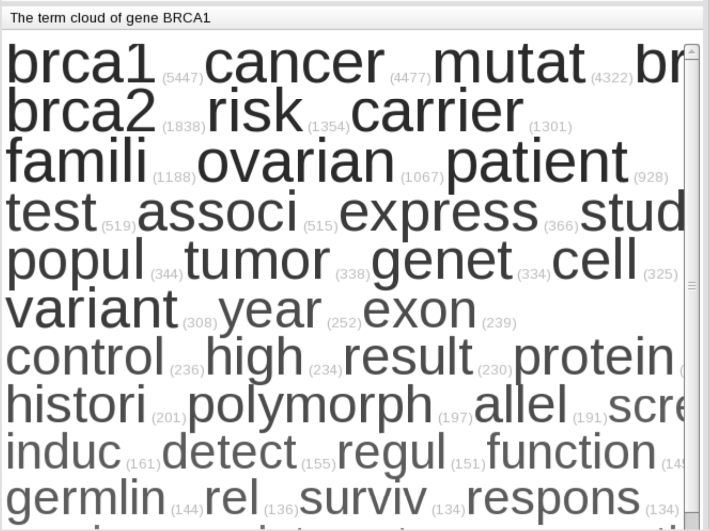
Similarity matrix

Hierarchical clustering

Project software available: http://aulne8.esat.kuleuven.be/TextPrior/

Application 2: Text Prior Project

Term cloud



The term cloud of gene BRCA1

brca1 (5447) cancer (4477) mutat (4322) br
brca2 (1838) risk (1354) carrier (1301)
famili (1188) ovarian (1067) patient (928)
test (519) associ (515) express (366) stud
popul (344) tumor (338) genet (334) cell (325)
variant (308) year (252) exon (239)
control (236) high (234) result (230) protein (
histori (201) polymorph (197) allel (191) scre
induc (161) detect (155) regul (151) function (14
germlin (144) rel (136) surviv (134) respons (134)

Conclusion and outlook

Conclusion

- Multi-view clustering based on multilinear algebra
  - Tensor model for multi-view data
  - Multi-view partitioning by tensor decomposition
  - Joint dimension reduction by multilinear projection
- Multi-view text mining
  - Scientific mapping and Text prior for biomedical application
  - Hybrid clustering of multi-view text mining data
  - Vertical observation for a specific domain

Conclusion and outlook

Outlook

- Extending to other multi-view learning tasks: classification, spectral embedding, collaborative filtering
- Multi-way learning by tensor analysis
- Missing data and multi-look clustering
- Outliers detection by multi-view clustering
- Text mining on medical report analysis

Thank you for your attention!