Dan-ping LIAO, Yun-tao QIAN, 2019. Paper evolution graph: multi-view structural retrieval for academic literature. *Frontiers of Information Technology & Electronic Engineering,* 20(2):187-205.
https://doi.org/10.1631/FITEE.1700105

Front Inform Technol Electron Eng

Paper evolution graph: multi-view structural retrieval for academic literature

Key words: Paper evolution graph; Academic literature retrieval; Metagraph factorization; Topic coherence

Corresponding author: Yun-tao QIAN
E-mail: ytqian@zju.edu.cn
ORCID: Yun-tao QIAN, http://orcid.org/0000-0002-7418-5891

Motivation

1. The existing academic literature retrieval systems, such as Google Scholar, Scopus, and Web of Science, can return articles that are most likely to match users' queries. However, the returned articles are displayed in a listed and isolated way. In other words, the underlying relationships between the retrieved articles remain unknown to users.

2. We present the retrieved results in a way by which the evolutionary relationship between the papers is explicitly shown.

Front Inform Technol Electron Eng

Main idea

1. We propose a concept of paper evolution graph and formalize the criteria to evaluate evolution graphs.

2. Three types of queries are supported and efficient methods are provided to construct evolution graphs given different types of queries.

3. We integrate user preferences into the framework to generate graphs describing the multi-view relationships among articles.

Front Inform Technol Electron Eng

Method

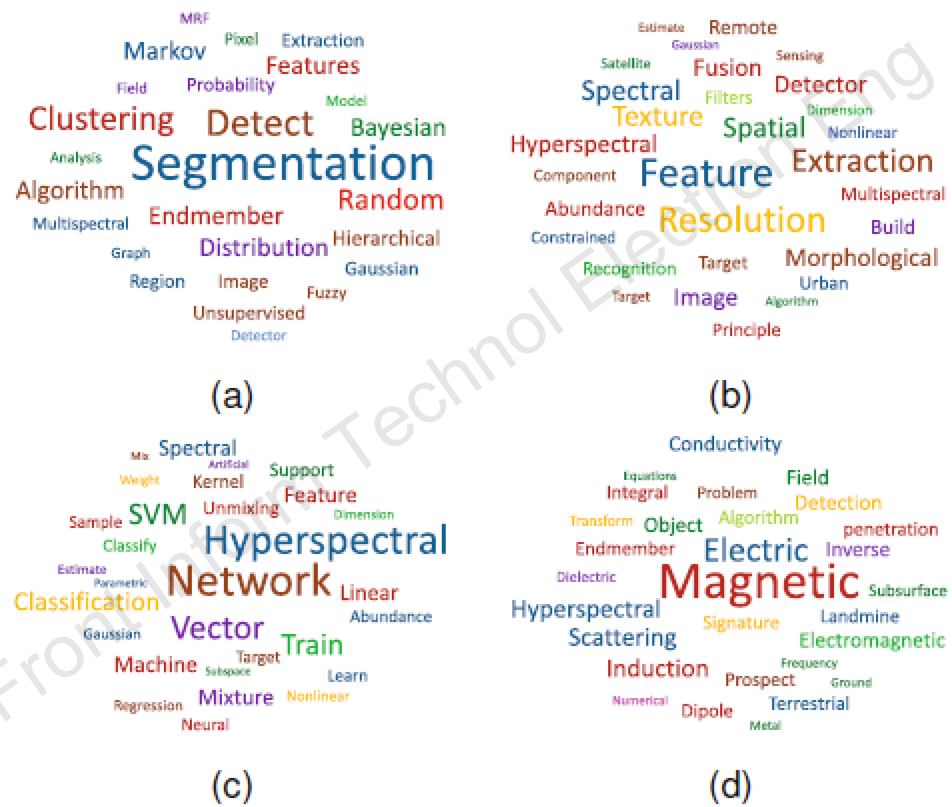
1. To fully cover the topics of a query paper, the topic distribution of papers in the dataset is obtained by multi-relational factorization of the metagraph, and then articles in the dataset are soft-clustered into communities based on their topic distribution.

2. From each community that relates to the query, the most cohesive chain is extracted based on topic coherence. The topic coherence between two papers is calculated by the word influence propagating algorithm.

3. The extracted chains are combined to form a paper evolution graph. Each chain focuses on one aspect of the query. Combining the chains gives us comprehensive and holistic retrieval results.

Front Inform Technol Electron Eng

Major results



Front Inform Technol Electron Eng

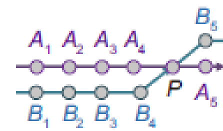
Fig. 8 Word clouds for various communities: (a) word cloud of community 9; (b) word cloud of community 11; (c) word cloud of community 14; (d) word cloud of community 30

Major results (Cont'd)

A retrieval example of the single-paper query

Input: Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles

Output:



Chain 1

A_1: Morphological transformations and feature extraction of urban data with high spectral and spatial resolution. (July 2003)

A_2: Decision level fusion in classification of hyperspectral data from urban areas. (Sept. 2004)

A_3: Classification of hyperspectral data from urban areas based on extended morphological profiles. (March 2005)

A_4: High-resolution multispectral image classification over urban areas by image segmentation and extended morphological profile. (July 2006)

P: Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. (Nov. 2008)

A_5: Classification of hyperspectral images with extended attribute profiles and feature extraction techniques. (July 2010)

Chain 2

B_1: Support vector machines for classification of hyperspectral remote-sensing images. (June 2002)

B_2: Source based feature extraction for support vector machines in hyperspectral classification. (Sept. 2004)

B_3: Transductive SVMs for semisupervised classification of hyperspectral data. (July 2005)

B_4: A combined support vector machines classification based on decision fusion. (July 2005)

P: Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. (Nov. 2008)

B_5: Spectral and spatial classification of hyperspectral data using SVMs and Gabor textures. (July 2011)

Fig. 9 Paper evolution graph of the query paper "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles"

Front Inform Technol Electron Eng

Major results (Cont'd)

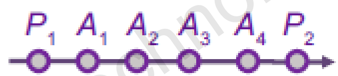
A retrieval example of the two-paper query

Input:

P_1: Kernel-based methods for hyperspectral image classification

P_2: Kernel nonparametric weighted feature extraction for hyperspectral image classification

Output:



P_1: Kernel-based methods for hyperspectral image classification. (June 2005)

A_1: Kernel orthogonal subspace projection for hyperspectral signal classification. (Dec. 2005)

A_2: Composite kernels for hyperspectral image classification. (Jan. 2006)

A_3: Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. (June 2008)

A_4: Kernel adaptive subspace detector for hyperspectral imagery. (March 2009)

P_2: Kernel nonparametric weighted feature extraction for hyperspectral image classification. (April 2009)

Fig. 11 Paper evolution graph of two-paper retrieval focusing on hyperspectral imagery classification

Front Inform Technol Electron Eng

Major results (Cont'd)

A retrieval example of the keyword query

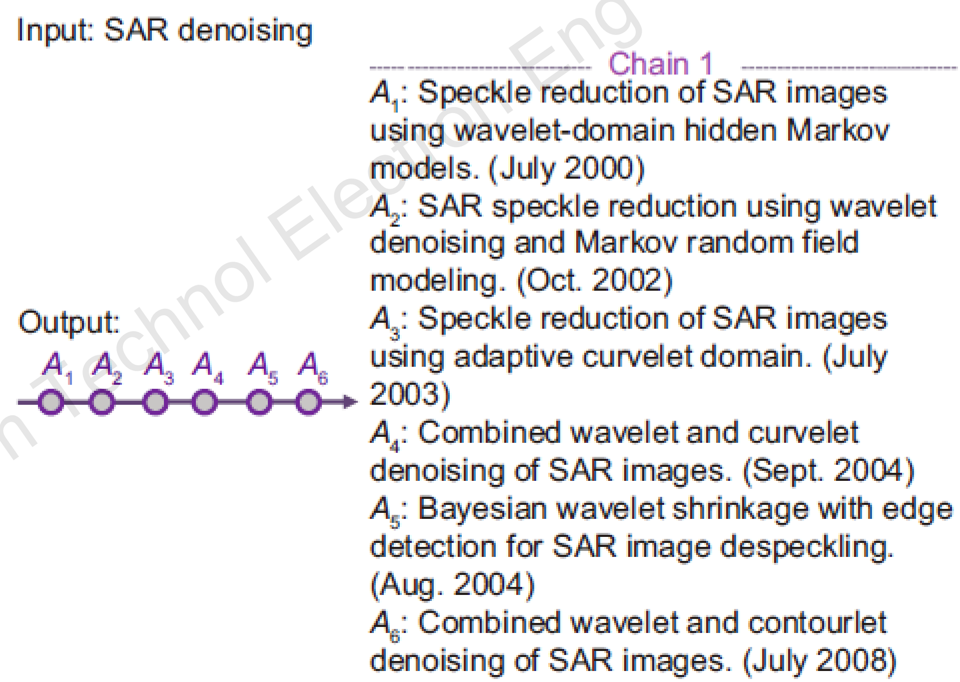


Fig. 14 Paper evolution graph for "SAR denoising" retrieval

Front Inform Technol Electron Eng

Major results (Cont’d)

Compared with Google Scholar, IEEE Xplore, and Web of Science in terms of topic coherence

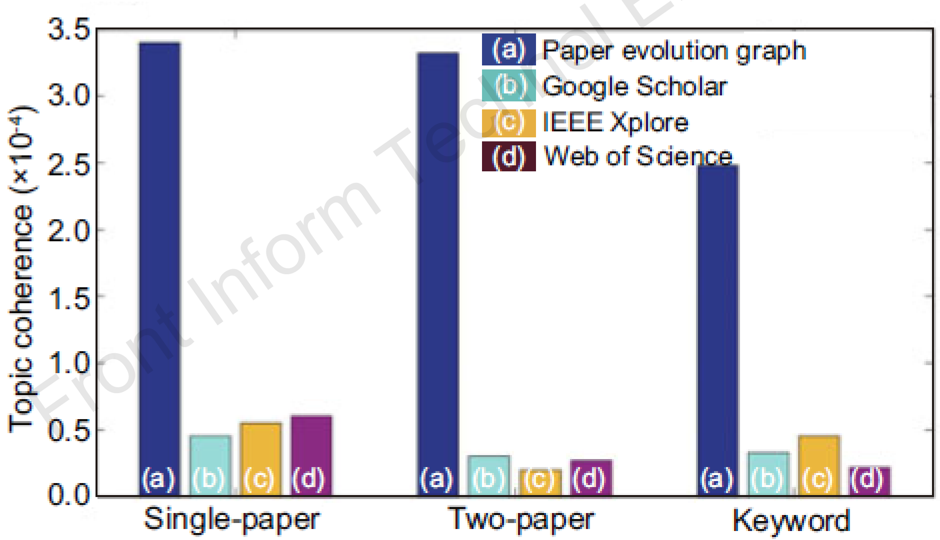


Fig. 17 Comparison of topic coherence

Front Inform Technol Electron Eng

Conclusions

1. The paper evolution graph (PEG) method has been proposed to create structured paper retrieval results. The PEG explicitly shows the multi-view relationships between the retrieved papers by a combination of a set of evolution chains.

2. Three types of information, i.e., content, author, and citation, are used in our system, to which users are allowed to attribute different weights to generate an evolution graph emphasizing different aspects.

3. Our system supports keyword search, single-paper search, and two-paper search to meet different user requirements.

Front Inform Technol Electron Eng