Skip to Main Content
Wiley Online Library

Wiley Online Library

Search
This Journal

- This Journal
- Anywhere

- Search term Search
  Search
- Search term Search
  Search

Login / Register

STATISTICAL ANALYSIS AND DATA MINING

Research Article

Exploiting associations between word clusters and document classes for cross-domain text categorization†

Fuzhen Zhuang

Corresponding Author

E-mail address:zhuangfz@ics.ict.ac.cn

The Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences

Graduate University of Chinese Academy of Sciences, Beijing 100039, China

The Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences
Search for more papers by this author
Ping Luo

Hewlett Packard Labs China

Search for more papers by this author
Hui Xiong

MSIS Department, Rutgers University

Search for more papers by this author
Qing He

The Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences

Search for more papers by this author
Yuhong Xiong

Innovation Works

Search for more papers by this author
Zhongzhi Shi

The Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences

Search for more papers by this author
Fuzhen Zhuang

Corresponding Author

E-mail address:zhuangfz@ics.ict.ac.cn

The Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences

Graduate University of Chinese Academy of Sciences, Beijing 100039, China

The Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences
Search for more papers by this author
Ping Luo

Hewlett Packard Labs China

Search for more papers by this author
Hui Xiong

MSIS Department, Rutgers University

Search for more papers by this author
Qing He

The Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences

Search for more papers by this author
Yuhong Xiong

Innovation Works

Search for more papers by this author
Zhongzhi Shi

The Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences

Search for more papers by this author
First published: 30 November 2010
https://doi.org/10.1002/sam.10099
Cited by: 17
†

†This is an invited submission from the Best of SDM 2010.

About

- Access
- Related
- Information

PDF

PDF
Access

- Log in to get access

INSTITUTIONAL
LOGIN
>

**Personal
login**

Wiley Online Library

Log in to your account

Email or Customer ID
Enter your email
Password
Enter your password
Forgot password?
Log In
NEW USER
>

- Purchase instant access
- Recommend to a librarian

PDF
PDF
PDF Tools

- Request permission
- Export citation
- Add to favorites
- Track citation

Share

Give access

Share

full
text
access

Share
full
text
access

Please review our Terms and Conditions of Use and check box below to share full-text version of article.

I have read and accept the Wiley Online Library Terms and Conditions of Use.

Shareable Link

Use the link below to share a full-text version of this article with your friends and colleagues. Learn more.

Copy URL

Share
a
link

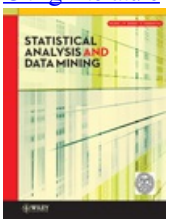
- Email to a friend
- Facebook
- Twitter

- Linkedin
- Google+
- Reddit
- CiteULike

Abstract

Cross-domain text categorization targets on adapting the knowledge learnt from a labeled source domain to an unlabeled target domain, where the documents from the source and target domains are drawn from different distributions. However, in spite of the different distributions in raw-word features, the associations between word clusters (conceptual features) and document classes may remain stable across different domains. In this paper, we exploit these unchanged associations as the bridge of knowledge transformation from the source domain to the target domain by the non-negative matrix tri-factorization. Specifically, we formulate a joint optimization framework of the two matrix tri-factorizations for the source- and target-domain data, respectively, in which the associations between word clusters and document classes are shared between them. Then, we give an iterative algorithm for this optimization and theoretically show its convergence. The comprehensive experiments show the effectiveness of this method. In particular, we show that the proposed method can deal with some difficult scenarios where baseline methods usually do not perform well. © 2010 Wiley Periodicals, Inc. Statistical Analysis and Data Mining 4: 100–114, 2011

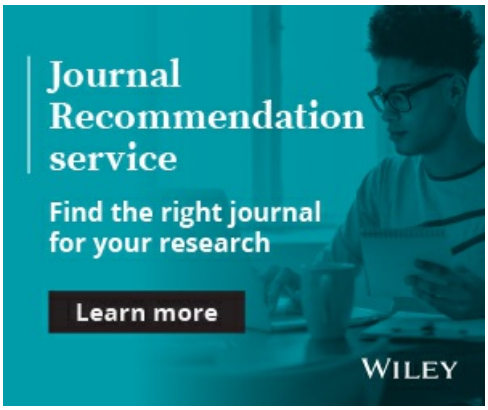
Citing Literature



Volume4, Issue1

February 2011

Pages 100-114



- Acces
- Related
- Information

Caption

Additional links

About Wiley Online Library

- Privacy Policy
- Terms of Use
- Cookies
- Accessibility

Help & Support

- Contact Us

Opportunities

- Subscription Agents
- Advertisers & Corporate Partners

Connect with Wiley

- The Wiley Network
- Wiley Press Room

Copyright © 1999-2018 John Wiley & Sons, Inc. All rights reserved

WILEY

Wiley Online Library

Log in to Wiley Online Library

Email or Customer ID

Enter your email

Password

Enter your password

Forgot password?

Log In

NEW USER > INSTITUTIONAL LOGIN >

Wiley Online Library

Change Password

Old Password

New Password

Too Short Weak Medium Strong Very Strong Too Long

Submit

Congrats!

Your password has been changed

Create a new account

Email or Customer ID

Register

Returning user

Wiley Online Library

Forgot your password?

Enter your email address below. If your address has been previously registered, you will receive an email with instructions on how to reset your password. If you don't receive an email, you should register as a new user

Email or Customer ID

Enter your email

RESET PASSWORD

Please check your email for your password reset instructions.

Request Username

Can't sign in? Forgot your username?

Enter your email address below and we will send you your username

Email or Customer ID

Submit

Close

If the address matches an existing account you will receive an email with instructions to retrieve your username