Archive of SID

فصلنامه علمی پیشرفت‌های مالی و سرمایه‌گذاری
سال دوم، بهار ۱۴۰۰- شماره ۲
صفحات ۵۷- ۲۳

# مروری بر پژوهش‌های انجام‌شده در خصوص خوشه‌بندی سری‌های زمانی مالی رویکرد نگاشت دانش

مرضیه نوراحمدی[۱]
فاطمه راستی[۲]
حجت‌الله صادقی[۳]

تاریخ دریافت: ۱۳۹۹/۱۰/۱۷
تاریخ پذیرش: ۱۴۰۰/۰۵/۱۰

## چکیده

میزان اطلاعات و داده‌هایی که ما بازیابی و استفاده می‌کنیم، به‌سرعت در حال افزایش هستند. داده‌کاوی فرایند استخراج داده‌های مربوط از حجم زیادی از داده‌ها و روش کشف و پیدا کردن الگوی مناسب از حجم زیادی از مجموعه داده‌ها است. خوشه‌بندی یکی از روش‌های معمول تجزیه‌وتحلیل داده‌های آماری و همچنین یکی از بهترین رویکردهای داده‌کاوی است. این رویکرد به‌عنوان یکی از روش‌های یادگیری بدون نظارت، با به‌کارگیری الگوریتم‌هایی، داده‌های سری‌های زمانی را برحسب معیارهای متفاوتی طبقه‌بندی می‌کند. هدف از پژوهش حاضر بررسی انواع کاربردهای خوشه‌بندی و شبکه‌سازی در حوزه‌های مختلف مالی ازجمله ریسک، معاملات الگوریتمی، بانکداری و دیگر موضوعات پرکاربرد در این حوزه است. در این پژوهش با استفاده از پکیج bibliometrix به‌مرور کلیه پژوهش‌های انجام‌شده در خصوص خوشه‌بندی پرداخته می‌شود. ضمن استخراج انواع معیارها و رویکردهای خوشه‌بندی به بررسی کاربردهای آن پرداخته‌شده است. این پژوهش با مروری جامع بر کلیه پژوهش‌های این حوزه می‌تواند به‌عنوان جعبه‌ابزاری در جهت ارائه انواع روش‌های خوشه‌بندی محققان را در ایده پردازی و انتخاب روش مناسب در طبقه‌بندی و تحلیل داده‌های مالی یاری دهد.

**کلیدواژه‌ها:** خوشه‌بندی، سری‌های زمانی مالی، شبکه‌های مالی، معیارهای فاصله‌ای، نگاشت دانش.

طبقه‌بندی موضوعی: ***G0, G11, G15***

---

[۱] دانشجوی دکتری مهندسی مالی، دانشگاه یزد، دانشکده اقتصاد، مدیریت و حسابداری، ایران. Nourahmadim@yahoo.com
[۲] کارشناسی ارشد، مدیریت مالی، دانشگاه یزد، دانشکده اقتصاد، مدیریت و حسابداری، ایران. Fatima.rasti@yahoo.com
[۳] استادیار مدیریت مالی، دانشگاه یزد، دانشکده اقتصاد، مدیریت و حسابداری، ایران. sadeqi@yazd.ac.ir



www.SID.ir

Archive of SID



## ۱. مقدمه

در سال‌های اخیر مسئله خوشه‌بندی و کشف روابط و روندهای موجود در سری‌های زمانی، موردتوجه بسیاری از پژوهشگران در حوزه‌های مختلف علوم ازجمله پژوهش‌های مالی قرارگرفته است. خوشه‌بندی سری‌های زمانی امکان تجزیه‌وتحلیل حجم زیادی از داده‌ها را در مقاطع زمانی مختلف، فراهم می‌کند (چاندرا و همکاران[1]،۲۰۰۸) به‌طورکلی خوشه‌بندی یکی از روش‌های داده‌کاوی است که در آن، داده‌های مشابه در گروه‌های مرتبط یا همگن قرار می‌گیرند(رای و همکاران[2]،۲۰۱۰). فرایند خوشه‌بندی شامل بخش‌بندی N داده در k گروه یا خوشه است. اگر مجموعه‌ای از داده‌ها به‌صورت $X = \{x_1, x_2, x_3, \ldots x_N\}$ در نظر گرفته شود، داده‌هایی که دارای بیشترین شباهت به یکدیگر هستند در یک خوشه و داده‌هایی که شباهت کمتری دارند در دیگر خوشه‌ها طبقه‌بندی می‌شوند و بسته به اینکه از کدام روش خوشه‌بندی استفاده می‌شود ترکیب خوشه‌ها می‌تواند متفاوت باشد(ایزوگوو و همکاران[3]،۲۰۲۰). خوشه‌بندی داده‌های سری زمانی معمولاً برای کشف الگوهای موجود در مجموعه سری‌های زمانی مورداستفاده قرار می‌گیرد(وانگ و همکاران[4]،۲۰۰۲). این فرایند خود به دو بخش مجزا تقسیم می‌شود، بخش اول شامل یافتن الگوهایی است که به‌صورت مکرر در مجموعه‌های سری زمانی رخ می‌دهد(فو و همکاران[5]،۲۰۰۱)، (چیو و همکاران[6]،۲۰۰۳) و بخش دوم روش‌هایی است که به کشف الگوهایی که به‌ندرت در سری‌های زمانی رخ می‌دهند می‌پردازد و همچنین اتفاقاتی که دارای تأثیرات شگفت‌آوری بر روند سری‌های زمانی هستند را بررسی می‌کند(کاگ و همکاران[7]،۲۰۰۲)، (لنگ و همکاران[8]، ۲۰۰۹). به‌طورکلی تجزیه‌وتحلیل سری‌های زمانی به سه هدف انجام می‌شود(پولتز و همکاران[9]،۲۰۰۳).

۱- شناخت تغییرات پویا در سری‌های زمانی: تشخیص همبستگی بین سری‌های زمانی، به‌عنوان‌مثال: در پایگاه داده‌های مالی می‌توان از آن برای یافتن شرکت‌هایی با حرکت مشابه سهام استفاده کرد (هی و همکاران[10]، ۲۰۱۲).

---

[1] Chandra, Gupta, and Gupta 2008
[2] Rai and Singh 2010
[3] Ezugwu et al. 2020
[4] Wang et al. 2002
[5] Fu et al. 2001
[6] Chiu, Keogh, and Lonardi 2003
[7] Keogh, Lonardi, and Chiu 2002
[8] Leng et al. 2009
[9] Polz, Hortnagl, and Prem 2003
[10] He et al. 2012



www.SID.ir

Archive of SID


۲- پیش‌بینی و توصیه: یک روش ترکیبی از ادغام خوشه‌بندی و تقریب عملکرد در هر خوشه می‌تواند به تحلیلگران در پیش‌بینی و توصیه کمک کند به‌عنوان‌مثال، در پایگاه‌های اطلاعاتی علمی، می‌تواند در حل مشکلاتی مانند یافتن الگوهای باد مغناطیسی خورشیدی برای پیش‌بینی الگوی امروز کمک کند (استفسس و همکاران[1]،۲۰۰٤)، (گریوز و همکاران[2]،۲۰۱۰).

۳- کشف الگو: باهدف کشف الگوهای جالب در پایگاه داده. به‌عنوان‌مثال، در پایگاه داده بازاریابی، می‌توان الگوهای مختلف فروش روزانه یک محصول خاص را در یک فروشگاه کشف کرد (آقابزرگی و همکاران[3]، ۲۰۱۵).

هدف از این پژوهش، مروری سیستمیک بر پژوهش‌های انجام‌شده در خصوص خوشه‌بندی و شبکه‌های مالی با تأکید بر موضوعات مالی و بررسی و تحلیل کاربردهای مختلف این روش در حوزه مالی با استفاده از روش bibliometrics می‌باشد. استفاده از bibliometrics یا نگاشت دانش امروزه در همه رشته‌ها توسعه‌یافته است.

ازآنجایی‌که تعداد مقالات علمی منتشرشده در حوزه‌های مختلف علمی به‌طور روزافزون در حال افزایش است و حتی ممکن است به‌روزرسانی دائمی اطلاعات با انتشار مقالات جدید امری غیرممکن باشد علاوه بر این، سهم پژوهش‌های تجربی که منجر به جریان‌های تحقیقاتی گسترده شده است مورد تأکید است (برینر و دنیر[4]،۲۰۱۲). همچنین مرور متون نقش مهمی را در ترکیب یافته‌های پژوهش‌های گذشته به‌منظور استفاده مؤثر از پایگاه دانش موجود، پیشبرد یک تحقیق و ارائه بینشی مبتنی بر شواهد، همچنین قابل‌آزمایش و قضاوت حرفه‌ای ایفا می‌کند (روسو[5]،۲۰۱۲).

پژوهشگران از روش‌های کمی و کیفی متفاوتی به‌منظور بررسی رویکردها برای درک و سازمان‌دهی یافته‌های اولیه استفاده می‌کنند. در این میان، bibliometrics بالقوه توانایی معرفی یک فرایند مروری نظام‌مند، شفاف و تکرارپذیری را بر اساس اندازه‌گیری آماری علم، متخصصان و فعالیت‌های علمی را دارد (برادوس[6]،۱۹۸۷)،

[1] Sfetsos and Siriopoulos 2004
[2] Graves and Pedrycz 2010
[3] Aghabozorgi, Shirkhorshidi, and Wah 2015
[4] Briner & Denyer,
[5] Rousseau,2012
[6] Broadus, R. N. (1987)



www.SID.ir

Archive of SID



(پریتچارد[1]،۱۹۶۹)، (دیوداتو[2]،۱۹۹٤) برخلاف روش‌های دیگر bibliometrics، تحلیل‌های عینی و قابل اطمینان‌تری را فراهم می‌کند، درواقع حجم عظیم اطلاعات جدید، پیشرفت‌های مفهومی و داده‌ها، محیطی هستند که در آن bibliometrics با ارائه یک تحلیل ساختاریافته برای مجموعه بزرگی از اطلاعات، به‌منظور استنباط روندها در طول زمان، موضوعات موردپژوهش، شناسایی تغییرات و همچنین شناسایی بسیاری از محققان و مؤسسات و به‌طورکلی ارائه «تصویر بزرگ» از تحقیقات موجود بسیار مفید خواهد بود (کرین[3]،۱۹۷۲).

در بخش بعدی، ابتدا به معرفی روش رویکرد نگاشت دانش پرداخته و سپس با استفاده از پکیج bibliometrix از کتابخانه R خروجی‌های مرتبط با موضوع پژوهش استخراج می‌شود. در بخش سوم، با استفاده از تحلیل نتایج مهم‌ترین مقالات این حوزه، انواع الگوریتم‌های خوشه‌بندی، انواع رویکردهای خوشه‌بندی پرداخته شده است. در بخش چهارم مقاله به کاربردهای خاص خوشه بندی در حوزه مالی پرداخته شده و نهایتاً در بخش آخر به جمع‌بندی و نتیجه‌گیری ارائه‌شده است.

## ۲. روش انجام کار

مسئله خوشه‌بندی شامل مجموعه‌ای از اشیاء و مجموعه‌ای از ویژگی‌های مرتبط با آن اشیاء است. هدف این است که با استفاده از این ویژگی‌ها گروه‌هایی (خوشه‌هایی) تشکیل شوند که دارای شباهت‌های درون‌گروهی حداکثر و بین گروهی حداقل باشند. ازآنجایی‌که نمونه‌ای برای کمک به حل این الگوریتم ارائه نمی‌شود این روش یکی از روش‌های یادگیری بدون نظارت است.

در این پژوهش از رویکرد نگاشت دانش به‌منظور بررسی پیشینه پژوهش‌های انجام‌شده در خصوص کاربرد سیستم‌های توصیه‌کننده در مباحث مالی استفاده می‌شود. نحوه اجرای فرایند نگاشت علم توسط (بورنر و همکاران[4]،۲۰۰۳) توصیف‌شده است. (کوبو و همکاران[5]،۲۰۱۱) نرم‌افزارهای نگاشت علم را با استفاده از فرایندی مشابه مقایسه نمودند. رویکرد استاندارد این فرایند شامل پنج مرحله است (زوپیک و همکاران[6]،۲۰۱۵):

[1] Pritchard, A. (1969)
[2] Diodato,1994
[3] Crane, 1972
[4] Börner, Chen, and Boyack 2003
[5] Cobo et al. 2011
[6] Zupic and Čater 2015



www.SID.ir

Archive of SID


۱. طرح مطالعه؛

۲. جمع‌آوری داده‌ها؛

۳. تجزیه‌وتحلیل داده‌ها؛

٤. مصورسازی داده‌ها؛

۵. تفسیر (آریا و همکاران[1]،۲۰۱۷).

در طرح مطالعه، محققان سؤال (های) پژوهش را تعریف کرده و روش‌های مناسب کتاب‌شناختی که می‌توانند پاسخگوی سؤال (ها) باشند را انتخاب می‌نمایند. سه نوع کلی از سؤالات پژوهش را می‌توان با استفاده از کتاب‌شناختی برای نگاشت علم پاسخ داد:

۱) شناسایی پایگاه دانش موضوع یا زمینه تحقیق و ساختار ذهنی آن؛

۲) بررسی حیطه پژوهش (ساختار مفهومی) موضوع یا زمینه تحقیق؛ و

۳) ایجاد یک ساختار شبکه اجتماعی از یک جامعه علمی خاص (والتمن[2]،۲۰۱٦).

سؤال اصلی این پژوهش این است که موارد استفاده از خوشه‌بندی و تحلیل‌های شبکه‌ای در مالی کدام است؟ تاکنون چه پژوهش‌هایی در این حوزه انجام‌شده است؟

مراحل پیاده‌سازی نگاشت علم برای کاربرد خوشه‌بندی در حوزه مالی به شرح زیر است:

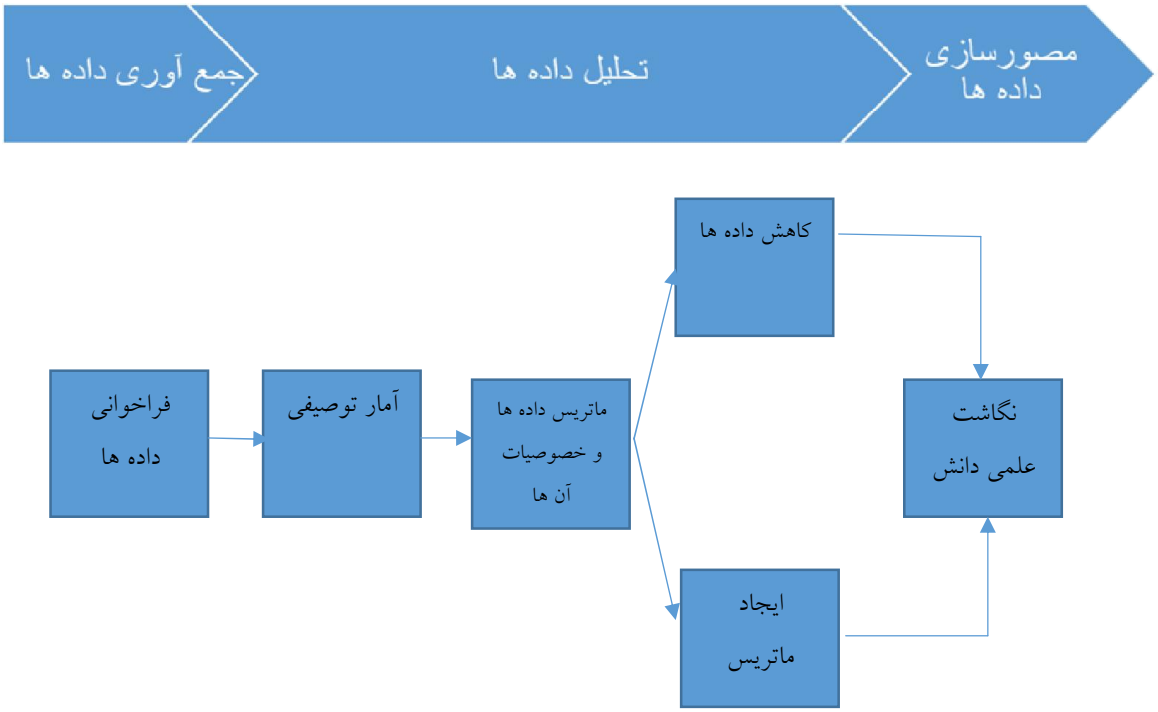
[1] Aria and Cuccurullo 2017
[2] Waltman 2016



www.SID.ir

Archive of SID






شکل (۱) مراحل نگاشت علم (آریا و همکاران،۲۰۱۷)

با توجه به شکل (۱) مراحل نگاشت علم، شامل سه مرحله جمع‌آوری داده‌ها، تجزیه‌وتحلیل داده‌ها و مصورسازی داده‌ها است که در ادامه به تشریح این مراحل پرداخته می‌شود:

جمع‌آوری داده شامل سه مرحله است. ابتدا بازیابی داده‌ها انجام می‌شود. بسیاری از پایگاه‌های کتاب‌شناختی آنلاین که پایگاه ذخیره داده‌ها هستند، می‌تواند به‌عنوان منبع کتاب‌شناختی مورداستفاده قرار گیرند، مانند (کوبو و همکاران،۲۰۱۱):

Clarivate Ana-lytics Web of Science (WoS at http://www.webofknowledge.com), Scopus (http://www.scopus.com), Google Scholar(http://scholar. google.com), and .Science Direct (http://www.sciencedirect.com/)



www.SID.ir

Archive of SID


برای درک بهتر موضوع کاربرد خوشه‌بندی و شبکه‌سازی مالی ابتدا کلیدواژه "clustering" AND "financial time series" AND "network" را در اسکوپوس در تاریخ ۷ نوامبر ۲۰۲۰ جستجو کرده و با استفاده از پکیج bibliometrix در نرم‌افزار R خروجی‌های زیر را دریافت شده است.

آمار توصیفی پژوهش‌های انجام‌شده در خصوص این موضوع به شرح جدول زیر است:

| توصیف | نتایج |
|---|---|
| **اطلاعات اصلی در مورد مقاله‌های مستخرج شده از اسکوپوس** | |
| **دوره زمانی** | ۱۹۹۸:۲۰۲۰ |
| **نوع منبع (ژورنال، کتاب و ...)** | ٤١ |
| **اسناد** | ٦٤ |
| **میانگین انتشار مقاله در هرسال** | ۷/۹۵ |
| **میانگین ارجاع دهی به مقالات در هرسال** | ۱۱/۲ |
| میانگین ارجاع به هر مقاله | ۱/۱۷۲ |
| منابع | ١٤٣٧ |
| **نوع مقاله‌ها** | |
| مقاله | ۳۱ |
| کنفرانسی | ۹ |
| کنفرانس مروری | ۲۳ |
| مروری | ۱ |
| **محتوای اسناد** | |
| کلیدواژه | ۳۹٤ |
| **نویسندگان** | |
| نویسنده | ۱۲٦ |
| نویسندگان دارای یک مقاله | ۳ |
| نویسندگان دارای چند مقاله | ۱۲۳ |

جدول (۱) آمار توصیفی مرتبط با مقاله

با توجه به جدول (۱)، ٦٤ مدرک توسط ۱۳۲ نویسنده بین سال‌های ۱۹۹۸ تا ۲۰۲۰ با محوریت خوشه‌بندی و شبکه‌سازی سری‌های زمانی مالی انجام شده است که از این بین ۳۱ مورد آن مقاله است.

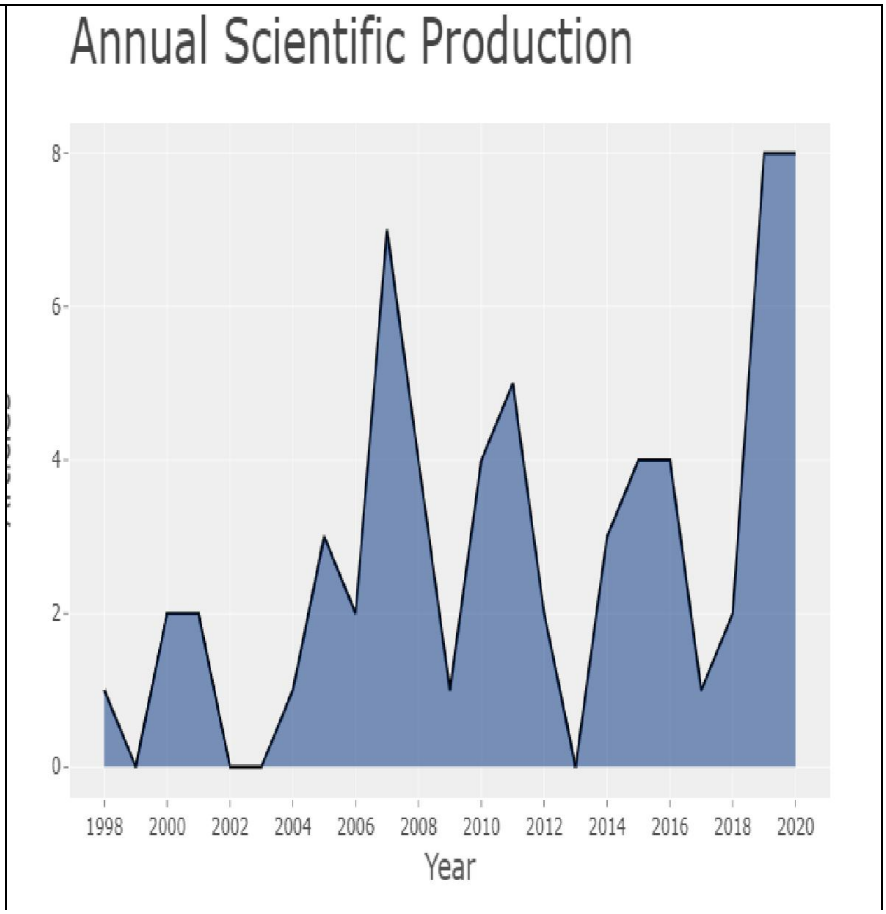


www.SID.ir

Archive of SID



| سال | مقاله |
|---|---|
| ۱۹۹۸ | ۱ |
| ۲۰۰۰ | ۲ |
| ۲۰۰۱ | ۲ |
| ۲۰۰۴ | ۱ |
| ۲۰۰۵ | ۳ |
| ۲۰۰۶ | ۲ |
| ۲۰۰۷ | ۷ |
| ۲۰۰۸ | ۴ |
| ۲۰۰۹ | ۱ |
| ۲۰۱۰ | ۴ |
| ۲۰۱۱ | ۵ |
| ۲۰۱۲ | ۲ |
| ۲۰۱۴ | ۳ |
| ۲۰۱۵ | ۴ |
| ۲۰۱۶ | ۴ |
| ۲۰۱۷ | ۱ |
| ۲۰۱۸ | ۲ |
| ۲۰۱۹ | ۸ |
| ۲۰۲۰ | ۸ |




نمودار (۱) تعداد پژوهش‌های انجام‌شده در هرسال

در نمودار (۲) پژوهش‌های انجام‌شده به تفکیک هر کشور نشان داده‌شده است. همان‌گونه که ملاحظه می‌شود بیشتر پژوهش‌ها توسط چین (۳۹ مورد) و سپس آلمان و ایتالیا (٦ مورد) انجام‌شده است.

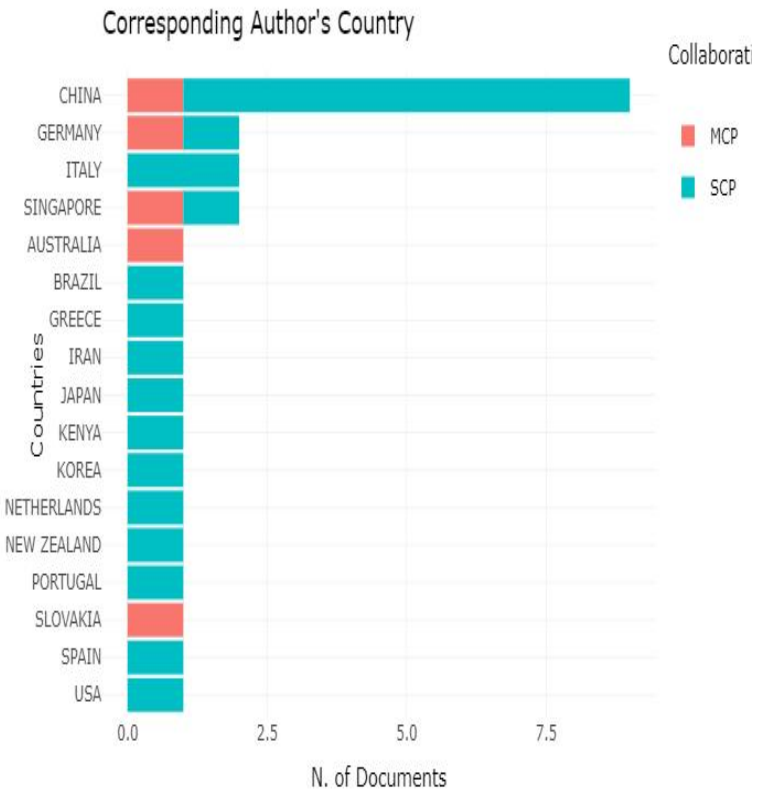


www.SID.ir

Archive of SID



| کشور | فراوانی |
|---|---|
| چین | ۳۹ |
| آلمان | ٦ |
| ایتالیا | ٦ |
| ژاپن | ٤ |
| ترکیه | ٤ |
| آمریکا | ٤ |
| کنیا | ۳ |
| کره جنوبی | ۳ |
| استرالیا | ۲ |
| برزیل | ۲ |
| ایران | ۲ |
| نیوزیلند | ۲ |
| پرتغال | ۲ |
| سنگاپور | ۲ |
| اسپانیا | ۲ |
| انگلستان | ۲ |
| کانادا | ۱ |




نمودار (۲) پژوهش‌های انجام‌شده به تفکیک هر کشور

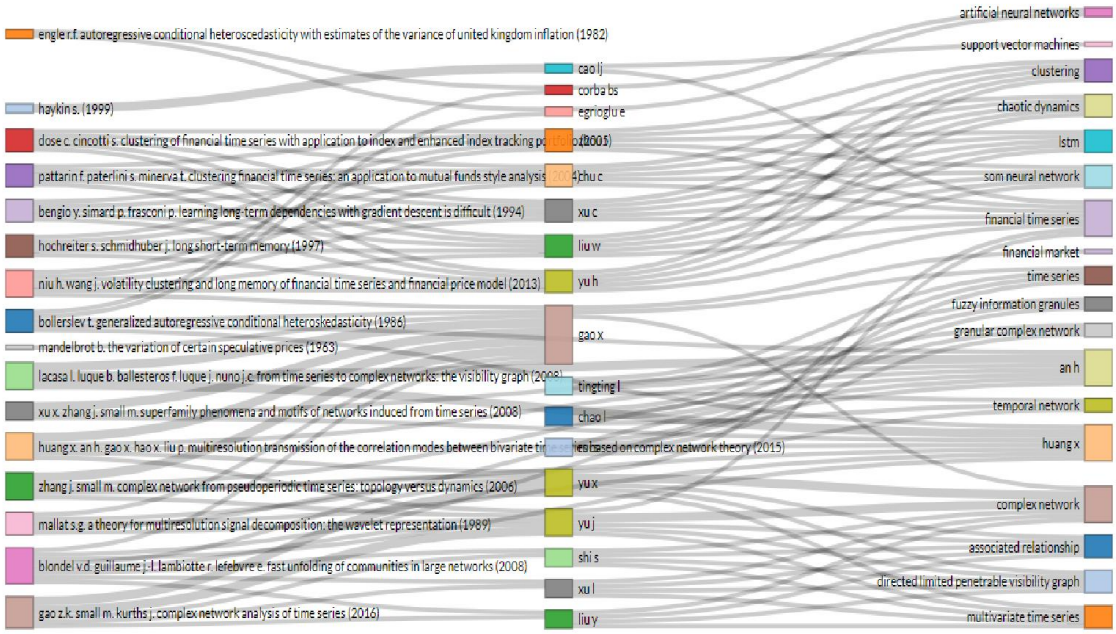
در نمودار (۳) شبکه ارتباط بین نویسندگان، واژگان کلیدی و عناوین ارائه‌شده است به‌نحوی‌که در سمت چپ کلیدواژه‌ها، در وسط عناوین و در سمت راست نویسندگان مربوط به پژوهش‌های برتر در این زمینه ارائه‌شده است.



www.SID.ir

Archive of SID






نمودار (۳) شبکه ارتباط بین نویسندگان، واژگان و عناوین پژوهش‌های انجام‌شده



www.SID.ir

Archive of SID



| کلیدواژه | دفعات تکرار |
|---|---|
| تحلیل سری زمانی | ۲۴ |
| پردازش داده‌های مالی | ۲۳ |
| سری زمانی | ۱۵ |
| سری زمانی مالی | ۱۴ |
| شبکه‌های عصبی | ۱۳ |
| مالی | ۱۱ |
| پیش‌بینی | ۱۰ |
| الگوریتم خوشه‌بندی | ۸ |
| الگوریتم ژنتیک | ۷ |
| مدل‌های ریاضی | ۷ |
| تحلیل خوشه | ۶ |
| تجارت | ۶ |
| شبکه‌های عصبی جلو رونده | ۶ |
| بازارهای مالی | ۶ |
| خوشه‌بندی | ۵ |
| شبکه‌های پیچیده | ۵ |
| برنامه‌ریزی احتمال پویا | ۵ |
| برنامه‌ریزی ژنتیک | ۵ |
| بهینه‌سازی کلی | ۵ |
| سرمایه‌گذاری | ۵ |

نمودار (٤) مهم‌ترین کلیدواژه‌های بکار رفته در مقالات

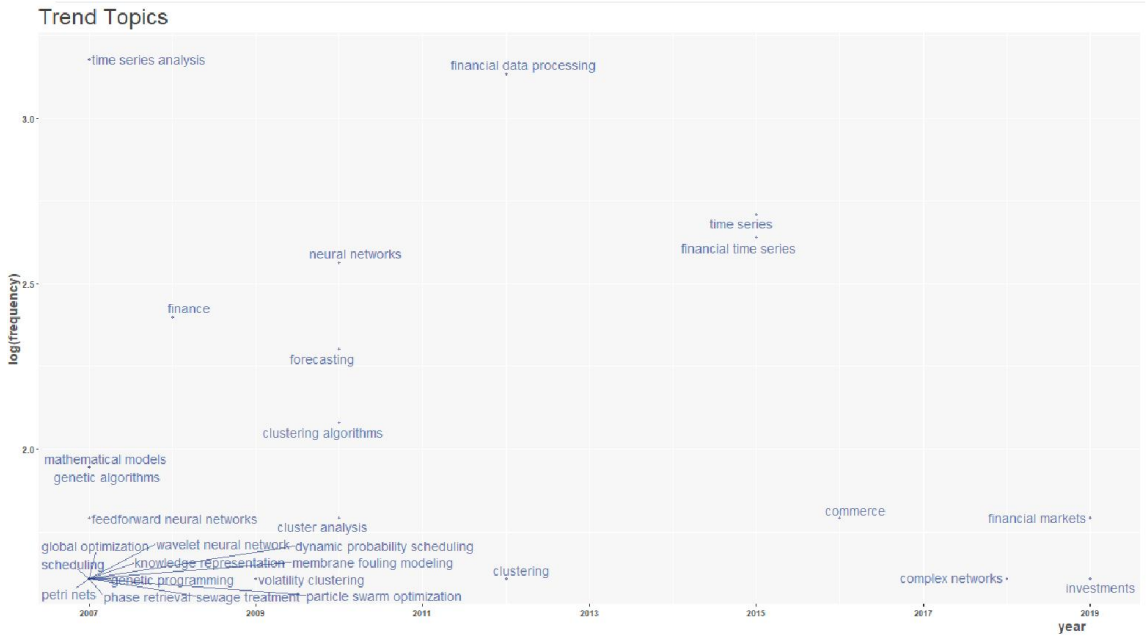
در نمودار (۵) سیر روند توجه نویسندگان به کلید واژگان در طی دوره‌های زمانی مختلف نشان داده‌شده است.



www.SID.ir

Archive of SID






نمودار (۵) روند اهمیت کلیدواژه‌ها برحسب دوره زمانی

## ۳. یافته‌ها

### ۳.۱. معیارهای اندازه‌گیری فاصله در پژوهش‌ها

یکی از مسائل کلیدی در امر خوشه‌بندی انتخاب مناسب‌ترین معیار فاصله/(عدم) شباهت هست چراکه نتایج خوشه‌بندی به‌طور قابل‌ملاحظه‌ای تحت تأثیر معیار به کار گرفته‌شده برای اندازه‌گیری شباهت میان سری‌های زمانی موردبررسی، قرار می‌گیرد. این معیارها که به معیارهای متریک نیز شهرت دارند به تعیین میزان فاصله‌ی میان دو داده‌ی موردبررسی در یک فضای متریک می‌پردازند. به‌علاوه فواصلی که با استفاده از معیارهای متریک در فضای موردبررسی تحت عنوان فضای متریک محاسبه‌شده‌اند، لزوماً شرایط زیر را برآورده می‌سازند:

۱. مثبت بودن: $d(A,B) \geq 0 ; d(A,B) = 0 \text{ if } A = B$

۲. متقارن بودن: $d(A,B) = d(B,A)$

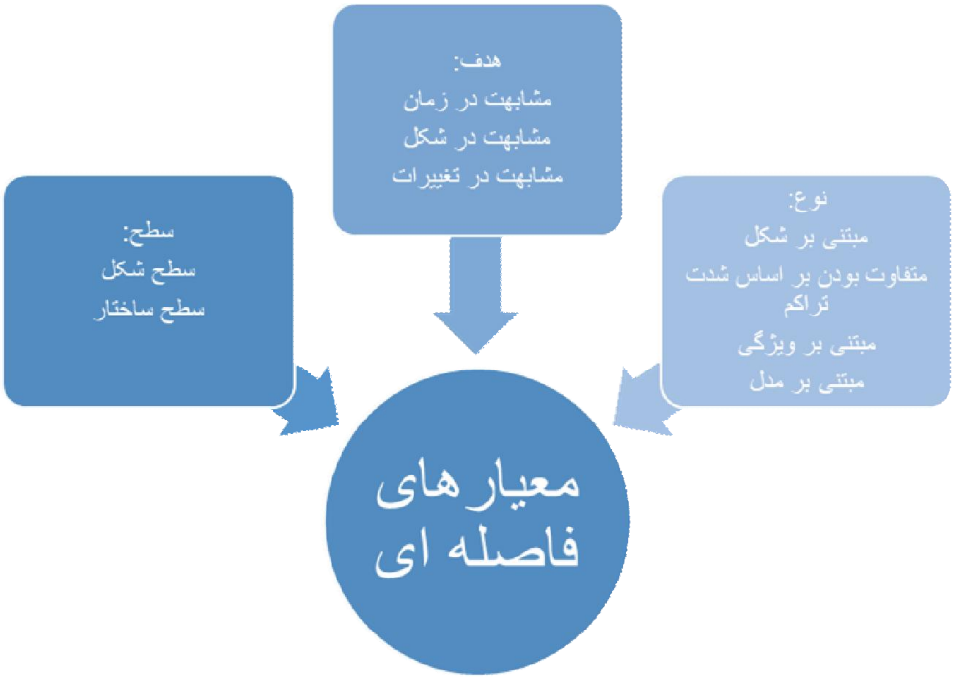
۳. برقراری نامساوی مثلثی: $d(A,C) \leq d(A,B) + d(B,C)$



www.SID.ir

Archive of SID


در شکل زیر دسته‌بندی معیارهای فاصله‌ای بر اساس نوع، هدف و سطح نشان داده‌شده است.




شکل (۲) معیارهای اندازه‌گیری فاصله در تحقیقات (آقابزرگی و همکاران، ۲۰۱۵)

### ۳.۲. انواع الگوریتم خوشه‌بندی

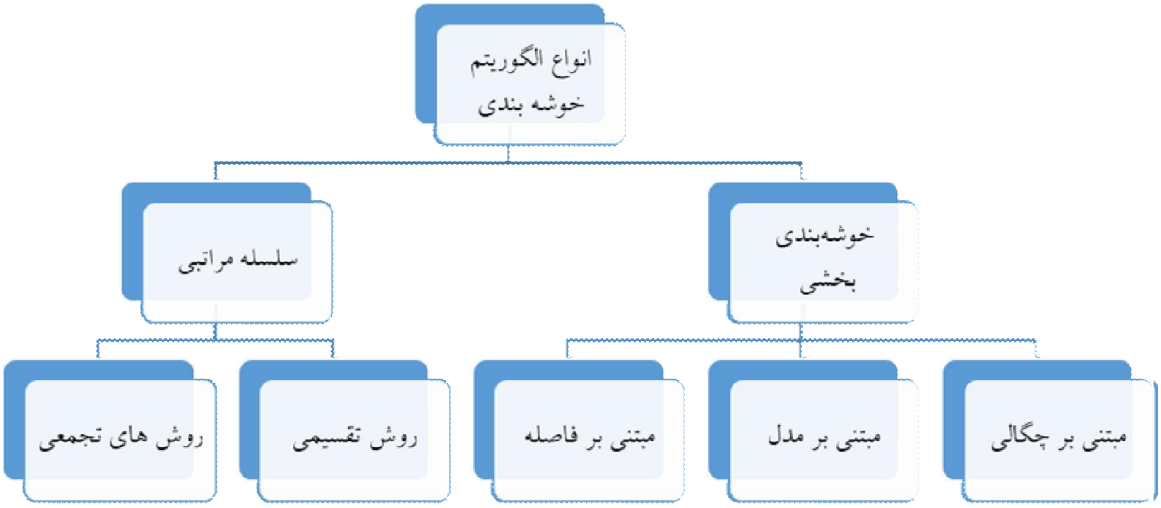
در این قسمت رویکردهای کلی الگوریتم‌های خوشه‌بندی ارائه می‌گردد.



www.SID.ir

Archive of SID






شکل (۳) انواع الگوریتم‌های خوشه‌بندی (ساکسنا و همکاران[1]،۲۰۱۷)

خوشه‌بندی به دودسته بخشی (جزء بندی) و سلسله مراتبی تقسیم‌بندی می‌شود که در ادامه به تعریف و دسته‌بندی هرکدام می‌پردازیم.

**خوشه‌بندی بخشی (جزء بندی):**

تقسیم مجموعه داده‌ها به زیرمجموعه‌های بدون همپوشانی به‌گونه‌ای که هر داده دقیقاً در یک زیرمجموعه قرار داشته باشد.

**خوشه‌بندی سلسله مراتبی به دودسته تقسیم می‌شود:**

- **روش‌های تجمعی (Agglomerative) (روش پایین به بالا)**

با هر داده در یک خوشه شروع می‌کند. به‌طوری تکراری، در هر مرحله خوشه‌های نزدیک به هم را ترکیب می‌کند تا درنهایت یک خوشه باقی بماند.

- **روش‌های تقسیمی (Divisible) (روش بالا به پایین)**

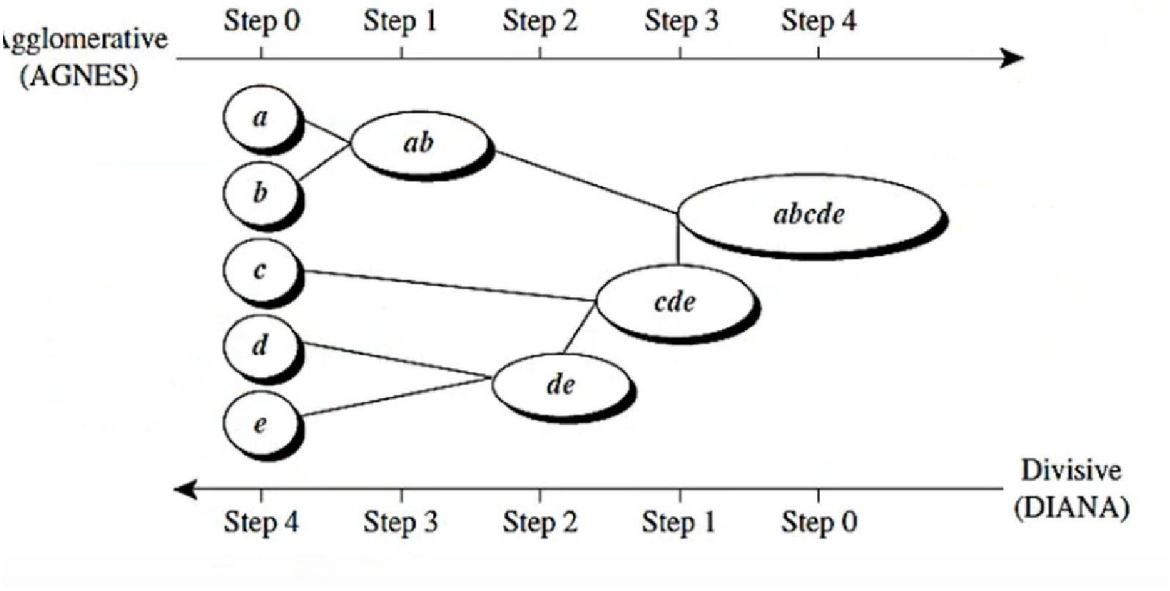
[1] Saxena et al. 2017



www.SID.ir

Archive of SID


با کل داده‌ها به‌عنوان یک خوشه شروع می‌کند. به‌طور تکراری، داده‌ها را دریکی از خوشه‌ها تقسیم می‌کند تا هنگامی‌که فقط یک داده در هر خوشه باشد.




شکل (٤) خوشه‌بندی سلسله مراتبی

### ۳.۳. انواع رویکردهای خوشه‌بندی

فریلی و رفتری در پژوهشی رویکردهای خوشه‌بندی را به دو گروه خوشه‌بندی سلسله مراتبی و خوشه‌بندی بخشی (جزء بندی) طبقه‌بندی کردند (فریلی و رفتری[1]،۱۹۹۸). روش‌های خوشه‌بندی بخشی، یک سطح (بدون تودرتو) از اشیا را ایجاد می‌کند (هر شیء فقط به یک خوشه تعلق دارد). روش‌های سلسله مراتبی یک توالی تودرتو از اشیا را با یک خوشه کلی از بالا و خوشه‌های انفرادی تا پایین تشکیل می‌دهند. این الگوریتم‌ها می‌تواند تقسیمی و یا تجمیعی باشند. الگوریتم تقسیمی از رویکرد از بالا به پایین استفاده می‌کند به این صورت که خوشه‌ی حاوی همه‌ی اشیا را تا زمانی که هر شیء به یک خوشه تعلق گیرد به خوشه‌های کوچک‌تر تجزیه می‌کند. الگوریتم تجمیعی رویکردی از پایین به بالا دارد، به صورتی که خوشه‌بندی‌ها با یک شی انفرادی آغازشده و با ادغام خوشه‌ها به خوشه‌های بزرگ‌تر ادامه می‌یابد تا زمانی که درنهایت در یک خوشه قرار داشته

[1] Fraley and Raftery 1998



www.SID.ir

Archive of SID



باشند. روش خوشه‌بندی سلسله مراتبی معمولاً منجر به تشکیل دندروگرام می‌شود. روش خوشه‌بندی سلسله مراتبی را می‌توان بر اساس معیارهای مشابهت و یا پیوند در سه‌طبقه‌ی: خوشه‌بندی پیوند واحد، خوشه‌بندی کامل و خوشه‌بندی میانگین پیوند تقسیم کرد همچنین با محدود کردن رشد درخت سلسله مراتبی، می‌توان خوشه‌بندی سلسله مراتبی را به خوشه‌بندی بخشی تبدیل کرد اما عکس آن وجود ندارد (دی پرادو[۱]،۲۰۲۰)، (جین و همکاران[۲]،۱۹۹۹).

رویکرد مبتنی بر شکل به‌عنوان یک رویکرد مبتنی برداده خام شناخته‌شده است. این رویکرد معمولاً از روش‌های مرسوم خوشه‌بندی که با داده‌های استاتیک سازگارند استفاده می‌کند. در رویکرد مبتنی بر ویژگی، از محاسبه‌ی فاصله اقلیدسی بردارهای سری‌های زمانی استفاده می‌شود(هااوتاماکی و همکاران[۳]،۲۰۰۸). روش‌های مبتنی بر شبکه،با کمی کردن فضای جسم به تعداد محدودی از سلول‌ها که ساختار شبکه‌ای را تشکیل می‌دهند و تمام عملیات خوشه‌بندی بر روی آن انجام می‌شود ، اقدام به خوشه‌بندی داده‌ها می‌کند. روش‌های مبتنی بر مدل ابتدا برای هر یک از خوشه‌ها مدلی را در نظر گرفته و داده‌ها را به بهترین وجه با مدل فرضی تطبیق می‌دهد. از روش‌های عمده مبتنی بر مدل می‌توان به رویکرد آماری و روش شبکه عصبی اشاره کرد(چیسمن و همکاران[۴]،۱۹۹٦)، که از تحلیل آماری بیزی برای تخمین تعداد خوشه‌ها استفاده می‌شود همچنین روش‌های خوشه‌بندی شبکه عصبی به روش یادگیری رقابتی و روش نقشه‌های ویژگی‌های خودسازمان‌دهی تقسیم می‌شوند که در خوشه‌بندی مورداستفاده قرار می‌گیرند(کوهینن[۵]،۱۹۹۰).

الگوریتم K-means یکی از شناخته‌شده‌ترین الگوریتم‌های خوشه‌بندی است که بیشتر برای حل مشکلات خوشه‌بندی استفاده می‌شود .در این روش مجموعه داده‌ها از طریق تعداد خوشه‌های تعریف‌شده توسط کاربر طبقه‌بندی می‌شود (تئودوریس و همکاران[٦]،۲۰۱۳)، (مک کوئین و همکاران[۷]،۱۹٦۷).

[1] De Prado 2020
[2] Jain, Murty, and Flynn 1999
[3] Hautamaki, Nykanen, and Franti 2008
[4] Cheeseman and Stutz 1996
[5] Kohonen 1990
[6] Theodoridis and Chellappa 2013
[7] MacQueen 1967



www.SID.ir

Archive of SID

نمودار (٦) رویکردهای خوشه‌بندی سری‌های زمانی (آقابزرگی و همکاران،۲۰۱۵)

در پژوهش‌های متعدد روش‌های گوناگونی به‌منظور بررسی انواع فاصله‌ی سری‌های زمانی استفاده‌شده است. در این بخش به مروری بر رویکردهای مورداستفاده در نمایش داده‌های سری‌های زمانی پرداخته می‌شود.

| روش | ویژگی (ها) | معیار فاصله | مقاله |
|---|---|---|---|
| مبتنی بر مدل | نه‌تنها قادر به درک وابستگی بین متغیرها است، بلکه همبستگی پیاپی در اندازه‌گیری‌ها را مشخص می‌کند. | مدل‌های پنهان مارکوف[1] | (Smyth 1997) |
| | تقویت نوفه[2] | LCSS<br>طولانی‌ترین زیر دنباله مشترک[3] | (Vlachos, Kollios, and Gunopulos 2002), (Banerjee and Ghosh 2001) |
| | معیار مبتنی بر آستانه اندازه‌گیری، در نظر گرفتن فواصل زمانی، | جستار آستانه‌ای[1] | (Aßfalg et al. |

[1] Hidden Markov models
[2] Noise robustness
[3] Longest Common Sub-Sequence



www.SID.ir

Archive of SID



| | | | |
|---|---|---|---|
| | مقایسه سری‌های زمانی هنگامی‌که از مقدار آستانه‌ی مشخصی فراتر رود بجای آنکه از مقادیر مشخصی استفاده شود. | | 2006) |
| | معیار مبتنی بر الگو | فاصله مونتاژ فضایی[۲] | (Chen et al. 2007) |
| مبتنی بر شکل | معیار منعطف (از یکی به همه، از یکی به هیچ‌کدام) مناسب برای رانش زمانی<br><br>دارای دقت بهتر از فاصله اقلیدسی<br><br>کارایی کمتر نسبت به فاصله اقلیدسی و شباهت مثلث. | پیچش زمانی پویا[۳] | (Sakoe 1971), |
| | معیار همگام (یک‌به‌یک) با استفاده از شاخص سازی، خوشه‌بندی، طبقه‌بندی، حساس به مقیاس‌گذاری | فاصله اقلیدسی[٤] | (Faloutsos, Ranganathan, and Manolopoulos 1994) |
| | کاهش نوفه، قادر به خلاصه کردن ساختار موقت | همبستگی متقابل مبتنی بر فواصل[٥] | (Golay et al. 1998) |
| | تقویت نوفه، انتقال و مقیاس‌گذاری داده، از نقطه مرجع ثابت استفاده‌شده است. | فاصله ویرایش با جریمه واقعی[٦] | (Chen and Ng 2004) |
| | به‌طور خودکار از داده‌های پرت جلوگیری می‌کند. | مشابه حداقل واریانس[۷] | (Chen, Özsu, and Oria 2005) |
| | معیار منعطف (یکی به همه، یکی به هیچ‌کدام)<br><br>استفاده از الگوی آستانه‌ای. | ویرایش فاصله در توالی‌های حقیقی | (Latecki et al. 2005) |
| | استفاده از هیستوگرام سری‌های زمانی چند مقیاسی | مبتنی بر هیستوگرام[۸] | (Chen and Özsu 2005) |
| | مناسب برای نمونه‌گیری‌های مختلف | DISSIM | (Frentzos, Gratsias, and Theodoridis 2007) |
| | نمره تشابه هم بر اساس پاداش تطابق و هم در جریمه عدم تطابق | مدل هم‌راستایی توالی موزون[۱] | (Morse |

[1] Threshold Queries
[2] Spatial Assembling Distance (SpADe)
[3] Dynamic Time Warping (DTW)
[4] Euclidean distance
[5] Cross-correlation based distances
[6] Edit Distance with Real Penalty (ERP)
[7] Minimal Variance Matching
[8] Histogram-based



www.SID.ir

Archive of SID



| | | | |
|---|---|---|---|
| | | | and Patel 2007) |
| | ازنظر مقیاس و موقعیت مکانی داده‌ها بی‌تغییر است | ضریب همبستگی پیرسون و فواصل مرتبط | (Lee Rodgers and Nicewander 1988) |
| | این روش شامل بازه‌های زمانی یا "پنجره‌ها" با اندازه‌های مختلف است؛ اما نحوه تعیین این "پنجره‌ها" مشخص نیست. | نرمال‌سازی جزئی | (Indyk, Koudas, and Muthukrishnan 2000) |
| | اندازه‌گیری طیفی که معکوس تبدیل فوریه[2] از طیف دامنه لگاریتمی کوتاه‌مدت است | کپستروم[3] | (Kalpakis, Gada, and Puttagunta 2001) |
| | قادر به خوشه‌بندی الگوهای فصلی | فاصله مبتنی بر احتمال[4] | (Kumar, Patel, and Woo 2002) |
| **متفاوت بودن بر اساس شدت تراکم** | در پژوهش koegh و همکاران، روش اندازه‌گیری فاصله پارامتر مبتنی بر تئوری پیچیدگی v پیشنهادشده است. | معیار متفاوت بودن بر اساس شدت تراکم | (Keogh et al. 2007) |
| | Lang و همکاران درجه فشرده‌سازی دیکشنری را برای اندازه‌گیری شباهت ایجاد کردند و پیشنهاد روش فشرده‌سازی مبتنی بر دیکشنری را برای محاسبه شباهت سری‌های زمانی طولانی‌مدت ارائه دادند. | متفاوت بودن مبتنی بر دایره المعارف[5] | (Lang, Morse, and Patel 2009) |
| **مبتنی بر ویژگی** | حساس به مقیاس‌گذاری<br>علیرغم مقادیر قطعی قادر به درک سوابق موقت است. | فاصله سری‌های زمانی کوتاه | (Möller-Levet et al. 2003) |

جدول (۱) مروری بر رویکردهای مورداستفاده در نمایش داده‌های سری‌های زمانی

[1] SequenceWeighted Alignment model
[2] Fourier transform
[3] Cepstrum
[4] Probability-based distance
[5] Dictionary-based compression



www.SID.ir

Archive of SID



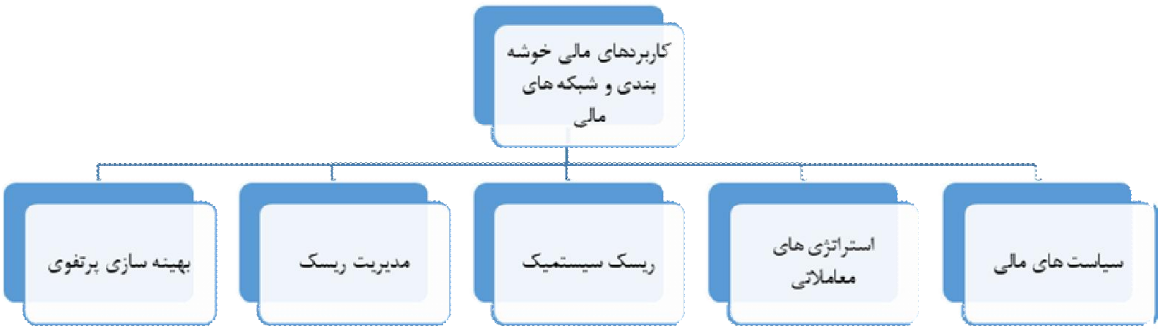
## ٤. کاربردهای مالی و خوشه‌بندی

روش خوشه‌بندی به‌طور گسترده‌ای در علوم مختلف و برای مقاصد گوناگونی به کار گرفته می‌شود که در ادامه به بیان برخی از کاربردهای آن می‌پردازیم:

| کاربرد | توضیح |
|---|---|
| **بازاریابی** | خوشه‌بندی مشتریان با توجه به رفتارها و نیازهای آن‌ها از طریق مجموعه ویژگی‌های مشتریان و نیز خرید آن‌ها. |
| **زیست‌شناسی** | خوشه‌بندی حیوانات و گیاهان با توجه به ویژگی‌های آن‌ها. |
| **کتابداری** | خوشه‌بندی کتاب‌ها بر اساس ویژگی‌های کتاب‌شناسی. |
| **نقشه‌برداری شهری** | خوشه‌بندی خانه‌ها بر اساس نوع و موقعیت جغرافیایی آن‌ها. |
| **مطالعات زلزله‌نگاری** | تشخیص مناطق حادثه خیز بر اساس مشاهدات قبلی. |
| **وب** | خوشه‌بندی اسناد (برای مثال اخبار) یا خوشه‌بندی مشتریان وب‌سایت‌ها. |
| **تشخیص گفتار** | خوشه‌بندی گفتار بر اساس گویندگان آن‌ها و یا فشرده‌سازی گفتار. |
| **تقسیم‌بندی تصاویر** | تقسیم‌بندی تصاویر پزشکی یا ماهواره‌ای. |
| **مطالعات مالی** | پیش‌بینی قیمت سهام، تعیین ارتباط میان سهام یا صنایع موجود در بورس اوراق بهادار و یا ارزهای موجود در بازار فارکس، تشکیل سبد سرمایه‌گذاری. |

جدول (۲) انواع کاربرد خوشه‌بندی در علوم مختلف

ازجمله کاربردهای خوشه‌بندی در حوزه مالی به شرح نمودار زیر است:




نمودار (۷) کاربردهای مالی خوشه‌بندی و شبکه‌های مالی، منبع: یافته‌های پژوهش



www.SID.ir

Archive of SID



### ۴.۱. کاربردهای خوشه‌بندی در پرتفوی

در جدول زیر به بررسی پیشینه پژوهش‌های انجام‌شده در خصوص کاربرد روش‌های خوشه‌بندی بر پرتفوی پرداخته‌شده است.

| عنوان مقاله | نویسنده (گان) | نکات کلیدی |
|---|---|---|
| **برابری ریسک سلسله مراتبی: حسابداری وابستگی‌های ناشی از تخصیص چندعاملی چند دارایی.** | (Lohre, Rother, and Schäfer 2020) | بررسی استراتژی‌های متنوع سازی بر اساس خوشه‌بندی سلسله مراتبی |
| **پویایی‌های همبستگی بازار: طبقه‌بندی و تجزیه‌وتحلیل پرتفوی** | (Onnela et al. 2003) | جنبه‌های فنی وابستگی اندازه پنجره و جنبه متنوع سازی پرتفوی بهینه موردبررسی قرار می‌گیرد. |
| **استراتژی پرتفوی پویا با استفاده از رویکرد خوشه‌بندی** | (Ren et al. 2017) | استراتژی پرتفوی پویا مبتنی بر ساختارهای متغیر شبکه‌های MST در بازار سهام چنین ارائه می‌شود. |
| **رویکرد شبکه‌ای برای انتخاب پرتفوی** | (Peralta and Zareei 2016) | هدف عمده این مقاله به‌کارگیری شبکه‌های مالی به‌عنوان ابزاری مفید برای انتخاب پرتفوی باهدف قرار دادن گروهی از دارایی‌ها با توجه معیار مرکزیت آن‌هاست. |
| **انتخاب پرتفوی مبتنی بر شبکه‌ها: آیا با پرتفوی بهینه مارکویتز همسو است؟** | (Hüttner, Mai, and Mineo 2018) | این مقاله رابطه ابتکاری بین مرکزیت شبکه و MVP از یک شباهت ساختاری بین دو مکانیسم انتخاب پرتفوی نشاءت نمی‌گیرد بلکه به دلیل ویژگی‌های خاص ماتریس‌های همبستگی مشاهده‌شده است. |
| **خوشه‌بندی سری‌های زمانی مالی با کاربرد آن در افزایش ردیابی شاخص پرتفوی** | (Dose and Cincotti 2005) | نتایج مبتنی بر خوشه‌بندی با نتایج تکنیک‌های تصادفی مقایسه می‌شود و اهمیت خوشه‌بندی را در کاهش نویز و پایداری روش‌های پیش‌بینی به‌ویژه برای ردیابی شاخص پیشرفته نشان می‌دهد. |
| **ریسک ارتباط متقابل و مدیریت پرتفوی فعال** | (Baitinger and Papenbrock 2016) | در این پژوهش به‌طور تجربی مطالعات را بررسی می کننده که چگونه امتیازهای معیار مرکزیت مختلف با یکدیگر و با ورودی‌های بهینه‌سازی پرتفوی ارتباط دارند. سپس چگونگی تلفیق مرکزیت در یک چارچوب مبتنی بر ریسک مانند چارچوب مبتنی بر بازده و ریسک را بیان می‌نمایند. |
| **الگوریتم‌های خوشه‌بندی برای ساخت پرتفوی ریسک تعدیل‌شده** | (León et al. 2017) | در این مقاله عملکرد هفت پرتفوی ایجادشده با استفاده از تکنیک‌های تجزیه‌وتحلیل خوشه‌بندی برای طبقه‌بندی دارایی‌ها به دسته‌ها و سپس استفاده از بهینه‌سازی کلاسیک در داخل هر خوشه برای انتخاب بهترین دارایی‌ها در داخل هر طبقه را بررسی می‌نماید. |
| **انتخاب دارایی مبتنی بر خوشه‌بندی سلسله مراتبی** | (Raffinot 2017) | این مقاله یک روش تخصیص دارایی مبتنی بر خوشه‌بندی سلسله مراتبی پیشنهاد می‌نماید که از تئوری شبکه و تکنیک‌های یادگیری |



www.SID.ir

Archive of SID



| | | |
|---|---|---|
| | | ماشین استفاده می‌نماید. نتایج تجربی آن‌ها نشان می‌دهد که پرتفوی مبتنی بر خوشه‌بندی سلسله مراتبی پایدار و واقعاً متنوع هستند و نسبت به تکنیک‌های معمول بهینه‌سازی ازنظر آماری عملکرد بهتری را برای تعدیل ریسک به دست می‌آورند. |
| **ریسک برابر سلسله مراتبی برای ساخت پرتفوی** | (Raffinot 2018) | نتایج نشان می‌دهد که پرتفوی HERC[1] مبتنی بر معیارهای ریسک نزولی، ازنظر آماری عملکرد بهتری نسبت به معیارهای CDaR برای تعدیل ریسک به دست می‌آورند. |
| **ساخت پرتفوی متنوع سازی شده که در مقایسه با نمونه عملکرد بهتری دارد** | (De Prado 2016) | در این مقاله، نویسنده رویکرد HRP[2] را برای جلوگیری از سه نگرانی عمده در بهینه‌سازی و CLA[3] به‌ویژه بی‌ثباتی، تمرکز و عملکرد پایین معرفی می‌نماید. |
| **آیا پرتفوی مبتنی بر روش‌های یادگیری ماشینی می‌تواند از پرتفوی مبتنی بر ریسک سنتی بهتر عمل کنند؟** | (Jain and Jain 2019) | در این مقاله تأثیر انتخاب غلط ماتریس کوواریانس را بر عملکرد روش‌های مختلف تخصیص بررسی نموده و سپس بررسی می‌کند که آیا عملکرد HRP مبتنی بر روش‌های یادگیری ماشین بهتر از پرتفوی مبتنی بر روش‌های تعدیل ریسک سنتی است یا نه؟ |
| **منبع: یافته‌های پژوهش** | | |

جدول (۳) کاربردهای خوشه‌بندی در پرتفوی

### ۴.۲. کاربردهای خوشه‌بندی و شبکه‌ها در مدیریت ریسک

در جدول زیر به بررسی پیشینه پژوهش‌های انجام‌شده در خصوص کاربرد شبکه‌ها در ریسک سیستمیک و مدیریت ریسک پرداخته‌شده است.

| عنوان | نویسنده (گان) | خلاصه |
|---|---|---|
| وابستگی متقابل شبکه‌های بازار اقتصادی و ریسک سیستمیک | (Harmon et al. 2010) | شبکه پویای روابط بین شرکت‌ها می‌تواند زمینه‌ساز شکست‌ها و بحران‌های اقتصادی شود.<br>بخش خدمات مالی با برقراری پیوند بین بخش‌های ضعیف اقتصادی، ریسک اقتصادی را افزایش می‌دهد و اگر خدمات مالی به‌نحوی‌که مانع رشد اقتصادی نشود محدود شوند می‌تواند ریسک سیستمیک را کاهش دهد. |
| ریسک سیستمیک در بازارهای | (Lautier and Raynaud | با استفاده از تئوری گراف می‌توان رفتار قیمت را بررسی کرد. |

[1] Hierarchical Equal Risk Contribution Portfolio (HERC)
[2] Hierarchical Risk Parity (HRP)
[3] critical line algorithm (CLA)



www.SID.ir

Archive of SID



| | | |
|---|---|---|
| مشتق انرژی: تجزیه‌وتحلیل تئوری گراف | 2012) | با استفاده از درخت پوشای کمینه می‌توان محتمل‌ترین و کوتاه‌ترین مسیر انتقال شوک‌های قیمتی را شناسایی کرد |
| استخراج شبکه بر اساس توضیح پویایی بین خوشه‌بندی و ارتباط بازارهای متقابل در منطقه آسیا: رویکرد درخت پوشای کمینه و خوشه‌بندی سلسله مراتبی | (Bhattacharjee, Shafi, and Acharjee 2019) | نتایج خوشه‌بندی سلسله مراتبی با استفاده از رویکرد k-means نشان داد که در طول بحران مالی و پس‌ازآن متوسط ارتفاع خوشه‌ها کاهش می‌یابد. همچنین گره‌های مرکزی نیز تغییر می‌کنند. MST تشکیل‌شده بر اساس بازارهای آسیایی در طول بحران مالی ۲۰۰۸ در مقایسه با قبل از آن دارای فواصل کوتاه‌تری است که بعد از پایان بحران به‌تدریج به حالت قبل بازمی‌گردد. |
| بررسی تکامل پیوند پویا در بازارهای سهام با استفاده از مدل‌ها و معیارهای شبکه: مطالعه‌ای در ادغام بازار سهام آسیا | (Bhattacharjee, Shafi, and Acharjee 2017) | مدل‌های شبکه‌های موزون همراه با معیارهای شبکه برای کشف ساختارهای پیوند متقابل بین بازارهای آسیایی به کار گرفته شدند. همچنین ویژگی‌های مکان‌شناسی یک شبکه با استفاده از معیارهای مرکزیت قابل توصیف هستند. نویسندگان این پژوهش ادعا می‌کنند که با استفاده از رویکردهای داده محور می‌توان در مدیریت ریسک سیستمیک مؤثر و برای انتخاب پرتفوی بهینه به‌خوبی متنوع شده که در مقابل شوک‌های سطح سیستم مقاوم باشند استفاده نمود. |
| ریسک سیستمیک و پویایی بازار مسکن ایالات‌متحده | (Meng et al. 2014) | مطالعه اهمیت بازارهای مسکن در اقتصاد و نقش مؤثر آن‌ها در رکود اقتصادی.<br>بررسی ریسک سیستمیک و پویایی بازار مسکن ایالات‌متحده با استفاده از تئوری ماتریس تصادفی.<br>مشاهده افزایش ریسک سیستمیک که ابزاری در جهت تشخیص زودهنگام حباب‌های مسکن است. |
| ریسک سیستمیک و رابطه علیت پویا در بازار بین‌المللی حمل‌ونقل | (Zhang et al. 2014) | بر اساس مطالعات مختلف، عاملی که باعث افزایش ریسک سیستمیک در اقتصاد می‌شود میزان همبستگی بسیار زیاد موجود بین بخش‌های مختلف بازارهای اقتصادی است. در این پژوهش ریسک سیستمیک در سه بخش بازار حمل‌ونقل موردبررسی قرارگرفته است. نتایج این پژوهش نشان می‌دهد شبکه‌های همبستگی طی سه دوره زمانی، قبل از بحران مالی، در طی بحران و بعدازآن، درخت پوشای کمینه و درخت سلسله مراتبی، هر دو دارای پیچیدگی پویایی هستند مثلاً بخش‌های مختلف بازار در طی بحران مالی دارای فواصل کمتری هستند. |
| اقدامات اقتصادسنجی از ارتباط و ریسک سیستمیک در بخش‌های مالی و بیمه | (Billio et al. 2012) | با بررسی معیارهای اقتصادسنجی بر اساس تحلیل اجزای اصلی و شبکه‌های مبتنی بر علیت گرانجر در ٤ بخش: بانک‌ها، کارگزاران، واسطه‌ها و شرکت‌های بیمه دریافتند که این بخش‌ها دارای همبستگی شدیدی با یکدیگر هستند و احتمالاً از طریق شبکه پیچیده و متغیر با زمان سطح ریسک سیستمیک را در صنایع مالی و بیمه افزایش می‌دهد. این معیارها همچنین می‌تواند دوره‌های بحران مالی را شناسایی و اندازه‌گیری کند. |



www.SID.ir

Archive of SID



| | | |
|---|---|---|
| شکنندگی بازار[1]، خطر سیستماتیک و انحنای ریچی[2] | (Sandhu, Georgiou, and Tannenbaum 2015) | اندازه‌گیری ریسک سیستمیک یا ضعف سیستم‌های مالی یکی از مسائل اساسی در تحلیل بهره‌وری بازار و تخصیص پرتفوی و همچنین پوشش ریسک مالی است. تلاش‌هایی صورت گرفته است که نشان می‌دهد چنین سیستم‌هایی می‌توانند به‌عنوان یک گراف موزون با مشخصه‌های شبکه پیچیده ارائه شوند. منحنی Ricci می‌تواند به‌عنوان شاخصی از بی‌ثباتی در شبکه‌های مالی مورداستفاده قرار گیرد. |
| مدیریت پویای ریسک‌پذیری و ریسک گریزی با رژیم‌های همبستگی و شبکه‌های همبستگی | (Papenbrock and Schwendner 2015) | در این پژوهش نظام همبستگی متمایزی بین سال‌های ۱۹۹۸ تا ۲۰۱۳ مشخص‌شده‌اند. نتیجه این پژوهش نشان می‌دهد که در طی بحران مالی ۲۰۰۸ نظام همبستگی نسبت به قبل به‌طور قابل‌توجهی تغییر کرده است. ردیابی خوشه‌ای نشان می‌دهد که کلاس‌های دارایی اکنون کمتر از یکدیگر تفکیک‌شده‌اند. شبکه‌های همبستگی نیز در شناسایی دارایی‌های "ریسکی" و "بدون ریسکی" در جهت مدیریت ریسک مالی، ساخت پرتفوی و تخصیص دارایی‌ها کمک‌کننده هستند. |
| شناسایی نظام بازارهای مالی | (Münnix et al. 2012) | بازارهای مالی را می‌توان در قالب یک سیستم پیچیده مورد تجزیه‌وتحلیل قراردادند. یافته‌های این پژوهش نشان داد که طیف گسترده‌ای از الگوهای ساختار همبستگی وجود دارد و همچنین شناسایی وضعیت بازار بر اساس همبستگی به‌منظور ایجاد "سیستم هشدار سریع" برای بازارهای مالی مفید است. این سیستم می‌تواند با مقایسه وضعیت فعلی باحالت‌های مشابه قبلی یا نظارت بر تغییرات سریع در ساختار همبستگی پیاده‌سازی شود. |
| متنوع سازی ریسک: مطالعه پایداری با رویکرد شبکه‌های همبستگی فیلتر شده | (Musmeci, Aste, and Di Matteo 2014) | در این پژوهش سیر زمانی تکامل ساختار همبستگی از بازده سهام، خوشه‌ها و ماندگاری آن‌ها را مطالعه موردمطالعه قرار گرفت. نتایج نشان می‌دهد که ساختار خوشه‌بندی در اوایل دهه ۲۰۰۰ کاملاً پایدار است و قبل از گسترش بحران ۲۰۰۷–۲۰۰۸ به پایداری کمتری می‌رسد. ساختار همبستگی درنهایت پس از بحران تداوم می‌یابد و ساختار جدید و متمایزی از ساختار قبل از بحران می‌سازد. |
| چشم‌انداز شبکه مالی در هم بخشی ریسک سیستمیک مؤسسات مالی | (Huang et al. 2016) | هدف از این پژوهش بررسی تأثیرات مؤسسات مالی در مکان‌شناسی ساختار شبکه‌های مالی بر سهم ریسک سیستمیک آن‌ها است. پس از مطالعه شبکه‌های مالی تشکیل‌شده از همبستگی شرطی پویا و طراحی گراف از طریق روش درخت پوشای کمینه دریافتند که مؤسسات مالی دارای قدرت گره بزرگ‌تر، مرکزیت گره بیشتر، معیار مرکزیت نزدیکی بیشتر و ضریب خوشه‌بندی بیشتر باشند، ریسک سیستمیک بیشتری را دارا هستند. |

[1] Market Fragility
[2] Ricci Curvature



www.SID.ir

Archive of SID



| | | |
|---|---|---|
| سیگنال‌های هشداردهنده زودهنگام سقوط، در مکان‌شناسی شبکه‌های بین‌بانکی | (Squartini, Van Lelyveld, and Garlaschelli 2013) | بحران مالی به‌وضوح نشان‌دهنده‌ی اهمیت توصیف سطح ریسک سیستمیک مرتبط با کل شبکه مالی بجای بررسی مؤسسات به‌صورت انفرادی است. |
| نظریه پیچیدگی و مقررات مالی | (Battiston et al. 2016) | نظریات اقتصادی سنتی قادر به توضیح سقوط سیستم‌های مالی و پیش‌بینی اثرات بلندمدت آن بر اقتصاد جهانی نمی‌باشد ازاین‌رو بعد از بحران مالی ۲۰۰۸ علاقه‌ی فزاینده‌ای نسبت به استفاده از نظریه شبکه‌های پیچیده و کاربرد آن در شبیه‌سازی بازارها به وجود آمده است. |
| آنتروپی ساختاری[1]: نظارت بر شبکه‌های مبتنی بر همبستگی در طول زمان با استفاده از بازارهای مالی | (Almog and Shmueli 2019) | مفهوم "تنوع ساختاری" یک شبکه به سطحی از عدم تشابه بین عوامل مختلفی که در سیستم عمل می‌کنند اشاره دارد و به‌طورمعمول به‌عنوان تعداد اجزای متصل در شبکه تفسیر می‌شود همچنین ساختار آنتروپی به‌سادگی احتمال کدگذاری اندازه متناسب خوشه‌ها در شبکه است. در این پژوهش پیشنهاد می‌شود که از آنتروپی ساختاری برای نظارت بر ساختار شبکه‌های مبتنی بر همبستگی در طول زمان استفاده شود. |
| تحلیل ریسک سیستمیک در اقتصادهای بازسازی‌شده و شبکه‌های مالی | (Cimini et al. 2015) | در این مقاله روش جدیدی موردبررسی قرار گرفت که این امکان را می‌دهد که یک شبکه موزون جهت‌دار را بازسازی کرده و مکان‌شناسی آن را فقط با استفاده از اطلاعات جزئی در مورد الگوهای اتصال آن و همچنین دو ویژگی مرتبط با هر گره، تخمین بزند. آزمایش‌ها بر روی شبکه‌های تجربی و نیز شبکه‌های ترکیبی تولیدشده از طریق مدل پیکربندی ناشی از شایستگی نشان می‌دهند که این روش برای غلبه بر فقدان اطلاعات مکان‌شناسی که اغلب مانع برآورد ریسک سیستماتیک در سیستم‌های اقتصادی می‌شود بسیار ارزشمند است. |
| خوشه‌بندی سری‌های زمانی مالی در شرایط ریسکی | (Durante, Pappadà, and Torelli 2014) | در این پژوهش روشی برای خوشه‌بندی سری‌های زمانی مالی با توجه به پیوستگی توزیع آن‌ها ارائه‌شده است این روش بر اساس محاسبه ضرایب همبستگی اسپیرمن استخراج‌شده است. نتایج این تحقیق می‌تواند برای ایجاد پرتفوی مالی با حداقل ریسک استفاده شود. |

[1] Structural Entropy



www.SID.ir

Archive of SID



| | | |
|---|---|---|
| ساختار وابستگی و ویژگی‌های پیمایشی سری‌های زمانی با یکدیگر مرتبط هستند. | (Morales, Di Matteo, and Aste 2014) | در این مقاله ساختار سلسله مراتبی از ریسک‌هایی که می‌تواند مسئول ساختار وابستگی چند متغیره سهام و رفتار تک متغیره به وجود آمده و سپس مدلی که مشاهدات تجربی را بازتولید کند ارائه می‌دهد (درهم تنیدگی چند مقیاس یک متغیره و خواص همبستگی چند متغیره سری‌های زمانی مالی). |
| ریسک سیستمیک در بازارهای مشتق انرژی: تحلیل تئوری گراف | (Lautier and Raynaud 2012) | با استفاده از تئوری گراف می‌توان رفتار قیمت را بررسی کرد.<br>با استفاده از درخت پوشای کمینه می‌توان محتمل‌ترین و کوتاه‌ترین مسیر انتقال شوک‌های قیمتی را شناسایی کرد |
| اثر متقابل چند مقیاسی و وابستگی سهام | (Buonocore et al. 2020) | نویسندگان این پژوهش یک وابستگی غیرخطی بین شاخصی از درجه چند سنجشی سری زمانی قیمت ورود سهام و میانگین همبستگی آن‌ها را کشف کردند. |
| طبقه‌بندی آماری صنایع | (Kakushadze and Yu 2016) | در این پژوهش الگوریتم‌هایی به‌منظور طبقه‌بندی صنعت آماری ارائه‌شده است و از صنایع می‌توان به‌عنوان عوامل ریسک در مدل‌های چندعاملی ریسک استفاده نمود. |
| شبکه‌های پرداخت سازمان‌ها و رتبه‌بندی ریسک اعتباری | (Letizia and Lillo 2019) | مشاهدات این پژوهش نشان می‌دهد که ریسک جامع در سیستم‌های پیچیده وابسته به دو ویژگی است: ریسک ویژه عناصر و مکان‌شناسی شبکه تعاملات بین آن‌ها. همچنین همبستگی قابل‌توجهی بین ویژگی‌های مکان‌شناسی هر گره (شرکت) و ریسک آن مشاهده شد. |
| تحلیل خوشه در بهینه‌سازی پرتفوی | (Tola et al. 2008) | در این پژوهش مسئله‌ی عدم اطمینان آماری ماتریس همبستگی را در بهینه‌سازی پرتفوی در نظر می‌گیرد. این پژوهش نشان می‌دهد که با فرض شرایط ایدئال با استفاده از الگوریتم‌های خوشه‌بندی می‌توان قابلیت اطمینان پرتفوی‌ها را ازنظر نسبت ریسک پیش‌بینی‌شده و ریسک تحقق‌یافته بهبود بخشید. |
| انقباض و فیلتر طیفی[1] ماتریس‌های همبستگی: مقایسه‌ای از طریق فاصله Kullback-Leibler. | (Tummino, Lillo, and Mantegna 2007) | در این پژوهش از فاصله‌ی Kullback-Leibler به‌منظور اندازه‌گیری عملکرد الگوریتم‌های فیلتر در بازیابی ماتریس همبستگی که متغیرها توسط یک توزیع گاوسی چندمتغیره توصیف می‌شوند استفاده شد. |

جدول (٤) کاربردهای خوشه‌بندی و شبکه‌ها در مدیریت ریسک

[1] Shrinkage and spectral filtering



www.SID.ir

Archive of SID


## ۵. نتیجه‌گیری

زمانی‌که هیچ دانش اولیه‌ای در مورد طبقه‌بندی داده‌های بزرگ وجود ندارد، خوشه‌بندی راه‌حلی برای این موضوع است. خوشه‌بندی در حوزه‌های مختلف علوم به‌منظور کشف الگوها و تحلیل داده‌ها استفاده می‌شود شامل الگوریتم‌ها و روش‌های متفاوتی ازجمله: روش کاهش ابعاد سری زمانی مانند روش‌های DFT,DWT، الگوریتم‌های خوشه‌بندی مانند خوشه‌بندی سلسله مراتبی و بخش‌بندی، روش‌های اندازه‌گیری میزان شباهت بین سری‌های زمانی مانند فاصله اقلیدسی و LCSS و ... و همچنین روش‌های مبتنی بر نمونه شامل معیارهای مرکزیت و میانگین‌ها می‌باشد.

اگرچه پژوهش‌های مختلفی با موضوع خوشه‌بندی سری‌های زمانی انجام‌شده‌اند اما ویژگی‌های منحصربه‌فرد داده‌های زمانی موانعی هستند که در اکثر الگوریتم‌های خوشه‌بندی مرسوم برای ارائه نتیجه مناسب، خوشه‌بندی سری‌های زمانی را با شکست مواجه می‌کند. به‌طور خاص، ابعاد بالا، همبستگی زیاد و به‌طورمعمول مقدار زیادی نویز که داده‌های سری زمانی را توصیف می‌کنند، به‌عنوان یک چالش تحقیقاتی، در خوشه‌بندی سری‌های زمانی در نظر گرفته‌شده‌اند.

در این پژوهش با استفاده از روش نگاشت علم به بررسی کاربردهای خوشه‌بندی در زمینه‌های مختلف پرداخته‌شده است. در بخش دوم، مراحل پیاده‌سازی نگاشت علم برای کاربرد خوشه‌بندی در حوزه مالی توضیح داده شد و نتایج پیاده‌سازی این روش که با استفاده از پکیج bibliometrix استخراج‌شده بود، ارائه گردید. روش نگاشت علم به‌تدریج در همه‌ی حوزه‌های علمی گسترش‌یافته و با تکیه‌بر معرفی یک فرایند، مروری کلی از پژوهش‌های انجام‌شده ارائه می‌دهد. این مرور که به‌صورت نظام‌مند، شفاف و تکرارپذیر است بر اساس اندازه‌گیری آماری علم، متخصصان و فعالیت‌های علمی را معرفی می‌کند(آریا و کوکورولو[1]،۲۰۱۷). در بخش سوم به دسته‌بندی انواع روش‌های خوشه‌بندی، الگوریتم‌های خوشه‌بندی پرداخته و در بخش چهارم مقاله ضمن ارائه کاربردهای خوشه‌بندی در حوزه‌های مختلف، به کاربردهای آن در حوزه مالی پرداخته و بیان شد که این مقوله در بررسی انواع کاربردهای خوشه‌بندی و شبکه‌سازی در حوزه‌های مختلف مالی ازجمله ریسک، معاملات الگوریتمی، بانکداری و دیگر موضوعات پرکاربرد در حوزه مالی قابل‌استفاده است. ازآنجایی‌که این پژوهش دربرگیرنده‌ی اطاعات بخش قابل‌توجهی از پژوهشگران و پژوهش‌هایی که در حوزه خوشه‌بندی داده‌ها

[1] Aria& Cuccurullo,2017



www.SID.ir

Archive of SID



به‌خصوص داده‌های مالی قابل‌دسترس می‌باشد به پژوهشگران در جهت تصمیم‌گیری و انتخاب روش‌ها و الگوریتم‌های موجود با توجه به پژوهش موردنظر خود کمک می‌کند.

## منابع


Aghabozorgi, Saeed, Ali Seyed Shirkhorshidi, and Teh Ying Wah. 2015. "Time-Series Clustering–a Decade Review." *Information Systems* 53:16–38.

Almog, Assaf, and Erez Shmueli. 2019. "Structural Entropy: Monitoring Correlation-Based Networks over Time with Application to Financial Markets." *Scientific Reports* 9(1):1–13.

Aria, Massimo, and Corrado Cuccurullo. 2017. "Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis." *Journal of Informetrics* 11(4):959–75.

Aßfalg, Johannes, Hans-Peter Kriegel, Peer Kröger, Peter Kunath, Alexey Pryakhin, and Matthias Renz. 2006. "Similarity Search on Time Series Based on Threshold Queries." Pp. 276–94 in *International Conference on Extending Database Technology*. Springer.

Baitinger, Eduard, and Jochen Papenbrock. 2016. "Interconnectedness Risk and Active Portfolio Management." *Journal of Investment Strategies, Forthcoming*.

Banerjee, Arindam, and Joydeep Ghosh. 2001. "Clickstream Clustering Using Weighted Longest Common Subsequences." P. 144 in *Proceedings of the web mining workshop at the 1st SIAM conference on data mining*. Vol. 143.

Battiston, Stefano, J. Doyne Farmer, Andreas Flache, Diego Garlaschelli, Andrew G. Haldane, Hans Heesterbeek, Cars Hommes, Carlo Jaeger, Robert May, and Marten Scheffer. 2016. "Complexity Theory and Financial Regulation." *Science* 351(6275):818–19.

Bhattacharjee, Biplab, Muhammad Shafi, and Animesh Acharjee. 2017. "Investigating the Evolution of Linkage Dynamics among Equity Markets Using Network Models and Measures: The Case of Asian Equity Market Integration." *Data* 2(4):41.

Bhattacharjee, Biplab, Muhammad Shafi, and Animesh Acharjee. 2019. "Network Mining Based Elucidation of the Dynamics of Cross-Market Clustering and Connectedness in Asian Region: An MST and Hierarchical Clustering Approach." *Journal of King Saud University-Computer and Information Sciences* 31(2):218–28.

Billio, Monica, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon. 2012. "Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors." *Journal of Financial Economics* 104(3):535–59.

Börner, Katy, Chaomei Chen, and Kevin W. Boyack. 2003. "Visualizing Knowledge Domains." *Annual Review of Information Science and Technology* 37(1):179–255.




www.SID.ir

Archive of SID


Briner, Rob B., and David Denyer. 2012. "Systematic Review and Evidence Synthesis as a Practice and Scholarship Tool." *Handbook of Evidence-Based Management: Companies, Classrooms and Research* 112–29.

Broadus, Robert N. 1987. "Toward a Definition of 'Bibliometrics.'" *Scientometrics* 12(5–6):373–79.

Buonocore, R. J., G. Brandi, R. N. Mantegna, and T. Di Matteo. 2020. "On the Interplay between Multiscaling and Stock Dependence." *Quantitative Finance* 20(1):133–45.

Chandra, B., Manish Gupta, and M. P. Gupta. 2008. "A Multivariate Time Series Clustering Approach for Crime Trends Prediction." Pp. 892–96 in *2008 IEEE International Conference on Systems, Man and Cybernetics*. IEEE.

Cheeseman, Peter C., and John C. Stutz. 1996. "Bayesian Classification (AutoClass): Theory and Results." *Advances in Knowledge Discovery and Data Mining* 180:153–80.

Chen, Lei, and Raymond Ng. 2004. "On the Marriage of Lp-Norms and Edit Distance." Pp. 792–803 in *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*.

Chen, Lei, and M. Tamer Özsu. 2005. "Using Multi-Scale Histograms to Answer Pattern Existence and Shape Match Queries." in *In SSDBM*. Citeseer.

Chen, Lei, M. Tamer Özsu, and Vincent Oria. 2005. "Robust and Fast Similarity Search for Moving Object Trajectories." Pp. 491–502 in *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*.

Chen, Yueguo, Mario A. Nascimento, Beng Chin Ooi, and Anthony K. H. Tung. 2007. "Spade: On Shape-Based Pattern Detection in Streaming Time Series." Pp. 786–95 in *2007 IEEE 23rd International Conference on Data Engineering*. IEEE.

Chiu, Bill, Eamonn Keogh, and Stefano Lonardi. 2003. "Probabilistic Discovery of Time Series Motifs." Pp. 493–98 in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*.

Cimini, Giulio, Tiziano Squartini, Diego Garlaschelli, and Andrea Gabrielli. 2015. "Systemic Risk Analysis on Reconstructed Economic and Financial Networks." *Scientific Reports* 5:15758.

Cobo, Manuel J., Antonio Gabriel López□Herrera, Enrique Herrera□Viedma, and Francisco Herrera. 2011. "Science Mapping Software Tools: Review, Analysis, and Cooperative Study among Tools." *Journal of the American Society for Information Science and Technology* 62(7):1382–1402.

Crane, Diana. 1972. "Invisible Colleges; Diffusion of Knowledge in Scientific Communities."

Diodato, Virgil P., and Peter Gellatly. 2013. *Dictionary of Bibliometrics*. Routledge.



www.SID.ir

Archive of SID


Dose, Christian, and Silvano Cincotti. 2005. "Clustering of Financial Time Series with Application to Index and Enhanced Index Tracking Portfolio." *Physica A: Statistical Mechanics and Its Applications* 355(1):145–51.

Durante, Fabrizio, Roberta Pappadà, and Nicola Torelli. 2014. "Clustering of Financial Time Series in Risky Scenarios." *Advances in Data Analysis and Classification* 8(4):359–76.

Ezugwu, Absalom E., Amit K. Shukla, Moyinoluwa B. Agbaje, Olaide N. Oyelade, Adán José-García, and Jeffery O. Agushaka. 2020. "Automatic Clustering Algorithms: A Systematic Review and Bibliometric Analysis of Relevant Literature." *Neural Computing and Applications* 1–60.

Faloutsos, Christos, Mudumbai Ranganathan, and Yannis Manolopoulos. 1994. "Fast Subsequence Matching in Time-Series Databases." *Acm Sigmod Record* 23(2):419–29.

Fraley, Chris, and Adrian E. Raftery. 1998. "How Many Clusters? Which Clustering Method? Answers via Model-Based Cluster Analysis." *The Computer Journal* 41(8):578–88.

Frentzos, Elias, Kostas Gratsias, and Yannis Theodoridis. 2007. "Index-Based Most Similar Trajectory Search." Pp. 816–25 in *2007 IEEE 23rd International Conference on Data Engineering*. IEEE.

Fu, Tak-chung, F. L. Chung, Vincent Ng, and Robert Luk. 2001. "Pattern Discovery from Stock Time Series Using Self-Organizing Maps." Pp. 26–29 in *Workshop Notes of KDD2001 Workshop on Temporal Data Mining*. Citeseer.

Golay, Xavier, Spyros Kollias, Gautier Stoll, Dieter Meier, Anton Valavanis, and Peter Boesiger. 1998. "A New Correlation□based Fuzzy Logic Clustering Algorithm for FMRI." *Magnetic Resonance in Medicine* 40(2):249–60.

Graves, Daniel, and Witold Pedrycz. 2010. "Proximity Fuzzy Clustering and Its Application to Time Series Clustering and Prediction." Pp. 49–54 in *2010 10th International Conference on Intelligent Systems Design and Applications*. IEEE.

Harmon, Dion, Blake Stacey, Yavni Bar-Yam, and Yaneer Bar-Yam. 2010. "Networks of Economic Market Interdependence and Systemic Risk." *ArXiv Preprint ArXiv:1011.3707*.

Hautamaki, Ville, Pekka Nykanen, and Pasi Franti. 2008. "Time-Series Clustering by Approximate Prototypes." Pp. 1–4 in *2008 19th International Conference on Pattern Recognition*. IEEE.

He, Wenping, Guolin Feng, Qiong Wu, Tao He, Shiquan Wan, and Jifan Chou. 2012. "A New Method for Abrupt Dynamic Change Detection of Correlated Time Series." *International Journal of Climatology* 32(10):1604–14.



www.SID.ir

Archive of SID


Huang, Wei-Qiang, Xin-Tian Zhuang, Shuang Yao, and Stan Uryasev. 2016. "A Financial Network Perspective of Financial Institutions' Systemic Risk Contributions." *Physica A: Statistical Mechanics and Its Applications* 456:183–96.

Hüttner, Amelie, Jan-Frederik Mai, and Stefano Mineo. 2018. "Portfolio Selection Based on Graphs: Does It Align with Markowitz-Optimal Portfolios?" *Dependence Modeling* 6(1):63–87.

Indyk, Piotr, Nick Koudas, and Shanmugavelayutham Muthukrishnan. 2000. "Identifying Representative Trends in Massive Time Series Data Sets Using Sketches." Pp. 363–72 in *26th International Conference on Very Large Data Bases, VLDB 2000*.

Jain, Anil K., M. Narasimha Murty, and Patrick J. Flynn. 1999. "Data Clustering: A Review." *ACM Computing Surveys (CSUR)* 31(3):264–323.

Jain, Prayut, and Shashi Jain. 2019. "Can Machine Learning-Based Portfolios Outperform Traditional Risk-Based Portfolios? The Need to Account for Covariance Misspecification." *Risks* 7(3):74.

Kakushadze, Zura, and Willie Yu. 2016. "Statistical Industry Classification." *Journal of Risk & Control* 3(1):17–65.

Kalpakis, Konstantinos, Dhiral Gada, and Vasundhara Puttagunta. 2001. "Distance Measures for Effective Clustering of ARIMA Time-Series." Pp. 273–80 in *Proceedings 2001 IEEE international conference on data mining*. IEEE.

Keogh, Eamonn, Stefano Lonardi, and Bill'Yuan-chi' Chiu. 2002. "Finding Surprising Patterns in a Time Series Database in Linear Time and Space." Pp. 550–56 in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*.

Keogh, Eamonn, Stefano Lonardi, Chotirat Ann Ratanamahatana, Li Wei, Sang-Hee Lee, and John Handley. 2007. "Compression-Based Data Mining of Sequential Data." *Data Mining and Knowledge Discovery* 14(1):99–129.

Kohonen, Teuvo. 1990. "The Self-Organizing Map." *Proceedings of the IEEE* 78(9):1464–80.

Kumar, Mahesh, Nitin R. Patel, and Jonathan Woo. 2002. "Clustering Seasonality Patterns in the Presence of Errors." Pp. 557–63 in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*.

Lang, Willis, Michael Morse, and Jignesh M. Patel. 2009. "Dictionary-Based Compression for Long Time-Series Similarity." *IEEE Transactions on Knowledge and Data Engineering* 22(11):1609–22.

Latecki, Longin Jan, Vasilis Megalooikonomou, Qiang Wang, Rolf Lakaemper, Chotirat Ann Ratanamahatana, and Eamonn Keogh. 2005. "Elastic Partial Matching of Time



www.SID.ir

Archive of SID


Series." Pp. 577–84 in *European Conference on Principles of Data Mining and Knowledge Discovery*. Springer.

Lautier, Delphine, and Franck Raynaud. 2012. "Systemic Risk in Energy Derivative Markets: A Graph-Theory Analysis." *The Energy Journal* 33(3).

Lee Rodgers, Joseph, and W. Alan Nicewander. 1988. "Thirteen Ways to Look at the Correlation Coefficient." *The American Statistician* 42(1):59–66.

Leng, Mingwei, Xinsheng Lai, Guolv Tan, and Xiaohui Xu. 2009. "Time Series Representation for Anomaly Detection." Pp. 628–32 in *2009 2nd IEEE International Conference on Computer Science and Information Technology*. IEEE.

León, Diego, Arbey Aragón, Javier Sandoval, Germán Jairo Hernández, Andrés Arévalo, and Jaime Niño. 2017. "Clustering Algorithms for Risk-Adjusted Portfolio Construction." Pp. 1334–43 in *ICCS*.

Letizia, Elisa, and Fabrizio Lillo. 2019. "Corporate Payments Networks and Credit Risk Rating." *EPJ Data Science* 8(1):21.

Lohre, Harald, Carsten Rother, and Kilian Axel Schäfer. 2020. "Hierarchical Risk Parity: Accounting for Tail Dependencies in Multi-Asset Multi-Factor Allocations." *Machine Learning and Asset Management, Forthcoming*.

MacQueen, James. 1967. "Some Methods for Classification and Analysis of Multivariate Observations." Pp. 281–97 in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. Oakland, CA, USA.

Meng, Hao, Wen-Jie Xie, Zhi-Qiang Jiang, Boris Podobnik, Wei-Xing Zhou, and H. Eugene Stanley. 2014. "Systemic Risk and Spatiotemporal Dynamics of the US Housing Market." *Scientific Reports* 4(1):1–7.

Möller-Levet, Carla S., Frank Klawonn, Kwang-Hyun Cho, and Olaf Wolkenhauer. 2003. "Fuzzy Clustering of Short Time-Series and Unevenly Distributed Sampling Points." Pp. 330–40 in *International symposium on intelligent data analysis*. Springer.

Morales, Raffaello, T. Di Matteo, and Tomaso Aste. 2014. "Dependency Structure and Scaling Properties of Financial Time Series Are Related." *Scientific Reports* 4(1):1–9.

Morse, Michael D., and Jignesh M. Patel. 2007. "An Efficient and Accurate Method for Evaluating Time Series Similarity." Pp. 569–80 in *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*.

Münnix, Michael C., Takashi Shimada, Rudi Schäfer, Francois Leyvraz, Thomas H. Seligman, Thomas Guhr, and H. Eugene Stanley. 2012. "Identifying States of a Financial Market." *Scientific Reports* 2:644.

Musmeci, Nicoló, Tomaso Aste, and Tiziana Di Matteo. 2014. "Risk Diversification: A Study of Persistence with a Filtered Correlation-Network Approach." *ArXiv Preprint*



www.SID.ir

Archive of SID


*ArXiv:1410.5621*.

Onnela, J. P., Anirban Chakraborti, Kimmo Kaski, Janos Kertesz, and Antti Kanto. 2003. "Dynamics of Market Correlations: Taxonomy and Portfolio Analysis." *Physical Review E* 68(5):56110.

Papenbrock, Jochen, and Peter Schwendner. 2015. "Handling Risk-on/Risk-off Dynamics with Correlation Regimes and Correlation Networks." *Financial Markets and Portfolio Management* 29(2):125–47.

Peralta, Gustavo, and Abalfazl Zareei. 2016. "A Network Approach to Portfolio Selection." *Journal of Empirical Finance* 38:157–80.

Polz, E. P. Patrick, Erik Hortnagl, and E. Prem. 2003. "Processing and Clustering Time Series of Mobile Robot Sensory Data." *Austrian Research Institute for Artificial Intelligence: Systemic Intelligence for GrowiNgup Artefacts That Live-SIGNAL*.

de Prado, Marcos Lopez. 2020. *Machine Learning for Asset Managers*. Cambridge University Press.

De Prado, Marcos Lopez. 2016. "Building Diversified Portfolios That Outperform out of Sample." *The Journal of Portfolio Management* 42(4):59–69.

Pritchard, Alan. 1969. "Statistical Bibliography or Bibliometrics." *Journal of Documentation* 25(4):348–49.

Raffinot, Thomas. 2017. "Hierarchical Clustering-Based Asset Allocation." *The Journal of Portfolio Management* 44(2):89–99.

Raffinot, Thomas. 2018. "The Hierarchical Equal Risk Contribution Portfolio." *Available at SSRN 3237540*.

Rai, Pradeep, and Shubha Singh. 2010. "A Survey of Clustering Techniques." *International Journal of Computer Applications* 7(12):1–5.

Ren, Fei, Ya-Nan Lu, Sai-Ping Li, Xiong-Fei Jiang, Li-Xin Zhong, and Tian Qiu. 2017. "Dynamic Portfolio Strategy Using Clustering Approach." *PloS One* 12(1):e0169299.

Rousseau, Denise M. 2012. *The Oxford Handbook of Evidence-Based Management*. Oxford University Press.

Sakoe, Hiroaki. 1971. "Dynamic-Programming Approach to Continuous Speech Recognition." in *1971 Proc. the International Congress of Acoustics, Budapest*.

Sandhu, Romeil, Tryphon Georgiou, and Allen Tannenbaum. 2015. "Market Fragility, Systemic Risk, and Ricci Curvature." *ArXiv Preprint ArXiv:1505.05182*.

Saxena, Amit, Mukesh Prasad, Akshansh Gupta, Neha Bharill, Om Prakash Patel, Aruna Tiwari, Meng Joo Er, Weiping Ding, and Chin-Teng Lin. 2017. "A Review of Clustering Techniques and Developments." *Neurocomputing* 267:664–81.



www.SID.ir

Archive of SID


Sfetsos, Athanasios, and Costas Siriopoulos. 2004. "Time Series Forecasting with a Hybrid Clustering Scheme and Pattern Recognition." *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 34(3):399–405.

Smyth, Padhraic. 1997. "Clustering Sequences with Hidden Markov Models." Pp. 648–54 in *Advances in neural information processing systems*.

Squartini, Tiziano, Iman Van Lelyveld, and Diego Garlaschelli. 2013. "Early-Warning Signals of Topological Collapse in Interbank Networks." *Scientific Reports* 3:3357.

Theodoridis, Sergios, and Rama Chellappa. 2013. *Academic Press Library in Signal Processing: Signal Processing Theory and Machine Learning*. Academic Press.

Tola, Vincenzo, Fabrizio Lillo, Mauro Gallegati, and Rosario N. Mantegna. 2008. "Cluster Analysis for Portfolio Optimization." *Journal of Economic Dynamics and Control* 32(1):235–58.

Tumminello, Michele, Fabrizio Lillo, and Rosario Nunzio Mantegna. 2007. "Shrinkage and Spectral Filtering of Correlation Matrices: A Comparison via the Kullback-Leibler Distance." *ArXiv Preprint ArXiv:0710.0576*.

Vlachos, Michail, George Kollios, and Dimitrios Gunopulos. 2002. "Discovering Similar Multidimensional Trajectories." Pp. 673–84 in *Proceedings 18th international conference on data engineering*. IEEE.

Waltman, Ludo. 2016. "A Review of the Literature on Citation Impact Indicators." *Journal of Informetrics* 10(2):365–91.

Wang, Haixun, Wei Wang, Jiong Yang, and Philip S. Yu. 2002. "Clustering by Pattern Similarity in Large Data Sets." Pp. 394–405 in *Proceedings of the 2002 ACM SIGMOD international conference on Management of data*.

Zhang, Xin, Boris Podobnik, Dror Y. Kenett, and H. Eugene Stanley. 2014. "Systemic Risk and Causality Dynamics of the World International Shipping Market." *Physica A: Statistical Mechanics and Its Applications* 415:43–53.

Zupic, Ivan, and Tomaž Čater. 2015. "Bibliometric Methods in Management and Organization." *Organizational Research Methods* 18(3):429–72.



www.SID.ir

Archive of SID


# A Review of Research on Financial Time Series Clustering: A Bibliometrics Approach

marzeyeh norahmadi, fatemeh rasti, hojatollah sadeghi

**Abstract**

The amount of information and data we retrieve and use is growing rapidly. Data mining is the process of extracting relevant data from large volumes of data and the method of discovering and finding the appropriate pattern from large volumes of data sets.Clustering is one of the most common methods of statistical data analysis, and also one of the best data mining approaches. This approach, as a method of unsupervised learning, uses algorithms to classify time series data according to different criteria. The purpose of this study is to investigate the types of applications of clustering and networking in various financial fields, including risk, algorithmic trading, banking and other widely used topics in this field. In this research, using the bibliometrix package in the software, all the researches on clustering is reviewed. While extracting various criteria and clustering approaches, its applications have been studied. This study with a comprehensive review of all research in this field can help researchers as a toolbox to provide a variety of clustering methods in ideation and selection of appropriate methods in classifying and analyzing financial data.

Keywords: clustering, financial time series, financial networks, distance metrics, Bibliometrics Approach.

JEL Classification: G0,G11, G15



www.SID.ir