7th International Conference on Web Research
May 19-20, 2021; Tehran, IRAN ICWR2021

Using Cluster-led Machine Learning in Textual Analysis of Articles Indexed on the Web

(Case: Articles on Climate Resilience in the Field of Urban Planning)

Mahdi Suleimany

Undergraduate student in Urban Planning, Art University of Isfahan (AUI)
mi.suleimany@gmail.com

Abstract

The increasing growth of scientific studies indexed on the web and also the rise of articles in different fields of science have made it impossible to thoroughly study the material in traditional ways, as well as to investigate all relevant sources in terms of time, cost, and available resources. Hence, researchers have shifted to machine learning methods and smart analysis of literature using mathematical algorithms. One of the newest and most common types of these algorithms is cluster-led machine learning, which provides bibliographical and textual analysis of scientific papers in the form of interfaces and software such as VOS Viewer. The present article has been compiled as a promotional review study by the method of library and documentary research to explain this algorithm and the way of using the relevant software in the textual analysis of the articles. In this study, first, the scientific literature review methods are explained, and then in three steps, the use of VOS Viewer software and its algorithm on a case study have been investigated. The case studied in this research is articles related to climate resilience in the field of urban planning that have been indexed on the web from 2001 to 2021.

Keywords: Cluster-led machine learning, Bibliographical analysis, Textual analysis, Scientific articles, VOS Viewer

7th International Conference on Web Research
May 19-20, 2021; Tehran, IRAN ICWR2021

کاربست یادگیری ماشینی خوشه‌مبنا در تحلیل متنی مقالات علمی نمایه شده در وب
(نمونه پژوهی: مقالات مربوط به تاب‌آوری اقلیمی در حوزه برنامه‌ریزی شهری)

مهدی سلیمانی

دانشجوی مقطع کارشناسی مهندسی شهرسازی، دانشگاه هنر اصفهان
mi.suleimany@gmail.com

چکیده

رشد فزاینده مطالعات علمی نمایه شده در وب و افزایش چشم‌گیر شمار مقالات در حوزه‌های تخصصی علوم، بررسی کامل مطالب به روش‌های سنتی و همچنین مطالعه‌ی تمام ادبیات مرتبط را به لحاظ زمان، هزینه و منابع در دسترس غیرممکن ساخته است. به همین دلیل پژوهشگران به کاربست شیوه‌های یادگیری ماشینی و هوشمندسازی روند تحلیل مقالات علمی به وسیله‌ی الگوریتم‌های ریاضی روی آوردند. یکی از جدیدترین و متداول‌ترین این الگوریتم‌ها، یادگیری ماشینی خوشه‌مبناست که در قالب رابط و نرم‌افزارهایی چون VOS Viewer به تحلیل مرجعی و متنی اسناد می‌پردازد. مقاله حاضر به عنوان یک پژوهش علمی-ترویجی و به روش مطالعات کتابخانه‌ای، با هدف تشریح این الگوریتم و نحوه استفاده از نرم‌افزارهای مربوطه در تحلیل متنی مقالات علمی، تدوین گشته است. به همین جهت، در ابتدا به بررسی شیوه‌های مرور ادبیات علمی پرداخت شده و سپس در سه گام نحوه استفاده از نرم‌افزار VOS Viewer و الگوریتم آن بر روی یک نمونه موردی، بررسی می‌گردد. نمونه مورد مطالعه در این پژوهش، مقالات مربوط به تاب‌آوری اقلیمی در حوزه برنامه‌ریزی شهری است که در بازه ۲۰۰۱ تا ۲۰۲۱ در اینترنت نمایه شده‌اند.

کلمات کلیدی

یادگیری ماشینی خوشه‌مبنا، تحلیل مرجعی، تحلیل متنی، مقالات علمی، نرم‌افزار VOS Viewer

۱- مقدمه

با توسعه دهکده جهانی اینترنت و به موازات آن افزایش ضرورت برای انجام پژوهش در حوزه‌های علمی مختلف، سالانه چند صدهزار مقاله یا سند علمی در وب، توسط انتشارات و پایگاه‌های علمی گوناگون نمایه می‌گردد. همچنین در بسیاری از حوزه‌های تخصصی، مجموع مقالات نمایه شده به بیش از ده‌ها هزار مقاله می‌رسد [1]. پایگاه اطلاعاتی اسکوپوس (Scopus) در سال ۲۰۲۰، به تنهایی بیش از ۲۴۰۰۰ مقاله با عنوان‌های متفاوت را نمایه ساخته است. در پلتفورم پایگاه اطلاعاتی Web of Science، که یکی دیگر از پایگاه‌های علمی بزرگ جهان است، بیش از ۱۸۰۰۰ مقاله در همین سال نمایه گردید. البته این آمار بدون نظرگرفتن کتب و اختراعات ثبت شده در این پایگاه‌هاست، که با افزودن آن‌ها این اعداد به ۲۰۰۰۰۰ تا ۳۰۰۰۰۰ سند علمی خواهد رسید [2 - 3].

شمار بالای پژوهش‌های نمایه شده در وب و نبود امکان بررسی همه آن‌ها به لحاظ زمان، هزینه و منابع در دسترس، پژوهشگران را بر آن داشت تا به کاربست شیوه‌های یادگیری ماشینی و هوشمندسازی روند تحلیل مقالات علمی با استفاده از الگوریتم‌های ریاضی روی آورند. در این راستا الگوریتم‌های متعددی توسعه یافت که یکی از متداول‌ترین آن‌ها، یادگیری ماشینی خوشه‌مبناست. یادگیری ماشینی خوشه‌مبنا، شیوه‌ای از کاربست یادگیری ماشینی در تحلیل متنی مقالات علمی‌ست که بر مبنای خوشه‌بندی مطالعات و موضوع آن‌ها بر اساس کلمات کلیدی و میزان تکرارشان در اسناد، توسعه یافته است. استفاده از این شیوه ابتدا در حوزه مطالعات علم‌سنجی رواج یافت و سپس پژوهشگران سایر رشته‌های علمی نیز به کاربست آن روی آوردند.

همچنین جهت بهینه‌سازی رابط کاربری، نرم‌افزارهای متعددی در این زمینه طراحی شده که می‌توان به دو نمونه VOS Viewer (توسعه یافته توسط دانشگاه لیدن) و CitNet Explorer اشاره نمود [4 - 5].

۲- روش‌شناسی پژوهش

پژوهش حاضر یک مطالعه علمی-ترویجی بوده و به روش مطالعات کتابخانه‌ای، با هدف تشریح شیوه کاربست یادگیری ماشینی خوشه‌مبنا در تحلیل متنی مقالات علمی، تدوین شده است. در این پژوهش ابتدا به شرح شیوه‌های مرور ادبیات علمی پرداخت شده و سپس در سه گام نحوه استفاده از نرم‌افزار VOS Viewer و الگوریتم به کار رفته در آن بر روی یک نمونه موردی، بررسی می‌گردد. نمونه مورد مطالعه در این پژوهش، مقالات مربوط به تاب‌آوری اقلیمی در حوزه برنامه‌ریزی شهری است که در بازه زمانی ۲۰۰۱ تا ۲۰۲۱ در اینترنت نمایه شده‌اند. همچنین دلیل استفاده از نرم‌افزار VOS Viewer آن است که این نرم‌افزار عمومیت بیشتری نسبت به سایر نرم‌افزارهای مشابه دارد. برای مثال در نرم‌افزار CitNet Explorer تنها می‌توان مقالات نمایه شده در پایگاه Web of Science را تحلیل نمود.

۳- نرم‌افزار VOS Viewer و شیوه کاربست آن

نرم‌افزار VOS Viewer یک نرم‌افزار مبتنی بر کدنویسی جاوا (Java) است. این نرم‌افزار در سال ۲۰۱۰، توسط محققین دانشگاه لیدن توسعه یافت. از این نرم‌افزار برای ساخت و نمایش بصری داده‌های متنی و مرجع‌شناختی استفاده می‌شود. گام‌های استفاده از این نرم‌افزار در ادامه توضیح داده شده است [4]؛

۳-۱- گام صفر: تعیین سبک مرور مطالعات

مرور مطالعات علمی و ادبیات نظری مرتبط در هر پژوهش، بخش مهم و جدانشدنی از آن است. مطالعات علمی پیشین در واقع پایه‌های علمی مطالعات فعلی هستند که به پژوهشگران در بررسی دقیق‌تر و نتیجه‌گیری واضح‌تر کمک می‌نماید. به طور کلی سه روش برای مرور ادبیات علمی وجود دارد؛

- مرور نظام‌مند (Systematic Review):
یک شکل ساختاردار مرور منابع است که روی یک یا چند پرسش مشخص متمرکز بوده و سعی در ارائه پاسخ با استفاده از تجزیه و تحلیل همه شواهد معتبر موجود دارد. این نوع مرور با استفاده از جستجوی منابع، استفاده از معیارهای ورود و خروج از پیش تعیین شده، ارزیابی نقادانه شواهد، استخراج و تولید داده‌ها از شواهد و تولید یافته از آن‌ها انجام می‌شود [6].
- مرور بهترین شواهد (Best Evidence Synthesis):
یک روش حد وسط و جایگزین برای دو نوع دیگر مرور منابع علمی‌ست. این شیوه ترکیبی از روش‌های نظام‌مند کمّی با تاکید بر مطالعات منفرد و مباحث ماهوی از نقطه نظر روایتی، و با تمرکز بر روی بهترین شواهد در یک حیطه است. در این روش برای جلوگیری از تورش مرورگر، بر اهمیت وجود معیارهای ورود قیاسی شفاف و پایدار به جای معیارهای متداول فراتحلیل تاکید شده است. ضریب تاثیر (Impact Factor) در این شیوه، برخلاف مرور نظام‌مند یک معیار مهم در انتخاب منابع است [۱].
- مرور روایتی (Narrative Review):
مرور روایتی برای موضوعات جامع و مفصل کاربرد دارد. این روش، مطالعات اولیه و اصیل یک حیطه را خلاصه می‌کند. از نقاط قوت این شیوه دربر داشتن تکثر و تفاوت‌ها در عناوین پژوهش‌های آکادمیک و ایجاد فرصت برای ارائه دانش و تجربیات فرد مرورگر است. در مرور روایتی، بر خلاف انواع دیگر روش‌های مرور، وقتی یک فرضیه واحد وجود ندارد یا با روش‌های نظام‌مند قابل جمع‌بندی نیست، فرد مرورگر با استفاده از تجربیات خود، و مدل‌ها و فرضیه‌های موجود، یک نتیجه‌گیری کلی در حیطه مورد نظر ارائه می‌کند [6 - 7].

نحوه مرور ادبیات علمی توسط هر فرد، بر روش جستجوی مطالعات توسط او و به تبع داده‌های ورودی به نرم‌افزار، تاثیر می‌گذارد. معمولا کاربست یادگیری ماشینی خوشه‌مبنا در شیوه‌های مرور نظام‌مند و مرور بهترین شواهد، ساده‌تر و دقیق‌تر است. با این حال می‌توان از آن در شیوه مرور روایتی نیز بهره برد.

۳-۲- گام اول: تعیین شیوه جستجو

پس از تعیین موضوع پژوهش و روش مرور ادبیات در آن، گام بعدی جستجوی مطالعات مرتبط است. جستجوی مطالعات به دو شکل جستجوی نظام‌مند و جستجوی روایتی وجود دارد.

۳-۲-۱- جستجوی نظام‌مند

در جستجوی نظام‌مند، افراد با مراجعه به یکی از پایگاه‌های داده، ورودی‌های مرتبط با کار خود شامل کلیدواژه‌ها، بازه زمانی ثبت مقالات، و در صورت نیاز ژورنال یا نشریه مدنظر را انتخاب نموده و علاوه بر بارگیری سند متنی مقالات (pdf.)، یک سند علمی پیوست نیز که در ادامه به آن پرداخته می‌شود، با یکی از فرمت‌های txt. / csv. / ris. / و یا enw. بارگیری می‌گردد. این شیوه جستجو برای مرور نظام‌مند و تا حدودی مرور بهترین شواهد کاربرد دارد [8]. در جستجوی مقالات برای نمونه پژوهی مطالعه حاضر، به صورت زیر از این روش استفاده شد؛

- پایگاه‌های اطلاعاتی: Scopus و Web of Science
- کلیدواژه‌های مورد جستجو: ترکیب زیر

TITLE-ABS-KEY [(“Climate Resilience”) AND. (“Urban Planning/Plan” OR. “City Planning/Plan” OR. “Town Planning/Plan” OR. “Planning/Plan”)].

- بازه زمانی ثبت مقالات: ۲۰۰۱ تا ۲۰۲۱

- نشریه: تمام نشریات

۳-۲-۲- جستجوی روایتی

در جستجوی روایتی فرد منابع را از یک پایگاه خاص دریافت نمی‌کند و می‌تواند مطالب را از پایگاه‌های مختلف جمع‌آوری کرده و برای تحلیل آماده‌سازی نماید. استفاده از موتورهای جستجو مانند Google Scholar نیز در این شیوه امکان‌پذیر است [7]. این شیوه جستجو برای مرور نظام‌مند و مرور بهترین شواهد کاربرد دارد.

نکته حائز اهمیت در استفاده از این شیوه آن است که پژوهشگر تنها سند متنی مطالعات را دریافت می‌کند و باید سند علمی یا به اصطلاح مرجعی (Bibliographic) آن را خود ایجاد نماید. برای ساخت این سند، پژوهشگر باید از یک نرم‌افزار مدیریت اسناد علمی مانند Mendeley یا Endnote استفاده کند. به این صورت که ابتدا باید اسناد متنی را به طور کامل و در قالب یک مجموعه واردساخته و سپس سند علمی یا مرجعی متناظر با آن را خروجی بگیرد. توصیه می‌شود پژوهشگران پیش از تهیه سند علمی یا مرجعی در هر دو روش جستجو، ابتدا با بررسی چکیده مطالعات، مقالات تکراری و نامرتبط با موضوع پژوهش خود را حذف کرده تا نتایج نهایی تحلیل، واقع‌بینانه و دقیق‌تر باشد.

۳-۳- گام دوم: ورود اطلاعات و پردازش اولیه

گام بعدی، ورود اطلاعات به نرم‌افزار و پردازش اولیه بر روی آن است. در این‌جا باید بیان شود که حاصل جستجوی اولیه برای نمونه پژوهی مطالعه در بخش قبل، یافتن ۴۶۷ مقاله بود که پس از حذف عناوین نامرتبط با موضوع به ۳۷۵ مقاله تقلیل یافت. سند مرجعی مرتبط با این مقالات پس از بارگیری از دو پایگاه علمی بیان شده، در نرم‌افزار Mendeley تلفیق و به صورت یک سند علمی یکپارچه با نگارش ris. درآمد.

برای وارد کردن سند مرجعی به نرم‌افزار، پژوهشگر باید ابتدا با انتخاب گزینه Create شیوه ورود اطلاعات را مشخص نماید. پس از وارد کردن اطلاعات، نرم‌افزار دو گزینه برای تحلیل اولیه ارائه می‌دهد که عبارت است از؛

- تحلیل هم‌نگاری (Co-Authorship): این شیوه از تحلیل مبتنی بر همکاری میان مولفان پژوهش‌ها و استناد آن‌ها به یکدیگر است.
- تحلیل هم‌آیندی (Co-Occurrence): این تحلیل مبتنی بر کلمات کلیدی، تعداد تکرار آن‌ها در پژوهش‌ها و ارتباط آن‌ها با یکدیگر بر اساس هم‌آیندی در یک پژوهش است.

نرم‌افزار همچنین دو شیوه محاسبه را برای هر یک از این تحلیل‌ها، بر اساس انتخاب کاربر در نظر می‌گیرد؛

- محاسبه مبتنی بر شمارش کامل (Full Counting): این سبک از محاسبه، وزن یکسانی برای ارتباطات (Links) در نظر می‌گیرد.
- محاسبه مبتنی بر شمارش کسری (Fractional Counting): این سبک از محاسبه، وزن ارتباطات را بر تعداد گره‌ها (Nodes)، که همان نویسندگان یا کلمات کلیدی هستند، تقسیم می‌نماید [4 - 5]

نکته حائز اهمیت در این‌جا آن است که اگر نرم‌افزار نتواند یک شبکه یکپارچه بر اساس نویسندگان یا کلمات کلیدی (به انتخاب کاربر) تشکیل دهد، شبکه‌ها را به صورت جدا از هم و یا تنها بزرگترین شبکه (بر اساس تعداد گره‌ها) را نمایش می‌دهد.

۳-٤- گام سوم: الگوریتم، و تحلیل‌های پیشرفته

پس از انجام پردازش اولیه بر روی مطالعات و تعیین خوشه‌ها، نرم‌افزار امکان اجرای تحلیل‌های پیشرفته‌تری را در اختیار کاربر قرار می‌دهد. حالت‌هایی که در این‌جا در اختیار کاربر قرار می‌گیرد مربوط به شیوه نرمال‌سازی ارتباط میان گره‌هاست. به طور کلی چهار سبک برای نرمال‌سازی این ارتباط وجود دارد؛

- بدون نرمال‌سازی (No Normalization): با انتخاب این گزینه هیچ نوعی از نرمال‌سازی، بر داده‌ها اعمال نمی‌شود. استفاده از این روش معمولا توصیه نمی‌گردد
- نرمال‌سازی مبتنی بر شدت پیوستگی (Association Strength): با انتخاب این گزینه، شدت پیوستگی یا همبستگی گره‌ها به یکدیگر، بر اساس فرمول (۱) در نرمال‌سازی لحاظ می‌شود:

$$S_A(c_{ij}, s_i, s_j) = \frac{c_{ij}}{s_i s_j} \qquad (۱)$$

در فرمول (۱)، c_{ij} تعداد تکرار رابط ij میان دو گره i و j است که بر اساس ماتریس هم‌آیندی (ماتریس O) و از فرمول(۲) محاسبه می‌شود. در ستون ماتریس O عناصری که قرار است تحلیل شوند (مانند کلیدواژه‌ها، نویسندگان و...) قرار گرفته و در ردیف آن، اسناد علمی استقرار می‌یابد. ماتریس O یک ماتریس Binary است که در صورت رخداد عنصر i در سند k، مقدار آن ۱ و در صورت نبود رخداد صفر است.

$$c_{ij} = \sum_{k=1}^{m} o_{ki} o_{kj} \qquad (۲)$$

در فرمول (۲)، o_{ki} سلول متناظر ردیف k و ستون i در ماتریس هم‌آیندی O است (برای o_{kj} نیز به همین منوال است). در این فرمول s_i (s_j)، تعداد رخداد عنصر i (j) در کل اسناد علمی‌ست و خود از فرمول (۳) محاسبه می‌شود [9 - 10].

$$s_i = \sum_{k=1}^{m} o_{ki} \qquad (۳)$$

شیوه نرمال‌سازی مبتنی بر شدت پیوستگی، روش پایه در نرمال‌سازی است. این شیوه به صورت پیش‌فرض در نرم‌افزار فعال است و برای تحلیل‌های اولیه نیز از آن استفاده می‌شود.

- نرمال‌سازی مبتنی بر کسری‌سازی (Fractionalization): در این گزینه، روابط میان عناصر با استفاده از شیوه کسری‌سازی مطابق فرمول (۴) نرمال می‌شود [10]

$$S_G(c_{ij}, s_i, s_j) = \frac{1}{2}\left(\frac{c_{ij}}{s_i} + \frac{c_{ij}}{s_j}\right) \quad (۴)$$

متغیرهای به کار رفته در فرمول بالا مانند متغیرهای معرفی‌شده در بخش قبل است.

- نرمال‌سازی مدولار (LinLog/Modularity): نرمال‌سازی مدولار یکی از پیشرفته‌ترین شیوه‌های نرمال‌سازی است که توسط نواک (۲۰۰۹) توسعه یافت. این شیوه مبتنی بر نرمال‌سازی لگاریتمیک بوده که به دلیل حجم بالای محاسبات، به منبع اصلی جهت آشنایی بیشتر ارجاع داده می‌شود [11].

به طور کلی یکی از این چهار شیوه نرمال‌سازی در انجام محاسبات نرم‌افزار و الگوریتم اجرایی آن، بر اساس انتخاب کاربر استفاده می‌شود. شیوه نرمال‌سازی روابط میان گره‌ها، تاثیر مستقیمی بر خوشه‌بندی و نمایش بصری آن‌ها دارد. به همین خاطر به پژوهشگران توصیه می‌شود که با آگاهی کامل، نسبت به انتخاب هر یک از این شیوه‌های نرمال‌سازی اقدام نمایند.

برای خوشه‌بندی و شبیه‌سازی بصری نتایج، نرم‌افزار از الگوریتمی‌هایی استفاده می‌نماید که به ترتیب به آن‌ها، الگوریتم بهینه‌سازی خوشه‌بندی و الگوریتم بهینه‌سازی چیدمان VOS گفته می‌شود [5 - 10]. در این دو الگوریتم پارامترهای دیگری نیز وجود دارد که در جدول (۱) به آن‌ها پرداخته شده است. این پارامترها بر قدرت و جهت ارتباط گره‌ها اثر ندارد و تنها برای خوشه‌بندی گره‌ها و چیدمان آن‌ها نسبت به یکدیگر، به کار می‌رود. علاوه بر پارامترهای معرفی شده در جدول (۱)، نرم‌افزار گزینه‌های دیگری را (مانند شکل گره‌ها، اندازه فونت، رنگ‌بندی و...) در اختیار کاربر جهت تغییر نمایش بصری قرار می‌دهد، که استفاده از آن‌ها تا حدود زیادی سلیقه‌ایست.

جدول (۱): پارامترهای نرم‌افزار VOS Viewer **در خوشه‌بندی و شبیه‌سازی بصری روابط میان عناصر** [5 – 10]

پارامترها		توضیحات
جاذبه و دافعه	Attraction and Repulsion	این پارامترها بر موقعیت و جایگیری گره‌ها در خوشه‌بندی و نقشه نهایی اثر می‌گذارد. پارامتر جاذبه مقداری میان ۹- تا ۱۰+ و پارامتر دافعه مقداری میان ۱۰- تا ۹+ دارد. شدت پارامتر دافعه همواره کمتر از پارامتر جاذبه است. هرچه شدت جاذبه بیشتر باشد، گره‌ها نزدیکتر و هرچه کمتر، گره‌ها از یکدیگر دورتر است.
شروع تصادفی	Random Starts	این پارامتر تعداد دفعاتی را تعیین می‌کند که الگوریتم بهینه‌سازی خوشه‌بندی و الگوریتم بهینه‌سازی چیدمان VOS اجرا می‌گردد. هر بار که این دو الگوریتم اجرا می‌شود، ممکن است خوشه‌بندی و چیدمان متفاوتی به دست آید. بهترین چیدمان حاصل از الگوریتم بهینه‌سازی به عنوان چیدمان نهایی مورد استفاده قرار می‌گیرد. هرچه مقدار پارامتر شروع تصادفی بالاتر باشد، دقت خوشه‌بندی و چیدمان نهایی به دست آمده بیشتر بوده، اما به تبع زمان پردازش و محاسبات نیز افزایش می‌یابد.
بیشینه تکرار	Max. Iterations	این پارامتر حداکثر تعداد تکرارهای انجام‌شده توسط یک الگوریتم بهینه‌سازی خوشه‌بندی و یک الگوریتم بهینه‌سازی چیدمان VOS را تعیین می‌کند. هرچه مقدار این پارامتر بالاتر باشد، کیفیت خوشه‌بندی و چیدمان نهایی بالاتر است.
متغیرهای اندازه‌ی مرحله	Initial Step Size Step Size Reduction Step Size Convergence	این پارامترها اندازه محاسباتی هر مرحله از الگوریتم بهینه‌سازی چیدمان VOS را مشخص می‌کند. مقدار آن میان ۰.۰۰۰۰۰۱ تا ۱ است. مقدار پیش‌فرض بهینه ترین مقدار بر اساس زمان محاسبه، حجم داده‌ها و کیفیت چیدمان نهایی است
دقت جداسازی	Resolution	این پارامتر سطح جزئیات در نظرگرفته شده در الگوریتم بهینه‌سازی خوشه‌بندی را تعیین می‌کند. این پارامتر مقداری مثبت دارد. هر چه مقدار آن افزایش یابد تعداد خوشه‌های تولید شده بیشتر می‌شود. کارشناسان توصیه می‌نمایند که پژوهشگر مقادیر مختلفی را برای این پارامتر امتحان، و خوشه‌بندی بهینه را انتخاب نمایند.
حداقل اندازه خوشه‌ها	Min. Cluster Size	این پارامتر حداقل تعداد گره را در هر خوشه مشخص می‌نماید. هرچه مقدار این پارامتر بیشتر باشد، تعداد خوشه‌ها کمتر خواهد شد. در صورتی که نرم‌افزار نتواند یک خوشه با حداقل تعداد گره‌های مشخص ایجاد کند، گره‌ها را به صورت بدون خوشه (خاکستری) نشان داده و یا در صورت انتخاب گزینه تلفیق خوشه‌های کوچک (Merge Small Clusters) خوشه‌های با اندازه کمتر را ادغام می‌نماید.

٤- نتایج و یافته‌های پژوهش

در شکل ۱، نتیجه تحلیل اولیه هم‌نگاری بر روی نمونه مورد مطالعه ارائه شده است. در این تحلیل نویسندگانی که حداقل ۳ اثر از آن‌ها وجود داشت، مورد بررسی قرار گرفتند. این تحلیل حاکی از وجود دو دیدگاه کلان میان نویسندگان است که آن‌ها را به دو خوشه تقسیم نموده. بررسی متن مقالات نشان می‌دهد که این نتیجه‌گیری اولیه صحیح است. عده‌ای برنامه‌ریزی در مقیاس کلان و مباحث مدیریتی را اصلی‌ترین عامل در افزایش تاب‌آوری اقلیمی جوامع می‌دانند (طیف زرد رنگ)، در حالی که عده‌ی دیگر بر توانایی اجتماعی–محیطی جهت سازگاری با تغییرات اقلیمی به عنوان مهم‌ترین رکن تاب‌آوری اقلیمی تاکید دارند (طیف آبی رنگ).

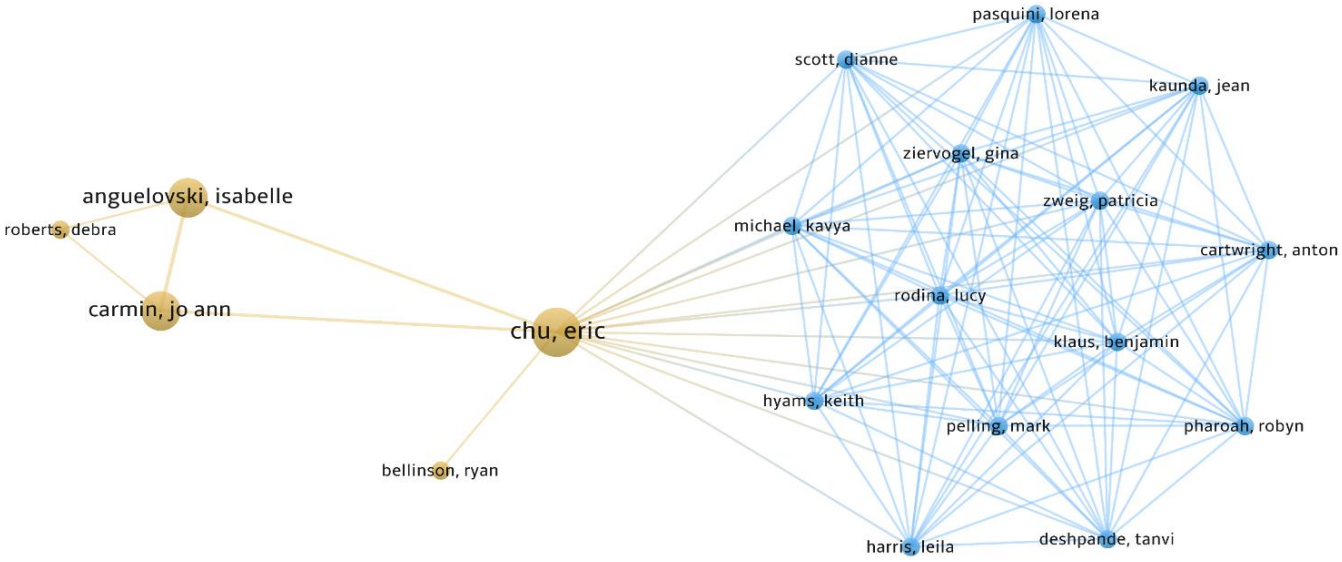
شکل ۲ نتیجه تحلیل اولیه هم‌آیندی بر روی نمونه مورد مطالعه است. در این تحلیل کلمات کلیدی با حداقل پنج بار تکرار در مجموع مطالعات، مورد بررسی قرار گرفته‌اند. در نتیجه این تحلیل کلمات کلیدی در هشت خوشه دسته‌بندی شدند. این نمودار نشان می‌دهد که برای دستیابی به جامعه‌ای تاب‌آور وجود یک برنامه جامع جهت مدیریت و پایش در کنار تقویت زیرساخت‌های اجتماعی و کالبدی ضروری است. به عبارت دیگر تحلیل هم‌آیندی نمونه مورد مطالعه، با نتایج اولیه تحلیل هم‌نگاری همسو است. لازم به ذکر است تعداد نویسندگان و کلمات کلیدی در این دو حالت، بیشتر از گره‌های نمودار شکل (۱) و (۲) است، اما چون نرم‌افزار نتوانست یک شبکه یکپارچه میان تمامی نویسندگان ایجاد کند، بزرگترین شبکه بر اساس تعداد گره نمایش داده شده است.

شکل (۳)، ابر واژگان (Word Cloud) نمونه‌ی مورد بررسی را نشان می‌دهد. بر اساس این تصویر، علاوه بر بحث تاب‌آوری اقلیمی، مطالعات به مباحثی چون سازگاری اقلیمی، دگردیسی اقلیمی، تغییرات اقلیمی و برنامه‌ریزی کاهشی به طور گسترده پرداخته‌اند.

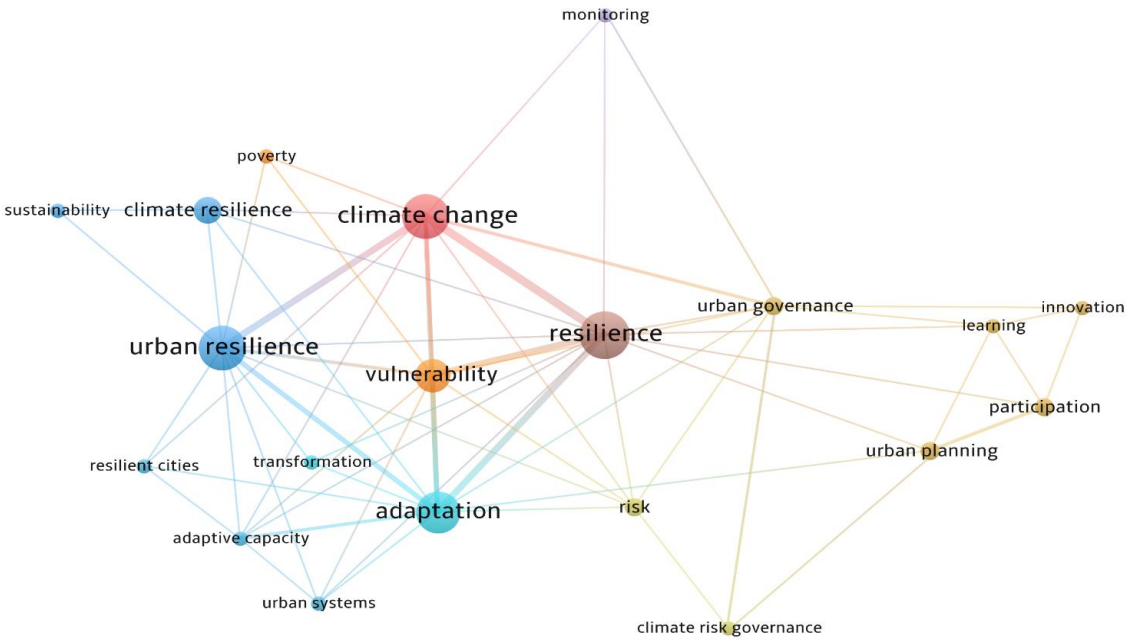
بررسی سیر تحول پژوهش‌ها در شکل (۴)، نشان می‌دهد که تاب‌آوری اقلیمی از اقدامات نهادی و کالبدی صرف در حال حرکت به سوی برنامه‌ریزی تلفیقی است. در این شکل نمایش بصری کلمات کلیدی بر پایه تحولات آن‌ها در گذر زمان ارائه شده است. نکته حائز اهمیت آن است که مطالعات در سال‌های اخیر، مسائل کارکردی را نیز مورد بررسی قرار داده و سعی در ارائه الگوی یکپارچه‌تری جهت افزایش تاب‌آوری اقلیمی شهرها دارند.

٥- جمع‌بندی و پیشنهادهایی برای مطالعات آینده

پژوهش حاضر در راستای ادای رسالت خود به عنوان یک مطالعه علمی–ترویجی، به تشریح شیوه کاربست یادگیری ماشینی خوشه‌مبنا و نرم‌افزار VOS Viewer در تحلیل متنی مقالات نمایه شده در وب، به همراه بررسی یک نمونه پرداخت. در این پژوهش تلاش شد تا با آموزش شیوه استفاده از نرم‌افزار VOS Viewer و بررسی الگوریتم‌های آن، فرایند علم‌سنجی و تحلیل متنی را برای پژوهشگران حوزه‌های علمی مختلف، تسهیل نماید. حال که شیوه کارکردن با این تکنیک به وضوح برای پژوهشگران مشخص شده، به آن‌ها پیشنهاد می‌گردد تا از این نرم‌افزار و الگوریتم آن در تشخیص کمبودهای مطالعاتی، شناسایی ایده‌های حاکم بر مولفان و همچنین بررسی کلان‌روندهای موجود در حوزه‌های تخصصی علوم استفاده نمایند. بدون شک کاربست این شیوه منجر به شناسایی بهتر نقاط قوت، ضعف‌ها و نیازها در حوزه‌های مختلف علمی و دقیق‌تر شدن نتیجه‌گیری پژوهش‌ها می‌شود

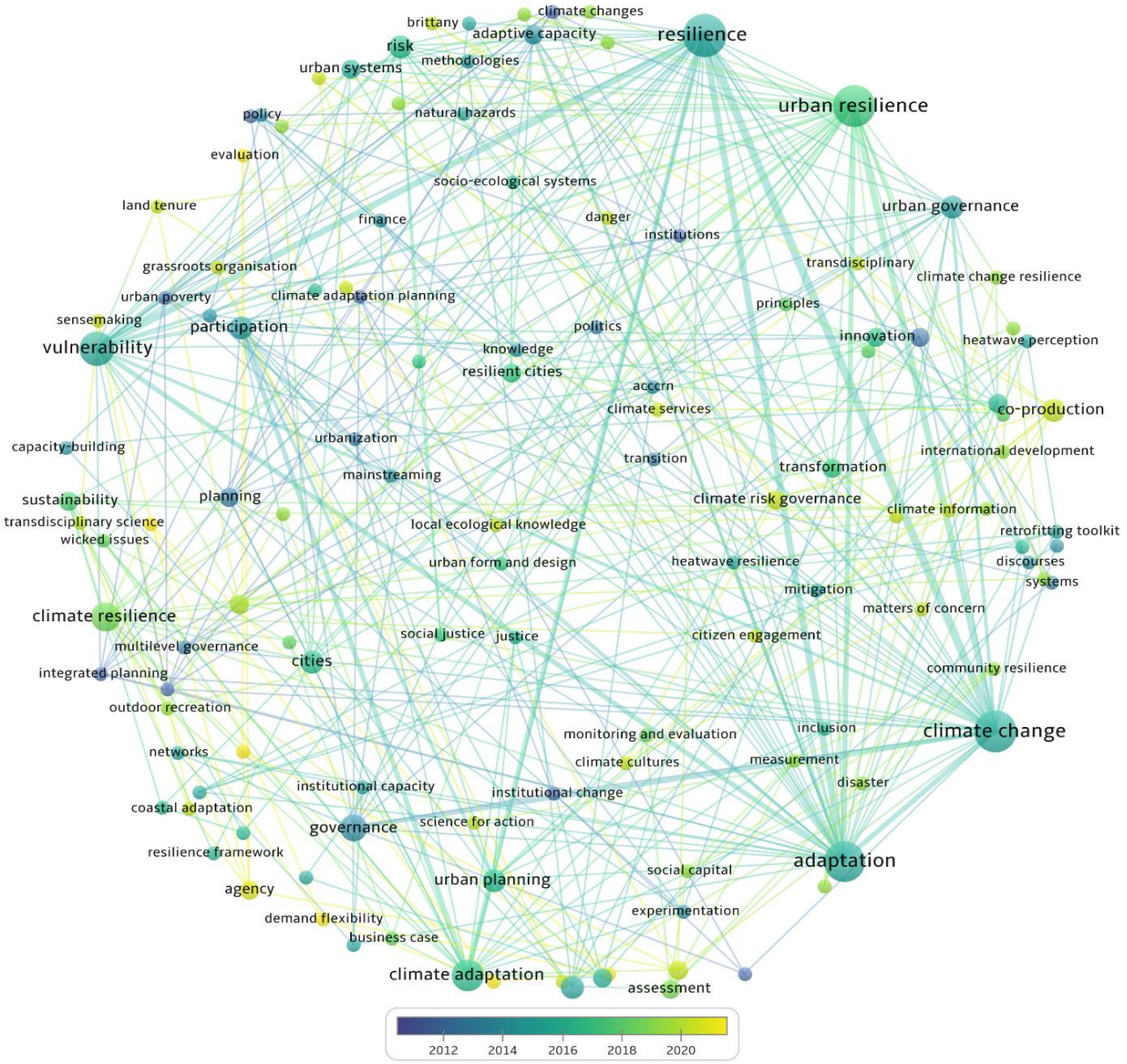


شکل (۱): تحلیل اولیه هم‌نگاری بر روی نمونه مورد مطالعه، با شرط N>2



شکل (۲): تحلیل اولیه هم‌آیندی بر روی نمونه مورد مطالعه، با شرط N>4

شکل (۳): نمودار ابر واژگان نمونه مورد مطالعه، با شرط N>5



شکل (٤): سیر تحول مطالعات مرتبط با نمونه مورد بررسی بر اساس کلمات کلیدی پژوهش‌ها، با شرط N>2

مراجع

[۱] سهرابی، محمدرضا. **"اصول نگارش مقالات مروری"**. نشریه پژوهنده، شماره ۲، صص: ۵۲-۵۶، دانشگاه شهید بهشتی، ۱۳۹۲.

[1] Lim, K.W., Buntine, W. *"Bibliographic analysis on research publications using authors, categorical labels and the citation network"*. Machine Learning. Vol 103, pp: 185–213, 2016. doi.org/10.1007/s10994-016-5554-z

[2] Scopus Publication. *"Scopus roadmap: what's coming up in 2020 & 2021?"* Website Report, 2021. Access on: https://blog.scopus.com/posts/scopus-roadmap-whats-coming-up-in-2020-2021

[3] Web of Science platform. *"Web of science: summary of coverage"*, Website Annual Report, 2021. Access on: https://clarivate.libguides.com/webofscienceplatform/coverage

[4] van Eck, N.J., Waltman, L. *"Citation based clustering of publications using CitNet Explorer and VOS Viewer"*, Scientometrics, Vol 111, pp: 1053-1070, 2017. doi.org/10.1007/s11192-017-2300-7

[5] van Eck N.J., Waltman L. *"Visualizing bibliometric networks"*. Measuring Scholarly Impact, pp: 285-320. 2014. doi.org/10.1007/978-3-319-10377-8_13

[6] Cooper, H., Hedges, V, *"Research synthesis as a scientific process"*, in H. Cooper, L. V. Hedges, and J. C. Valentine,

eds, Research Synthesis as A Scientific Process. Research Synthesis and Meta-Analysis, New York: Russell Sage Foundation, pp: 3-18, 2009

[7] Cipriani A, Geddes J. *"Comparison of systematic and narrative reviews: The example of the atypical antipsychotics"*. Epidemiologia Psichiatria Sociale. 2003, *10.1017/s1121189x00002918*

[8] Bramer, W., de Jonge, G., Rethlefsen, M.L., Mast, F., Kleijnen, J. *"A systematic approach to searching: an efficient and complete method to develop literature searches"*, Journal of the Medical Library Association, 2018, *10.5195/jmla.2018.283*

[9] van Eck N.J., Waltman L. *"How to normalize cooccurrence data? An analysis of some well-known similarity measures"*, Journal of the American Society for Information Science and Technology, Vol 60, pp: 1635-1651, 2009, *doi.org/10.1002/asi.21075*

[10] van Eck N.J., Waltman L. *"VOS Viewer manual"*, University of Liden Publication, Access on: *https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.8.pdf*

[11] Noack, A., *"Modularity clustering is force-directed layout"*, Journal of Physical Review, Vol 79, 2009, *10.1103/PhysRevE.79.026102*