Archive of SID

تاثیر روش‌های اندازه‌گیری فاصله در طبقه‌بندی گروه‌های بوم‌شناسی در جنگل‌های هیرکانی

نغمه پاک‌گهر[۱]، جواد اسحاقی‌راد[۲*]، غلامحسین غلامی[۳]، احمد علیجانپور[۲] و دیوید وی. رابرتز[۴]

۱) دکتری رشته جنگلداری، دانشکده منابع طبیعی، دانشگاه ارومیه، ارومیه، ایران.

۲) دانشیار گروه جنگلداری، دانشکده منابع طبیعی، دانشگاه ارومیه، ارومیه، ایران. *رایانامه نویسنده مسئول مکاتبات: j.eshagh@urmia.ac.ir

۳) استادیار گروه ریاضی، دانشکده علوم، دانشگاه ارومیه، ارومیه، ایران.

۴) استاد گروه بوم‌شناسی، دانشگاه ایالتی مونتانا، بازمن، ایالات متحده آمریکا.

تاریخ دریافت: ۱۴۰۰/۰۴/۰۵ تاریخ پذیرش: ۱۴۰۰/۰۶/۲۷

چکیده

امروزه به صورت گسترده از الگوریتم‌های خوشه‌بندی استفاده می‌شود، هرچند که تصمیم‌گیری برای انتخاب روش مناسب به دلیل روش‌های مختلف خوشه‌بندی و عوامل موثر بر آنها دشوار است. هدف این پژوهش مقایسه نتایج روش‌های خوشه‌بندی سلسله مراتبی و بررسی روش اندازه‌گیری فاصله موثر برای خوشه‌بندی است. برای این پژوهش، داده‌های جنگل‌های راش هیرکانی واقع در حوزه استحافظی اداره کل منابع طبیعی نوشهر مورد بررسی قرار گرفت. برای تعیین مراکز قطعات نمونه از روش منظم تصادفی با ابعاد شبکه ۲۰۰×۱۰۰ استفاده شد و در هر مرکز قطعه نمونه، قطعات نمونه صد مترمربعی (۱۰×۱۰ متر) برای بررسی گونه‌های علفی و چهارصد مترمربعی (۲۰×۲۰ متر) برای بررسی گونه‌های درختی و درختچه‌ای انتخاب شد. در مجموع ۱۲۰ قطعه نمونه اندازه‌گیری شد. برآورد فراوانی و پوشش گونه‌های درختی، درختچه‌ای و علفی بر اساس مقیاس براون بلانکه انجام پذیرفت. سه روش اندازه‌گیری فاصله بری کورتیس، هلینگر و منهتن و الگوریتم‌های خوشه‌بندی، روش میانگین، روش وارد، روش بتای انعطاف‌پذیر با مقدار بتای ۰/۱-، ۰/۲۵-، ۰/۴- برای پژوهش انتخاب شدند و با استفاده از شش شاخص ارزیابی‌کننده (شاخص میانگین سیلوئت، آنالیز پارتیشن، آنالیز گونه‌های معرف، آنالیز گونه‌های معرف خوشه‌ها برای کمینه کردن ثبات میانی، روش پاسخ چندگانه جایگشت و ضریب فی) مورد بررسی قرار گرفتند. نتایج هر معیار ارزیابی‌کننده از بهترین به بدترین رتبه دسته‌بندی شدند. یافته‌ها نشان داد روش خوشه‌بندی وارد و روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱- بهترین عملکرد را دارد و روش اندازه‌گیری فاصله هلینگر در داده‌های همگن بهتر از سایر روش‌های اندازه‌گیری فاصله است.

واژه‌های کلیدی: جنگل‌های هیرکانی، خوشه‌بندی، روش اندازه‌گیری فاصله، طبقه‌بندی سلسله‌مراتبی.

مقدمه

روش‌های طبقه‌بندی عددی یکی ابزارهای اصلی تحلیلی در بوم‌شناسی گیاهی و سایر علوم همچون روانشناسی، ژنتیک و یا جامعه‌شناسی است (Lengyel *et al.*, 2019). طبقه‌بندی در جوامع گیاهی سبب شناخت ترکیب جوامع می‌شود که این شناخت سرآغاز مدیریت بهینه و پایش جوامع و حفظ منابع است (Lengyel *et al.*, 2018)، به‌دلیل اهمیتی که پوشش‌گیاهی در بوم‌سازگان دارد، عمدتا برای طبقه‌بندی جوامع مورد استفاده قرار می‌گیرد زیرا نشان‌دهنده تغییرات عوامل زنده و غیرزنده بوم‌سازگان است (Peet & Roberts, 2013). اما با وجود هزاران گونه با ویژگی‌های اکولوژیکی متفاوت، طبقه‌بندی امری مشکل است (McKenna, 2003). طبقه‌بندی خوشه‌ای را می‌توان نوع خاصی از طبقه‌بندی پوشش گیاهی دانست، خوشه‌بندی یک الگوریتم طبقه‌بندی بدون ناظر است که نمونه‌ها را از یکدیگر جدا می‌کند و نمونه‌های مشابه را کنار یکدیگر قرار می‌دهد (Rodriguez *et al.*, 2019). در میان روش‌های مختلف طبقه‌بندی، روش خوشه‌بندی را می‌توان یکی از پرکاربردترین

www.SID.ir

Archive of SID

روش‌ها دانست. روش‌های مختلف خوشه‌بندی به بخش روش اصلی روش‌های خوشه‌بندی سخت[1] و خوشه‌بندی نرم[2] تقسیم می‌شوند (Wang *et al.*, 2019). خوشه‌بندی سخت خود به دو بخش خوشه‌بندی سلسله مراتبی و غیرسلسله مراتبی تبدیل می‌شود. خوشه‌بندی سلسله مراتبی که در این پژوهش به آن پرداخته شده است، یک ساختار درختی برای نمایش نتایج خوشه‌بندی ایجاد می‌کند. همچنین این روش‌ها وابسته به ماتریس تشابه/عدم‌تشابه یا معیار فاصله هستند (Zaki *et al.*, 2014; Hämäläinen *et al.*, 2017). روش‌های خوشه‌بندی سلسله مراتبی به روش‌های تجمعی و روش‌های مقسمی تفکیک می‌شوند (Hämäläinen *et al.*, 2017). روش‌های سلسله مراتبی تجمعی از یک قطعه نمونه به‌عنوان اولین خوشه شروع می‌شوند و سپس در روند انجام خوشه‌بندی سایر قطعه نمونه‌ها به خوشه‌ها اضافه می‌شوند تا زمانی که تمام قطعات نمونه در یک خوشه قرار گیرند. از سوی دیگر، در روش‌های مقسمی در هنگام شروع، تمامی قطعات نمونه در یک خوشه قرار دارند و در ادامه در خوشه‌ها قرار می‌گیرند تا زمانی که هر قطعه نمونه در یک خوشه منحصر به فرد قرار گیرد (Rolecek *et al.*, 2009; Fern´andez *et al.*, 2019). روش‌های خوشه‌بندی غیرسلسله مراتبی، قطعات نمونه را به‌طور مستقیم در داخل مراکز خوشه‌هایی قرار می‌دهند که از قبل تعیین شده‌اند، سپس با جابه‌جایی قطعه نمونه داخل خوشه‌ها و تغییر مراکز خوشه، خوشه‌بندی نهایی آماده می‌شود (Arora & Varshney, 2016). با وجود توسعه روش‌های خوشه‌بندی و روش‌های مختلف موجود در دسترس هنوز طبقه‌بندی گروه‌های اکولوژیکی وابسته به تصمیمات متعدد ذهنی است (De Cáceres *et al.*, 2015). از جمله این عوامل در آنالیز خوشه‌ای می‌توان به انتخاب تعداد بهینه خوشه‌ها، روش اندازه‌گیری ماتریس فاصله، انتخاب روش خوشه‌بندی، روش ادغام خوشه‌های مختلف و تبدیل داده‌ها اشاره کرد (Lengyel & Podani, 2015).

با وجود مطالعاتی که پژوهشگران انجام داده‌اند هنوز دستورالعمل واحدی برای انجام خوشه‌بندی تعیین نشده است از جمله این مطالعات می‌توان به پژوهش اسحاقی‌راد و همکاران (۱۳۸۷) اشاره کرد که تعداد بهینه خوشه‌ها را با استفاده از تجزیه و تحلیل خوشه‌ای در جوامع راش جنگل خیرودکنار نوشهر با استفاده از روش تجزیه و تحلیل گونه‌های معرف، مورد بررسی قرار دادند. نتایج نشان داد تعداد مناسب خوشه‌ها در جوامع راش منطقه مورد مطالعه چهار گروه است. در پژوهشی دیگر، با استفاده از شاخص تعلقه فی، گونه‌های معرف گروه‌های اکولوژیکی جنگل حفاظت شده خیبوس تعیین شدند، سپس با استفاده از دو شاخص تمایز و انحصارگرای ارزش گونه‌های معرف در ارزیابی کیفیت طبقه‌بندی گروه‌ها مشخص شد و نتایج نشان داد که روش عضویت‌گیری TPFIM سبب بهبود عملکرد آنالیز گونه‌های شاخص دوطرفه می‌شود (اسماعیل‌زاده و اسدی، ۱۳۹۳). Lengyel و Podani (۲۰۱۵) اهمیت نسبی انتخاب روش شامل اندازه قطعات نمونه و تجزیه و تحلیل داده‌ها در طبقه‌بندی پوشش گیاهی را مورد بررسی قرار دادند و به این نتیجه رسیدند که روش طبقه‌بندی، سطح خوشه، تبدیل داده‌ها و میانگین اندازه قطعات نمونه همواره از جمله مهم‌ترین متغیرها در طبقه‌بندی داده‌ها هستند. Lengyel و همکاران (۲۰۱۸) تعداد بهینه خوشه‌ها و تاثیر تبدیل داده‌ها برای بهینه کردن نتایج خوشه‌بندی را مورد بررسی قرار دادند. نتایج آنها نشان داد تعداد بهینه خوشه‌ها با تغییر توان در تبدیل توانی تغییر می‌کند. در نتیجه هر یک عوامل یاد شده بر نتایج تجزیه و تحلیل خوشه‌ای تاثیرات معنی‌داری دارد (De Cáceres *et al.*, 2015). به‌این ترتیب تشخیص درستی و نادرستی نتایج حاصل به‌وسیله پژوهشگر بسیار مشکل و گمراه‌کننده خواهد بود (Bertrand & Bel Mufti, 2006). با توجه به عوامل یاد شده و اهمیت طبقه‌بندی در اجتماعات گیاهی برای یافتن الگو نهفته در بوم‌سازگان و خلاصه کردن داده‌های پیچیده و چندمتغیره، این پژوهش در نظر دارد تاثیر روش‌های محاسبه ماتریس فاصله بر روی نتایج خوشه‌بندی‌های سلسله مراتبی تجمعی را به‌عنوان اولین گام شروع خوشه‌بندی مورد بررسی قرار دهد.

مواد و روش‌ها

منطقه مورد بررسی

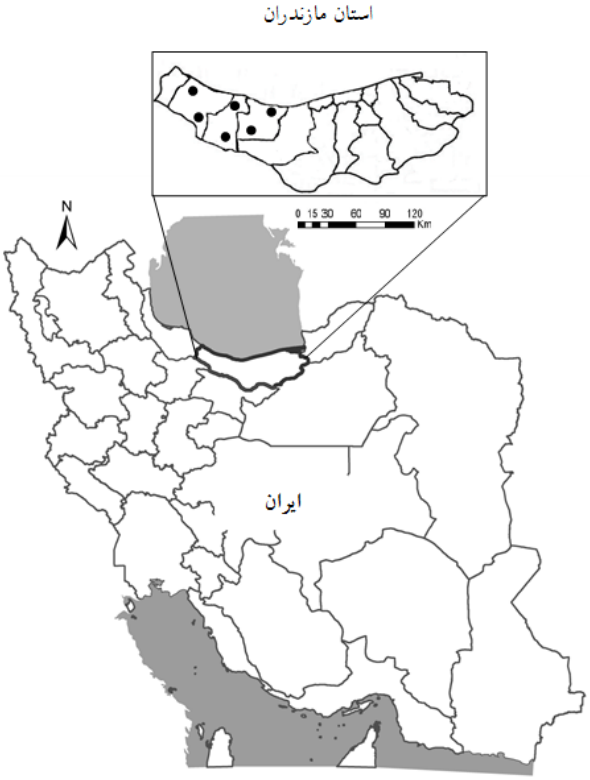
در این پژوهش با توجه به اهداف پژوهش حاضر و جنگل‌گردشی مناطق مورد مطالعه در سال ۱۳۹۵، داده‌های پوشش گیاهی از پارسل‌های ۲۹ (مدیریت‌شده) و ۲۶ (شاهد)

1 Hard clustering
2 Soft clustering

www.SID.ir

Archive of SID

از سری ۱۰ لالیس (طول‌های جغرافیایی "۳۰ ،'۲۳ ،°۵۱ تا '۲۸ ،°۵۱ و عرض جغرافیایی "۴۵ ،'۳۲ ،°۳۶)، پارسل‌های ٤١٤ (مدیریت‌شده) و ٤١٢ (شاهد) سری دهگا (طول‌های جغرافیایی "۳٤، '۵۲ ،°۵۰ تا "۱۵، '۷۵ ،°۵۰ و بین عرض‌های جغرافیایی "۱۵ ،'٤٠ ،°۳۶ تا "۲۵ ،'٤٤ ،°۳۶)، پارسل‌های ۱۳٤ (مدیریت‌شده) و ۱۳۵ (شاهد) از سری ۱ شیراکنس (طول‌های جغرافیایی "۰۰ ،'۳۰ ،°۵۱ تا "۲۸ ،'۳۳ ،°۵۱ و عرض‌های جغرافیایی"۳۵، '۲۷ ،°۳۶ تا "۰۰، '۳۰ ،°۳۶)، پارسل‌های ۵۳۰ (مدیریت‌شده) و ٥٤١ (شاهد) از سری ۵ لاکوبن (طول‌های جغرافیایی"۱۵، '۱٦، °۵۱ تا "۵۰، '۱۸، °۵۱ و عرض‌های جغرافیایی "۵۰، '۳٦، °۳٦ تا '٤٠ ،°۳٦)، پارسل‌های ۷۲۵ (مدیریت‌شده) و ۷۲٤ (شاهد) از سری ۷ واشمرد (طول‌های جغرافیایی '۱۸ ،°۵۱ تا '۲۲ ،°۵۱ و عرض‌های جغرافیایی '۳۷ ،°۳٦ تا '۳۹ ،°۳٦)، از سری ۳ جمند، پارسل‌های ۳۱۷ (مدیریت‌شده) و ۳۱۸ (مدیریت‌نشده) (طول جغرافیایی "۳۰ ،'۰۰ ،°۵۱ تا "۲۸ ،'۳۳ ،°۵۱ و عرض جغرافیایی "۰۰ ،'۳۰ ،°۳٦ تا "۲۷ ،'۳۵ ،°۳٦) واقع در حوزه استحفاظی اداره کل منابع طبیعی نوشهر استان مازندران برداشت شد. جنگل‌های مدیریت شده مورد مطالعه ۲ تا ۳ دوره مورد بهره‌برداری قرار گرفتند و در جنگل‌های مدیریت‌نشده در بازه زمانی ۳۰ تا ۵۰ سال گذشته، بهره‌برداری انجام نشد. موقعیت سری‌های مورد مطالعه در شکل ۱ ارایه شده است (Khanalizadeh *et al.*, 2020). برای تعیین مراکز قطعات نمونه از روش منظم تصادفی با ابعاد شبکه ۲۰۰×۱۰۰ استفاده شد. در هر مرکز قطعه نمونه، قطعات نمونه ۱۰۰ مترمربعی (۱۰× ۱۰ متر) برای بررسی پوشش علفی، ٤۰۰ مترمربعی (۲۰×۲۰ متر) برای بررسی گونه‌های درختی و درختچه‌ای انتخاب شد. برآورد فراوانی و پوشش گونه‌های درختی، درختچه‌ای و علفی بر اساس مقیاس براون بلانکه انجام شد (اسحاقی‌راد و همکاران، ۱۳۸۸). میانگین هر طبقه (از جدول براون بلانکه) به‌عنوان میانگین درصد پوشش هر گونه در نظر گرفته شد. از هر سری ۲۰ قطعه‌نمونه برداشت شد (هر پارسل ۱۰ قطعه نمونه). در مجموع ۱۲۰ قطعه‌نمونه و ۱۰۰ گونه در این پژوهش مورد بررسی قرار گرفت.



شکل ۱. موقعیت سری‌های مورد مطالعه در ایران و استان مازندران

www.SID.ir

Archive of SID

روش‌های تجزیه و تحلیل داده‌ها

در این تحقیق برای تعیین گروه‌های اکولوژیکی، خوشه‌بندی سلسله مراتبی تجمعی گروه‌ها استفاده شد. ابتدا با استفاده از تبدیل داده هلینگر داده‌ها تبدیل شدند تا پراکنش داده‌ها دارای تقارن بیشتری باشد و تعداد صفرها کم شود (Legendre & Gallagher, 2001). در قدم بعدی، با توجه به ماهیت و همگنی داده‌ها تعداد خوشه‌ها (۲ تا ٦ خوشه) مشخص و خوشه‌بندی انجام شد. برای انجام خوشه‌بندی ابتدا ماتریس فاصله بین قطعات نمونه محاسبه شد، سپس بر اساس معیاری مشخص از فاصله بین قطعات نمونه یا گروه‌ها، دو گروه انتخاب و در هم ادغام شدند. بعد از انجام خوشه‌بندی کیفیت خوشه‌بندی‌ها با استفاده از معیارهای ارزیابی‌کننده مورد بررسی قرار گرفتند و نتایج ارزیابی رتبه‌بندی شدند. در نهایت میانگین هر ارزیابی‌کننده‌ها برای معرفی ماتریس فاصله و خوشه‌بندی بهتر محاسبه شد و تمامی محاسبات در نرم افزار R 3.6.1 انجام شد (R Core Team, 2019). در ادامه به تفضیل هر مرحله در تجزیه و تحلیل شرح داده می‌شود.

– تبدیل داده هلینگر برای داده‌های فراوانی بسیار مناسب است که وزن کمی به گونه‌های نادر می‌دهد. تبدیل داده‌ها شامل تقسیم هر مقدار به مجموع هر ردیف در ماتریس داده‌ها می‌شود و سپس مجذور هر مقدار محاسبه می‌گردد (Legendre & Gallagher, 2001).

– در این پژوهش بر اساس بررسی سایر مطالعات (Aho *et al.*, 2008; Schmidtlein *et al.*, 2010; Lotter *et al.*, 2013; Roberts, 2015) مقیاس‌های فاصله در آنالیزهای پوشش گیاهی و روش‌های طبقه‌بندی انتخاب شدند و برای اتصال خوشه‌ها نیز با بهره‌گرفتن از سایر مطالعات و نتایج آنها، روش‌های اتصال مذکور استفاده شد (جدول ۱) (Aho *et al.*, 2008; Schmidtlein *et al.*, 2010; Lotter *et al.*, 2013; Roberts, 2015; Lengyel & Podani, 2015; Lyons *et al.*, 2016). با توجه اهمیت انتخاب مقدار بتا در خوشه‌بندی بتای انعطاف‌پذیر، در این بررسی مقدار بتا ۰/۱–، ۰/۲۵– و ۰/٤– انتخاب شد تا محدوده موثر بتا مورد ارزیابی قرار گیرد (Lance & Williams 1967; Tichy *et al.*, 2011; Lotter *et al.*, 2013; Lengyel *et al.*, 2019).

روش‌های اندازه‌گیری فاصله

– ضریب فاصله اقلیدوسی (رابطه۱) بر اساس خواص اقلیدوسی مثلث قائم‌الزویه است که در آن مربع وتر برابر مجموع مربعات دو ضلع دیگر است (Orloci, 1974).

$$\mathrm{ED}_{ih} = \sqrt{\sum_{k=1}^{n}(x_{1k} + y_{2k})^2} \quad \text{رابطه ۱}$$

ED: برابر فاصله اقلیدوسی بین دو قطعه نمونه، n: تعداد گونه، x_{1k}: وفور گونه kام در قطعه نمونه اول و y_{2k}: وفور گونه kام در قطعه نمونه می‌باشد.

ضریب فاصله اقلیدوسی با استفاده از تبدیل داده هلینگر به تبدیل روش اندازه‌گیری فاصله هلینگر است (Legendre & Gallagher, 2001) که در ادامه به جای استفاده از ضریب فاصله اقلیدوسی از روش اندازه‌گیری هلینگر استفاده شده است.

– از روش اندازه‌گیری فاصله بری کورتیس (رابطه ۲) که بر اساس مطالعات متعدد روش اندازه‌گیری مناسب برای داده‌های جوامع اکولوژیکی گیاهی است، استفاده شد (Bray & Curtis, 1957; Roberts, 2015) و برای اتصال خوشه‌ها از روش اتصال بتای انعطاف‌پذیر استفاده شد. این روش توسط Lance و Williams (۱۹٦۷) معرفی شد و برای این منظور، آنها روش اتصال معمول خوشه‌ها را تغییر دادند و یک فاصله بین خوشه k و خوشه ij ایجاد کردند.

$$BC_{1,2} = \frac{2\sum_{i=1}^{m}\min(Y_{1K}, Y_{2K})}{\sum_{k=1}^{m}(Y_{1K}+Y_{2K})} \quad \text{رابطه ۲}$$

برای ماتریس داده‌ای با m گونه، مقدار kام گونه در iام قطعه نمونه برابر با Y_{ik} است.

– ضریب فاصله منهتن (رابطه ۳) فاصله مطلق بین دو بردار است (Craw, 2011).

$$Manhattan(x.y) = \sum_{i=1}^{n}|x_i - y_i| \quad \text{رابطه ۳}$$

فاصله بین دو نقطه $x=(x_1, x_2, \ldots\ldots x_n)$ و $y=(y_1,y_2,\ldots\ldots y_n)$ در فضای n بعدی برابر مجموع فواصل در هر بعد است.

www.SID.ir

Archive of SID

جدول ۱. روش‌های مختلف اتصال خوشه‌ها

روش‌ها	توضیحات
روش میانگین	فاصله متوسط بین دو نقطه، این روش حدفاصل بین روش نزدیکترین همسایه و دورترین همسایه است
روش وارد	این روش واریانس کل در داخل خوشه‌ها را کمینه می‌کند، تمایل به ایجاد اندازه‌های یکسان خوشه‌ها و خوشه‌های کروی دارد. حساس به داده‌های پرت است.
روش بتای انعطاف‌پذیر	زمانی که میزان بتا ۰/۲۵- باشد نتایج این روش همانند روش وارد است و خوشه‌های کروی ایجاد می‌کند. میزان بتا در این مطالعه ۰/۱-، ۰/۲۵- و ۰/٤- انتخاب شده است.

معیارهای ارزیابی کیفیت خوشه‌بندی

- شاخص میانگین سیلوئت: این معیار برای پیدا کردن تعداد بهینه خوشه‌ها و ارزیابی کیفیت خوشه‌ها استفاده می‌شود. این معیار از فشردگی (میانگین فاصله داخل خوشه‌ها) و جدایش (میانگین فاصله بین خوشه‌ها) استفاده می‌کند. میانگین سیلوئت s(i) به‌صورت رابطه ٤ تعریف می‌شود:

$$s(i) = \frac{b(i)-a(i)}{max\{a(i).b(i)\}} \quad \text{رابطه ٤}$$

که در آن a(i) عدم‌تشابه بین نمونه‌ها و b(i) عدم‌تشابه با نزدیک‌ترین خوشه است. محدوده این شاخص بین ۱- تا ۱ است. مقدار s(i) نزدیک ۱ نشان‌دهنده طبقه‌بندی خوب است. اگر s(i) صفر باشد طبقه‌بندی منطقی به‌نظر نمی‌رسد و اگر s(i) ۱- باشد نشان‌دهنده طبقه‌بندی اشتباه است (Rousseeuw, 1987).

- آنالیز گونه‌های معرف: برای شناسایی گونه‌هایی که شاخص گروه‌ها هستند، استفاده می‌شوند (Dufrêne & Legendre, 1997). بر اساس فراوانی داخل گونه‌ها و مقایسه وقوع آنها است (Legendre & Legendre, 1998). میزان این شاخص بین ۰ تا ۱ است اما به صورت درصد نیز بیان می‌شود. معنی‌داری گونه‌های معرف توسط آزمون مونت‌کارلو بررسی می‌شود. سپس از تعیین گونه‌های معرف در هر خوشه، خوشه‌های با بیشترین گونه‌های معرف انتخاب می‌شوند.

- ضریب فی: میزان وفادارای گونه‌ها را براساس داده‌های حضور و عدم‌حضور محاسبه می‌کند. این شاخص قادر است میزان وفاداری گونه‌ها را در داده‌های با تعداد رویشگاه‌های مختلف محاسبه کند. ضریب همبستگی فی (Φ) محدوده بین ۱- تا ۱+ است. میزان Φ مثبت برای شناسایی گونه‌های معرف مهم است و میزان منفی Φ برای پیداکردن میزان اختلاف بین جوامع اهمیت دارد (Tichy & Chytry, 2006).

- آنالیز گونه‌های معرف خوشه‌ها برای کمینه‌کردن ثبات میانی: تجزیه و تحلیلی است که ثبات گونه‌ها را بین خوشه‌ها محاسبه می‌کند (رابطه ۵)، گونه‌هایی که همیشه در خوشه‌ها حضور دارند یا همیشه بین گونه‌ها غایب هستند (Roberts, 2015).

$$I = \frac{\sum_{i=1}^{n}\{(2\sum_{k=1}^{g}|C_{ik}-0.5|)/g}{n} \quad \text{رابطه ۵}$$

I: میانگین شاخص، n: تعداد گونه‌ها، g (k = 1, 2, ..., g): تعداد خوشه‌ها (k = 1, 2, ..., g)، C_{ik}: ثبات گونه‌ها i در خوشه k . میزان شاخص بین ۱ و ۱- است و میزان بالای این شاخص نشان‌دهنده خوشه‌بندی بهتر است.

- آنالیز پارتیشین: این آنالیز (رابطه ٦) میزان نسبت تشابه درون‌خوشه‌ای به برون‌خوشه‌ای را محاسبه می‌کند (Aho *et al.*, 2008).

$$P = \frac{\sum_{z=1}^{c}\sum_{\substack{i=1\\ i\in z}}^{N-1}\sum_{\substack{j=i+1\\ j\in z}}^{N} S_{ij}/\sum_{z=1}^{c}(n_z^2-n_z)/2}{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} S_{ij}/\sum_{z=1}^{c-1}\sum_{k=z+1}^{c} n_z.n_k} \quad \text{رابطه ٦}$$

P: نسبت آنالیز پارتیشن، N: تعداد کل نمونه‌ها، C: تعداد خوشه‌ها، n_z: تعداد نمونه در خوشه zام (z = 1, 2, ..., C)، S_{ij}: تشابه بین نمونه i و j می‌باشد. بنابراین نشان می‌دهد که i و j نمونه‌هایی از خوشه‌های متفاوت هستند. مقادیر بیشتر این شاخص نشان‌دهنده خوشه‌بندی بهتر است.

- روش جایگشت پاسخ چندگانه: یک روش غیرپارامتری است که برای مقایسه واحدهای نمونه گروه‌بندی‌شده استفاده می‌شود (Mielke & Berry, 2007). روش پاسخ چندگانه جایگشت یک روش آماری برای بررسی اختلاف گروه‌ها با در نظر گرفتن کلیه متغیرها است. این روش همچنین دو آماره T و A را تولید می‌کند. آمارهٔ T برای تعیین مقدار اختلاف بین گروه‌های طبقه‌بندی‌شده استفاده می‌شود و آمارهٔ A همگنی داخل گروه‌ها را تعیین می‌کند.

نتایج

پنج روش خوشه‌بندی با تعداد مختلف خوشه (۲ تا ٦ خوشه -۵ تکرار) با سه روش اندازه‌گیری فاصله مورد بررسی

www.SID.ir

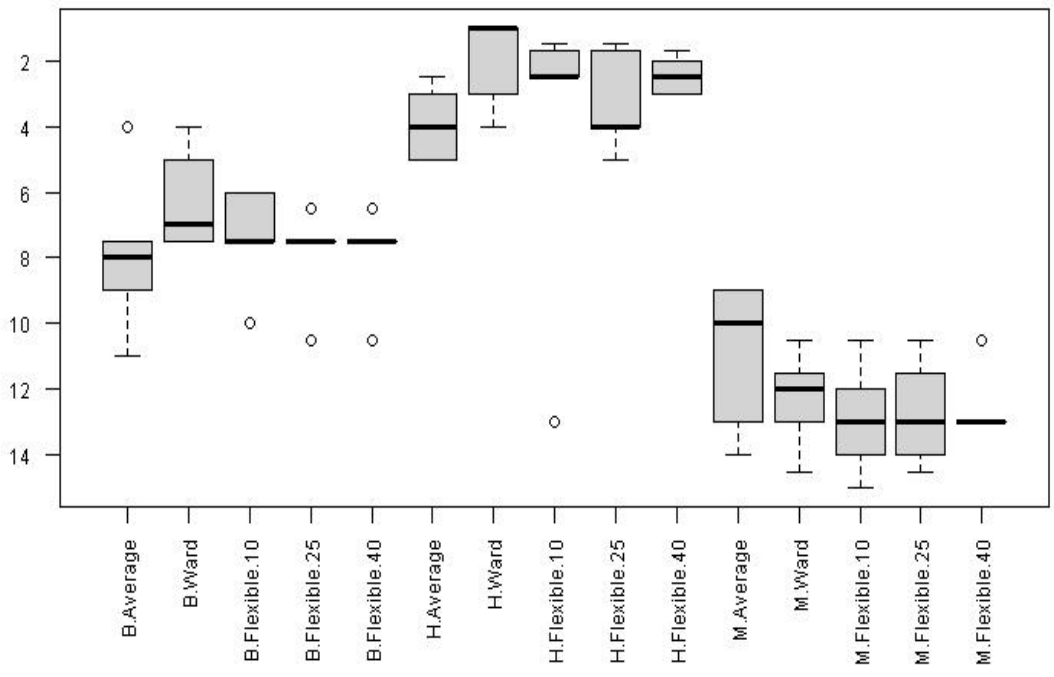
Archive of SID

قرار گرفت. در مجموع ۷۵ روش خوشه‌بندی با شش روش ارزیابی‌کننده، آزمون شد. برای مقایسه الگوریتم‌های مختلف خوشه‌بندی، نتایج رتبه‌بندی شدند و میانگین این رتبه‌ها برای هر الگوریتم محاسبه شد (جدول ۲). مطابق با جدول ۲ در روش‌اندازه‌گیری فاصله بری کورتیس روش خوشه‌بندی وارد و روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱- عملکرد بهتری داشتند. در روش‌اندازه‌گیری هلینگر روش خوشه‌بندی وارد و روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱- برتر از سایر خوشه‌بندی‌ها بودند و در روش اندازه‌گیری منهتن روش خوشه‌بندی میانگین و روش خوشه‌بندی وارد روش بهتری بودند. در هر سه روش اندازه‌گیری فاصله روش خوشه‌بندی وارد رتبه یک تا دو را کسب کردند. در روش اندازه‌گیری فاصله هلینگر، روش میانگین ضعیف‌ترین عملکرد را داشتند و در روش اندازه‌گیری ماتریس فاصله منهتن روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/٤- عملکرد ضعیفی داشته و کمترین رتبه را کسب کرده است. در ماتریس اندازه‌گیری فاصله روش‌اندازه‌گیری فاصله بری کورتیس روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۲۵- و ۰/٤۰-عملکرد بهتری داشتند.

مطابق با شکل ۱ ترکیب روش اندازه‌گیری فاصله هلینگر و خوشه‌بندی وارد بهترین عملکرد را در میان سایر خوشه‌بندی‌ها دارند و ترکیب روش اندازه‌گیری منهتن و خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/٤- بدترین عملکرد را داشتند. روش‌اندازه‌گیری فاصله هلینگر سازگاری بهتری با خوشه‌بندی‌های سلسله مراتبی دارد و روش اندازه‌گیری منهتن، کمترین سازگاری را با خوشه‌بندی‌های سلسله‌مراتبی دارد.

جدول ۲. میانگین رتبه الگوریتم‌های مختلف خوشه‌بندی بر اساس نتایج ارزیابی‌کننده‌ها

خوشه‌بندی	بری کورتیس	هلینگر	منهتن
میانگین	۳	۵	۱
وارد	۱	۱	۲
بتای انعطاف‌پذیر (با مقدار بتای ۰/۱-)	۲	۲	٤،۵
بتای انعطاف‌پذیر (با مقدار بتای ۰/۲۵-)	٤/۵	٤	٤،۵
بتای انعطاف‌پذیر (با مقدار بتای ۰/٤-)	٤/۵	۳	۵



شکل ۱. میانگین رتبه الگوریتم‌های مختلف خوشه‌بندی بر اساس نتایج ارزیابی‌کننده‌ها

www.SID.ir

Archive of SID

محور افقی نشان‌دهنده خوشه‌بندی‌های مختلف (که حروف قبل از اسامی خوشه‌بندی‌ها نماینده روش اندازه‌گیری فاصله است. B=بری کورتیس، H=هلینگر و M=منهتن) و محور عمودی نشان-دهنده رتبه‌ها است (رتبه یک برای بهترین خوشه‌بندی و رتبه ۱۵برای بدترین خوشه‌بندی است)

نتایج ارزیابی‌کننده‌ها نشان داد (جدول ۳) با توجه به نتایج ارزیابی‌کننده شاخص میانگین سیلوئت و آنالیز پارتیشین ترکیب روش ماتریس فاصله هلینگر و روش خوشه‌بندی میانگین بهترین روش خوشه‌بندی است. ارزیابی‌کننده آنالیز گونه‌های معرف و ارزیابی کننده ضریب فی ترکیب روش هلینگر با بتای انعطاف‌پذیر با مقدار بتای ۰/٤- بهترین روش معرفی می‌کند و ارزیابی‌کننده آنالیز گونه‌های معرف خوشه‌ها برای کمینه‌کردن ثبات میانی نیز ترکیب روش هلینگر با بتای انعطاف‌پذیر با مقدار بتای ۰/۱- روش بهتری معرفی می‌کند و بر اساس ارزیابی‌کننده روش جایگشت پاسخ چندگانه ترکیب روش ماتریس فاصله بری‌کورتیس و خوشه‌بندی وارد برتر عملکرد بهتری دارد.

جدول ۳. نتایج رتبه‌بندی الگوریتم‌های مختلف خوشه‌بندی بر اساس ارزیابی‌کننده‌های مختلف در داده‌های هیرکانی

ماتریس فاصله	خوشه‌بندی	میانگین سیلوئت	آنالیز پارتیشین	گونه معرف	گونه‌های معرف خوشه‌ها با ثبات میانی	ضریب فی	روش جایگشت چندگانه
بری کورتیس	میانگین	۳	۳	۱۵	۸	۷	۱٤
	وارد	۵	۸	۷	۷	۸	۱
	بتای انعطاف‌پذیر (با مقدار بتای ۰/۱-)	٦	۷	۵	٦	۹	۵/۵
	بتای انعطاف‌پذیر (با مقدار بتای ۰/۲۵-)	۸/۵	۹/۵	۹/۵	۹/۵	۱۰/۵	۳/۵
	بتای انعطاف‌پذیر (با مقدار بتای ۰/٤-)	۸/۵	۹/۵	۹/۵	۹/۵	۱۰/۵	۳/۵
هلینگر	میانگین	۱	۱	۱۳	۳	۵	۱۲
	وارد	۱۰	٦	۳	۲	٤	۲
	بتای انعطاف‌پذیر (با مقدار بتای ۰/۱-)	۲	۲	۸	۱	۳	۱۰
	بتای انعطاف‌پذیر (با مقدار بتای ۰/۲۵-)	۷	٤	۲	٤	۲	۷
	بتای انعطاف‌پذیر (با مقدار بتای ۰/٤-)	۱۵	۱۱	۱	۵	۱	۵/۵
منهتن	میانگین	٤	۵	۱٤	۱۵	٦	۱۵
	وارد	۱۲/۵	۱٤	۱۱	۱۲	۱٤	۸
	بتای انعطاف‌پذیر (با مقدار بتای ۰/۱-)	۱۱	۱۲	۱۲	۱٤	۱۳	۱۲
	بتای انعطاف‌پذیر (با مقدار بتای ۰/۲۵-)	۱۲/۵	۱۳	۱٤	۱۳	۱۵	۱۱
	بتای انعطاف‌پذیر (با مقدار بتای ۰/٤-)	۱٤	۱۵	٦	۱۱	۱۲	۹

بحث و نتیجه‌گیری

انتخاب روش اندازه‌گیری فاصله در مطالعات متعددی مورد بررسی قرار گرفته است (Legendre & Gallagher, 2001; Roberts, 2017; Legendre & Borcard, 2018). انتخاب روش اندازه‌گیری فاصله، به‌طور معنی‌داری نتایج تجزیه و تحلیل را تحت تاثیر قرار می‌دهد (Podani, 2005; Saraçli *et al*., 2013). انتخاب روش اندازه‌گیری فاصله به عنوان اولین قدم در خوشه‌بندی، بسیار پراهمیت است. در پژوهش حاضر، روش اندازه‌گیری فاصله هلینگر عملکرد بهتری نسبت به ماتریس فاصله منهتن و بری کورتیس دارد و عملکرد روش‌اندازه‌گیری فاصله بری کورتیس بهتر از روش منهتن است. ضریب فاصله هلینگر برای تجزیه و تحلیل داده‌های جوامع گیاهی به دو دلیل توصیه می‌شود: این ضریب فاصله، محدودیت‌های اقلیدوسی را ندارد و همچنین برای تجزیه و تحلیل‌هایی که به فضای متریک نیازمند هستند، مناسب است (Legendre & De Ca´ceres, 2013). در بسیاری از مطالعات روش اندازه‌گیری فاصله بری کورتیس به عنوان روش اندازه‌گیری مناسب برای داده‌های پوشش گیاهی معرفی شده است (Roberts, 2008; Lotter *et al*, 2013)، که این امر به‌دلیل زیاد بودن تعداد مشاهدات در جوامع اکولوژیکی است و همچنین زیاد بودن تعداد صفر در مجموعه داده‌ها است. از این‌رو، توزیع داده‌ها چوله به راست است (Ricotta & Podani, 2017) و تفاوت‌ها را در بین قطعات نمونه با گونه‌های عمومی بهتر نشان می‌دهد (Anderson, 2006). Legendre و De Ca´ceres (۲۰۱۳) همچنین عنوان کردند روش اندازه‌گیری فاصله منهتن برای داده‌های جوامع گیاهی مناسب نیست، زیرا ویژگی‌های همچون غیرمنفی بودن، تقارن و نابرابری مثلثی در روش اندازه‌گیری فاصله همچون روش اقلیدوسی و منهتن

www.SID.ir

Archive of SID

استفاده از آنها را محدود می‌کند، اما روش اندازه‌گیری فاصله بری کورتیس از این محدودیت‌ها پیروی نمی‌کند (Roberts, 2017).

نتایج پژوهش نشان داد در خوشه‌بندی سلسله مراتبی، روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱- و روش وارد موثرتر از سایر روش‌ها می‌باشد. این یافته مطابق با یافته‌های Aho و همکاران (۲۰۰۸) است که بیان کردند بر اساس ارزیابی‌کننده‌های هندسی، روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱- و روش وارد عملکرد بهتری در بین سایر روش‌های خوشه‌بندی دارند. Milligan (1980) اشاره کرد که روش وارد واریانس را حداقل می‌کند و برای داده‌های چندمتغیره با توزیع نرمال کروی موثر است اما در داده‌های خوشه‌بندی‌شده با پراکنش بیضی عملکرد ضعیفی دارد (Everitt *et al*., 2011). به‌علاوه، lotter و همکاران (۲۰۱۳) به نتایج مشابهی دست یافتند. همچنین روش وارد برای خوشه‌بندی داده‌هایی که دارای نقاط پرت هستند، نامناسب است (Milligan, 1980). نتایج نشان داد ترکیب روش خوشه‌بندی ماتریس اندازه‌گیری فاصله هلینگر و بری کورتیس با خوشه‌بندی میانگین عملکرد ضعیفی دارد اما ترکیب روش خوشه‌بندی ماتریس اندازه‌گیری فاصله منهتن و میانگین خوشه‌بندی میانگین سازگاتر از سایر خوشه‌بندی‌های موجود ترکیب شده با این ماتریس فاصله بود. Saraçli و همکاران (۲۰۱۳) روش‌های خوشه‌بندی سلسله مراتبی را مقایسه کردند و نتیجه‌گیری کردند که روش خوشه‌بندی اتصال میانگین در داده‌های کوچک نتیجه بسیار خوبی دارد اما خروجی این روش برای داده‌های بزرگ مناسب نیست و به‌دلیل اینکه روش خوشه‌بندی اتصال میانگین تغییرات زیادی نسبت به سایر روش‌های حفظ ساختار فضایی ایجاد می‌کند روش خوشه‌بندی اتصال میانگین کمتر مورد استفاده قرار می‌گیرد (McCune & Grace, 2002). نتایج ترکیب روش‌های اندازه‌گیری فاصله و روش خوشه‌بندی بتای انعطاف‌پذیر نشان داد که عملکرد خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱- و ۰/۲۵- بهتر از خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/٤- می‌باشد. Belbin و همکاران (1992) تاکید کردند مقادیر بتای منفی و نزدیک صفر کارکرد بهتری دارد. نتایج Tichy و همکاران (۲۰۱۱) نشان دادند خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای با ۰/۱-، ۰/۲۵- و ۰/٤- ثبات طبقه‌بندی بالایی دارند و بتای انعطاف‌پذیر با بتای ۰/۱- بسیار نزدیک به ثبات طبقه‌بندی روش k-means است.

ماتریس تشابه هلینگر در داده‌های هیرکانی مورد مطالعه که با تعداد گونه متوسط تطابق بهتری با روش‌های مختلف خوشه‌بندی انتخاب شده این پژوهش داشت، همچنین دو روش خوشه‌بندی وارد و بتای انعطاف‌پذیر با مقدار بتای ۰/۱- نیز عملکرد بهتری نسبت به سایر روش‌های خوشه‌بندی داشتند. لازم به ذکر است محققان باید نسبت به این موضوع آگاه باشند که تصمیات آنها برای انتخاب روش اندازه‌گیری فاصله و الگوریتم خوشه‌بندی به‌طور معنی‌داری نتایج و تفسیر آنان را تحت تاثیر قرار می‌دهند. در نتیجه در انتخاب روش‌های مناسب با توجه به ماهیت داده مورد استفاده‌شان باید دقت لازم را داشته باشند. همچنین برای بررسی نتایج روش خوشه‌بندی بهتر از چندین شاخص استفاده کنند، زیرا هر شاخص معایب مخصوص به خود را دارد و پیچیدگی داده‌ها نتایج را تحت تاثیر قرار می‌دهد.

منابع

اسحاقی‌راد، ج.، زاهدی‌امیری، ق. و متاجی، ا.(۱۳۸۷) تعیین تعداد بهینه گروه‌های اکولوژیک در طبقه‌بندی پوشش گیاهی، مطالعه موردی جنگل آموزشی و پژوهشی خیرودکنار نوشهر. فصلنامه علمی- پژوهشی تحقیقات جنگل و صنوبر ایران، ۱٦(۳): ٤٥٥-٤٦٦.

اسحاقی‌راد، ج.، زاهدی‌امیری، ق.، مروی‌مهاجر، م.ر.، اسدی، م. و متاجی، ا. (۱۳۸۸) ارتباط بین پوشش کف جنگل و خصوصیات فیزیکی و شیمیایی خاک در جوامع راش. فصلنامه علمی- پژوهشی تحقیقات جنگل و صنوبر ایران، ۱۷(۲): ۱۷٤–۱۸۷.

اسماعیل‌زاده، ا. و اسدی، ح. (۱۳۹۳) معرفی مدل مجموع شاخص تعلقهٔ فی(TPFIM) در تجزیه و تحلیل گروه‌های اکولوژیک گیاهی. مجلهٔ جنگل ایران، انجمن جنگلبانی ایران، ٦(۲): ۲۱۵-۲۳۲.

Aho, K., Roberts, D.W. and Weaver, T. (2008) Using geometric and non-geometric internal evaluators to compare eight vegetation classification results. Journal of Vegetation Science, 19(4): 549–562.

Anderson, M.J. (2006) Distance-based tests for homogeneity of multivariate dispersions. Biometrics, 62(1): 245-253.

www.SID.ir

Archive of SID

cluster number and abundance transformation for obtaining effective vegetation classifications. Journal of Vegetation Science, 29(2): 336-347.

Legendre, P., and De Ca´ceres, M. (2013) Beta diversity as the varianceof community data: dissimilarity coefficients and partitioning. Ecology Letter 16: 951–963.

Lengyel, A., Roberts, D.W. and Botta-Dukat, Z. (2019) Comparison of silhouette-based reallocation methods for vegetation classification. bioRxiv, p.630384.

Lotter, M.C., Mucina, L. and Witkowski, E.T.F. (2013) The classification conundrum: Species fidelity as leading criterion in search of a rigorous method to classify a complex forestdata set. Community Ecology, 14(1): 121–132.

Lyons, M.B., Keith, D.A., Warton, D.I., Somerville, M. and Kingsford, R.T. (2016) Model-based assessment of ecological community classifications. Journal of Vegetation Science, 27(4): 704-715.

Mielke, P.W. and Berry, K.J. (2007) Permutation methods: A distance function approach (2nd edition). Springer, New York, 353p.

Milligan, G.W. (1980) An examination of the effect of six types of error perturbation on fifteen clustering algorithms. Psychometrika, 45(3): 325–342.

McKenna, Jr.J.E. (2003) An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. Environmental Modelling and Software, 18(3): 205-220.

McCune, B., Grace, J. B. (2002) Analysis of ecological communities. MJM Software Design, Glenedon Beach,Oregon, USA

Orlóci, L. (1974) On information flow in ordination. Vegetatio, 29(11): 16-16.

Peet, R.K. and Roberts, D.W. (2013) Classification of natural and seminatural vegetation. In van der Maarel, E. and Franklin, J. (eds.) Vegetation ecology (2nd edition), Wiley-Blackwell, Oxford, pp: 26–62.

Podani, J. (2005) Multivariate exploratory analysis of ordinal data in ecology: pitfalls, problems and solutions. Journal of Vegetation Science, 16(5): 497-510.

Roleček, J., Tichý, L., Zelený, D. and Chytrý, M. (2009). Modified TWINSPAN classification in which the hierarchy respects cluster heterogeneity. Journal Vegetation Science, 20(4): 596-602.

Rousseeuw, P.J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20: 53–65.

Roberts, D.W. (2017) Distance, dissimilarity, and mean–variance ratios in ordination. Methods in Ecology and Evolution, 8(11): 1398-1407.

Roberts, DW. (2015) Vegetation classification by two new iterative reallocation optimization algorithms. Plant Ecology, 216(5):741-58.

Arora, P. and Varshney, S. (2016) Analysis of k-means and k-medoids algorithm for big data. Procedia Computer Science, 78: 507-512.

Bertrand, P. and Bel Mufti, G. (2006) Loevinger's measures of rule quality for assessing cluster stability. Computational statistics and data analysis, 50(4): 992-1015.

Bray, R.J. and Curtis, J.T. (1957) An ordination of the up and forest communities of southern Wisconsin. Ecological Monographs, 27: 325–349.

Belbin, L., Faith, D.P. and Milligan, G.W. (1992) A comparison of two approaches to beta-flexible clustering. Multivariate Behavioral Research, 27(3): 417-433.

Craw, S. (2011) Manhattan distance. In: C. Sammut and G.I. Webb (Eds.), Encyclopedia of machine learning, Boston, MA: Springer.

De Cáceres, M., Chytrý, M., Agrillo, E. and et al. (2015) A comparative framework for broad- scale plot- based vegetation classification. Applied Vegetation Science, 18(4): 60-543.

Dufreˆne, M. and Legendre, P. (1997) Species assemblages and indicator species: The need for a flexible asymmetric approach. Ecological Monographs, 67(3): 345–367.

Everitt, B.S., Landau, S., Leese, M. and Stahl, D. (2011) Cluster analysis (5th edition). ChiChester: Wiley and Sons, 290p.

Fernández, A. and Gómez, S. (2019) Versatile linkage: A family of space-conserving strategies for agglomerative hierarchical clustering. Journal of Classification, 37: 584–597

Hämäläinen, J. Jauhiainen, S. and Kärkkäinen, T. (2017) Comparison of internal clustering validation indices for prototype-based clustering. Algorithms, 10(3):105-105.

Khanalizadeh, A., EshaghiRad, J., Amiri, G.Z., Zare, H., Rammer, W. and Lexer, M.J. (2020) Assessing selected microhabitat types on living trees in Oriental beech (*Fagus orientalis* L.) dominated forests in Iran. Annals of Forest Science, 77(3): 1-13.

Lance, G.N. and Williams, W.T. (1967) A general theory of classificatory sorting strategies: 1. Hierarchical systems. The Computer Journal, 9(4): 373–380.

Legendre, P. and Borcard, D. (2018) Box–Cox-chord transformations for community composition data prior to beta diversity analysis. Ecography, 41(11): 1820-1824.

Legendre, P. and Gallagher E.D. (2001) Ecologically meaningful transformations for ordination of species data. Oecologia, 29(2): 271-280.

Legendre, P. and Legendre, L. (1998) Numerical ecology (2nd), English ed., Elsevier, Amsterdam, 583p.

Lengyel, A. and Podani, J. (2015) Assessing the relative importance of methodological decisions in classifications of vegetation data. Journal of Vegetation Science, 26(4): 804–815.

Lengyel, A., Landucci, F., Mucina, L., Tsakalos, J.L. and Botta-Dukát, Z. (2018) Joint optimization of

www.SID.ir

Archive of SID

Schmidtlein, S., Tichy, L., Feilhauer, H. and Faude, U. (2010) A brute-force approach to vegetation classification. Journal of Vegetation Science, 21(6): 1162–1171.

Tichý, L., Chytrý, M. and Smarda, P. (2011) Evaluating the stability of the classification of community data. Ecography, 34(5): 807-813.

Tichy, L. and Chytry, M. (2006) Statistical determination of diagnostic species for site groups of unequal size. Journal of Vegetation Science, 17(6): 809-818.

Wang, P., Liu, Q., Xu, G. and Wang, K. (2019) A Three-Way clustering method based on ensemble strategy and three-way decision. Information Journal, 10(2): 59-59.

Zaki, M.J. and Meira, W.Jr. (2014) Data mining and analysis: Fundamental concepts and algorithms. Cambridge University Press: New York, 586p.

Ricotta, C. and Podani, J. (2017) On some properties of the Bray Curtis dissimilarity and their ecological meaning. Ecological Complexity, 31: 201-205.

R Core Team, (2019) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Roberts, D.W. (2015) Vegetation classification by two new iterative reallocation optimization algorithms. Plant Ecology, 216(5): 741–758.

Roberts, DW. (2008). Statistical analysis of multidimensional fuzzy set ordinations. Ecology, 89:1246–1260

Rodriguez, M.Z., Comin, C.H., Casanova, D., Bruno, O.M., Amancio, D.R., Costa, L.D.F. and Rodrigues, F.A. (2019) Clustering algorithms: A comparative approach. PloS one,14(1): e0210236.

Saraçli, S. Doğan, N. and Doğan, İ. (2013) Comparison of hierarchical cluster analysis methods by cophenetic correlation. Journal of Inequalities and Applications, 2013(1): 203-203.

www.SID.ir

Archive of SID

The effect of distance measurement methods on the classification of ecological groups in Hyrcanian forests

Naghmeh Pakgohar[1], Javad Eshaghi Rad[2]*, Gholamhossein Gholami[3], Ahmad Alijanpour[2] and David W. Roberts[4]

1) Ph.D. of Forestry, Department of Forestry, Faculty of Natural Resources, University of Urmia, Urmia, Iran.
2) Associate Professor, Department of Forestry, Faculty of Natural Resources, University of Urmia, Urmia, Iran. *Corresponding Author Email Address: j.eshagh@urmia.ac.ir
3) Assistant Professor, Department of mathematics, Faculty of Science, University of Urmia, Urmia, Iran.
4) Professor, Department of Ecology, Montana State University, Bozeman MT, US.

Date of Submission: 2021/06/26 Date of Acceptance: 2021/09/18

Abstract

Nowadays, the application of clustering methods is widely increased, although choosing the right method due to the existence of different method and effective factors is difficult. The present study aimed to compare the results of widely used clustering algorithms and to determine the most effective methods according to the different evaluators and evaluate the effective distance measurement method for clustering algorithms. The data of Hyrcanian beech forests were examined in an area protected by the department of natural resources of Nowshahr. Random-systematic sampling method with regular grid of 100×200 m was used for determining the center of sample plots; 100-m^2 (10×10 m) sample plots had been used to check the shrub species and 400-m^2 (20×20 m) to check the herbaceous species. A total of 120 sample plots were measured. The abundance and coverage of tree, shrub and herbaceous species were estimated based on Braun-Blanquette scale. Three distance methods of measuring distance Bray Curtis, Hellinger and Manhattan were used and five clustering methods (Average method clustering methods, Ward method, flexible beta method with beta values of -0.1, -0.25, -0.4) with six evaluation indicators (silhouette evaluation criterion, PARATNA criterion, Indval criterion, ISAMIC criterion, MRPP criterion and Phi correlation coefficient) were examined. Different clustering algorithms were arranged from best to worst for each dataset. The comparison analysis revealed that Ward's and flexible-beta with beta value of -0.1 had the best performance. The present findings illustrated that Hellinger distance measurement method is better in homogeneous data than other distance measurement methods.

Keywords: Classification, Distance measures, Hierarchical clustering, Hyrcanian forest.

www.SID.ir