تحقیقات منابع آب ایران
Iran-Water Resources Research

سال هفتم، شماره ۱، زمستان و بهار ۹۰-۱۳۸۹
Volume 7, No. 1, Winter & Spring 2011 (IR-WRR)
۳۸-۴۶

خوشه‌بندی مکانی شبکه‌های آبیاری با استفاده از روش کلاسیک K-Means
(مطالعه موردی شبکه آبیاری قزوین)

محمدجواد منعم[۱*] و سیدمهدی هاشمی [۲]

چکیده

بهبود عملکرد شبکه‌های آبیاری از راهکارهای اساسی صرفه‌جویی در منابع آبی می‌باشد. اولین گام برای بهبود عملکرد شبکه‌ها، ارزیابی وضع موجود و سپس ارائه راهکار جهت رفع مشکلات می‌باشد. یک گام موثر و کاربردی در ارزیابی و بهبود عملکرد، استخراج مناطق همگن شبکه کانال‌ها بر اساس خصوصیات فیزیکی و فنی می‌باشد. هدف اصلی از این تحقیق، پهنه‌بندی مکانی و کاربرد یک روش کمی جهت استخراج مناطق همگن فیزیکی شبکه‌های آبیاری و نمایش قابلیت این روش در یک شبکه آبیاری واقعی می‌باشد. تکنیک مورد استفاده جهت پهنه‌بندی مکانی، روش خوشه‌بندی کلاسیک K-Means است. داده‌های مورد استفاده به عنوان ورودی مدل خوشه‌بندی به‌صورت یک ماتریس ۵×۱۶۲ بعدی است که همان خصوصیات فیزیکی و فنی کانال‌های انتقال شبکه آبیاری قزوین می‌باشند. بر اساس شاخص صحت‌سنجی خوشه‌بندی دیویس- بولدین (DB)، تعداد بهینه خوشه‌ها برابر ۱۰ خوشه بدست آمد. هر کدام از خوشه‌های بدست آمده معرف یک ناحیه همگن در سطح شبکه می‌باشد که مدیران شبکه را قادر خواهد ساخت دامنه تصمیم‌گیری‌های خود را از محدوده وسیعی در ابعاد یک شبکه به محدوده کوچک‌تری در ابعاد چند منطقه همگن محدود کاهش دهند. این امر سبب سهولت مدیریت و ارزیابی و تصمیم‌گیری در سطح مناطق همگن و نیز صرفه‌جوئی در زمان و هزینه مدیریت خواهد شد.

کلمات کلیدی: شبکه‌های آبیاری، خوشه‌بندی K-Means، پهنه‌بندی مکانی.

تاریخ دریافت مقاله: ۱۶ آذر ۱۳۸۷
تاریخ پذیرش مقاله: ۲۷ دی ۱۳۸۹

۱- دانشیار، گروه سازه‌های آبی، دانشگاه تربیت مدرس، تهران، ایران.
۲- دانشجوی دکتری، گروه سازه‌های آبی، دانشگاه تربیت مدرس، تهران، ایران.
*- نویسنده مسئول

Spatial Clustering of Irrigation Networks Using K-Means Method
(Case Study of Ghazvin Irrigation Network)

M. J. Monem [1] and S. M. Hashemi [2]*

Abstract

Improving the performance of water conveyance networks is one of the key issues in saving limited water resources. The first step for this improvement is performance evaluation and then presenting the solutions. One of the practical and efficient approches for performance improvement is to extract the homogenous area out of the irrigation network based on the physical and technical features. The main idea behind this research is to present a quantitative benchmark for exploring homogenous areas with similar physical attributes and present the abilites of this method for a real case study. K-Means clustering algorithm, is applied to spatial clustering of irrigation networks based on physical attributes. Data was arranged based on the "objects" and the "features" in the matrix language. Ghazvin irrigation network data was used to form the input matrix. This matrix consisted of 162 rows and 5 columns. Using Davies and Bouldin (DB) index as the cluster validity index, it has been shown that the optimum number of clusters is 10. Each cluster represented a homogenous area in the irrigation network district. Clustering reduces the dimension of assessments from a large extended irrigation district to a limited number of homogeneous regions and provide a context for better and easier decision making, performance evaluation, and allocation of facilities and budget to different regions.

Keywords: Irrigation networks, K-Means clustering, Spatial regionalization.

Received: December 6, 2008
Accepted: January 17, 2011

1- Associate Professor, Water Structure Dep., Tarbiat Modares University, Tehran, Iran. Email: monem_mj@modares.ac.ir
2- PhD Candidate, Water Structure Dep., Tarbiat Modares University, Tehran, Iran. Email: hashemy@modares.ac.ir
*- Corresponding Author

۱- مقدمه

امروزه کشت آبی نقش مهمی را در تولید مواد غذائی در دنیا دارد. به طوری‌که در حدود ۴۰ درصد از تولید مواد غذائی مربوط به نواحی فاریاب می‌باشد (Malano and Burton, 2001). بررسی‌های انجام شده نشان می‌دهد که عملکرد اکثر شبکه‌های آبیاری و زهکشی چه از نظر فرآیندی مثل کفایت، راندمان، عدالت و پایداری و چه از لحاظ خروجی، منظور تولید محصول، مطلوب نبوده است. بنابراین بهبود عملکرد شبکه‌های آبیاری یک ضرورت اجتناب ناپذیر است و اولین گام برای بهبود عملکرد شبکه‌ها، ارزیابی وضع موجود و نهایتاً ارائه راهکار جهت بهبود عملکرد می‌باشد (خلخالی و همکاران، ۱۳۸۷). بر این اساس، روش‌های متعدد کمی و کیفی برای ارزیابی عملکرد شبکه‌های آبیاری توسط محققین مختلف به‌کار گرفته شده ولی عدم مطلوبیت اغلب آن‌ها این است که منجر به بهبود قابل توجه عملکرد شبکه‌ها نشده است (حیدریان و همکاران، ۱۳۸۲؛ منعم و همکاران، ۱۳۸۱ ; Bruscoli and Bresci, 2001; Burt, 2001; Oad and Mc Cornick, 1989). بخشی از این عدم مطلوبیت ناشی از عدم توجه به تفاوت‌های مکانی مناطق مختلف شبکه از نظر فیزیکی و کارکردی می‌باشد. عملکرد سیستم و راه‌کارهای بهبود در مناطق مختلف یک شبکه به دلایل زیر کاملاً متفاوت هستند :

۱- تفاوت موقعیت قرارگیری هر قسمت شبکه نسبت به منبع آب و تفاوت در زمان جابه‌جائی جریان و تاخیر در تحویل آب .
۲- تفاوت وضعیت مسئولین بهره‌برداری در قسمت‌های مختلف شبکه (از لحاظ دانش، تجربه کار و میزان دقت آنها).
۳- تفاوت وضعیت کانال‌ها و سازه‌ها از نظر کیفیت کار و عملکرد، واسنجی سازه و تعمیرات و نگهداری.
۴- امکان وجود مدیریت چندگانه حاکم بر قسمت‌های مختلف شبکه (دولتی، تعاونی و خصوصی).
۵- تفاوت وضعیت بافت خاک، آب زیرزمینی، زهکشی و نیاز آبی گیاهان مختلف در مناطق مختلف به‌خصوص در شبکه‌های بزرگ.
۶- تفاوت دانش، تجربه و مهارت کشاورزان در مناطق مختلف، عادات و شرایط اقتصادی، اجتماعی و نوع محصول.

با توجه به عوامل مذکور استفاده از روش‌های مرسوم برای ارزیابی عملکرد و ارائه راه‌کارهای بهبود شبکه‌ها مستلزم تجزیه و تحلیل کامل مسائل خاص و جزئی در کلیه قسمت‌های شبکه می‌باشد، که امری پرهزینه و زمان‌بر خواهد بود. چه بسا پس از صرف زمان قابل توجه برای شناسائی و رفع این کاستی‌ها، صورت مساله تغییر کرده و راه‌کارهای بهبود دیگری را طلب نماید. با توجه به تفاوت‌های موجود در بخش‌های مختلف شبکه می‌توان با شناسائی مناطق همگن در شبکه‌ها، مطالعات ارزیابی و ارائه راه‌کارهای بهبود را در مناطق همگن متمرکز نمود و برخی از محدودیت‌های روش‌های مرسوم را کاهش داد. با استفاده از روش خوشه‌بندی، اجزائی از شبکه که براساس هر کدام از شاخص‌های مورد نظر ارزیابی دارای عملکرد مشابه هستند، در کلاس‌های متفاوت دسته‌بندی می‌شوند. با این کار ارائه راه‌کار موردی برای تک‌تک اجزای شبکه تبدیل به ارائه راه‌کار برای دسته‌ای از اجزای شبکه خواهد شد که دارای رفتار یا عملکرد مشابهی می‌باشند. به عبارت دیگر روش خوشه‌بندی قادر است اقدام به پهنه‌بندی مکانی شبکه آبیاری، براساس عوامل مورد نظر مدیران شبکه نماید.

روش خوشه‌بندی یکی از زیر مجموعه‌های علم داده‌کاوی است که هدف آن اکتشاف و پردازش پایگاه‌های داده‌ای، به منظور استخراج دانش از آنها است. خوشه‌بندی یک روش یادگیری غیرنظارتی[1] برای دسته‌بندی داده‌ها براساس مشابهت‌های آن‌ها می‌باشد. این تکنیک ابزاری توانمند جهت استخراج ساختار اصلی نهفته در مجموعه داده‌ها معرفی شده است (Valente and Pedrycz , 2007). روش‌های غیرنظارتی، مانند خوشه‌بندی، از شناسنده‌های[2] از پیش تعریف شده جهت انجام عملیات دسته‌بندی استفاده نمی‌کنند.

تنها تحقیق انجام شده در زمینه خوشه‌بندی عملکرد شبکه‌های آبیاری مربوط به استفاده از روش رتبه‌بندی و خوشه‌بندی بر روی داده‌های زمانی شبکه‌های آبیاری گلبورن[3] استرالیا و شی‌جین[4] در کشور چین می‌باشد. هدف از این مطالعه، توسعه و کاربرد تئوری فازی جهت رتبه‌گذاری و خوشه‌بندی عملکرد شبکه‌های آبیاری در یک دوره زمانی خاص بوده است. در این تحقیق آمار سالانه پارامترهائی مانند مساحت زمین‌های آبیاری شده، راندمان سیستم توزیع، عملکرد آبیاری مزرعه و میزان محصول سالانه جهت خوشه‌بندی زمانی عملکرد شبکه، با روش سلسله‌مراتبی فازی، استفاده شده است (Malano and Gao, 1992). در این تحقیق از خوشه‌بندی به عنوان یک ابزار مدیریتی، جهت تعیین سال‌های با عملکرد مشابه استفاده شده است. استفاده از روش‌های خوشه‌بندی به منظور تحلیل مکانی، امکان شناسائی مناطق همگن در سطح شبکه و به دنبال آن برنامه‌ریزی ارزیابی و بهبود عملکرد را متناسب با آن فراهم می‌نماید.

یکی دیگر از مشکلات پیش‌روی تحقیقات انجام شده در شبکه‌های آبیاری گستردگی زیاد و پراکندگی اطلاعات موجود در سطح شبکه‌ها

می‌باشد. این پراکندگی داده‌ها عملاً تجزیه تحلیل روش‌های ارزیابی کلاسیک را محدود، زمان‌بر و هزینه‌بر کرده و تحلیل نتایج را با مشکل روبه‌رو می‌سازد. خوشه‌بندی به عنوان یک زیرمجموعه از علم داده‌کاوی[5]، که اخیراً مورد توجه محققان علوم مختلف قرار گرفته است، ابزاری توانمند برای مدیریت داده‌ها و افزایش قدرت و بهبود تصمیم‌گیری مدیران شبکه خواهد شد.

هدف از این تحقیق بررسی قابلیت روش خوشه‌بندی K-Means در استخراج مناطق همگن فیزیکی در سطح شبکه‌های انتقال آبیاری است. برای این منظور از خصوصیات فیزیکی و فنی شبکه‌های آبیاری قزوین استفاده شده است. نتایج خوشه‌بندی به‌صورت نقشه‌های پهنه‌بندی از محدوده مورد مطالعه ارائه می‌گردد تا مدیران و تصمیم‌گیران بتوانند از این نتایج به عنوان یک ابزار پیش‌پردازش جهت به‌کارگیری روش‌های ارزیابی عملکرد و نیز در برنامه‌ریزی عملیات بهره‌برداری و نگهداری شبکه‌های آبیاری استفاده کنند.

۲- روش تحقیق و آزمایش

۲-۱- خوشه‌بندی

در این تحقیق از روش خوشه‌بندی برای دسته‌بندی اجزای مشابه شبکه‌های آبیاری استفاده شده است. ورودی مدل خوشه‌بندی ماتریسی از داده‌ها می‌باشد که این داده‌ها می‌توانند کمی (عددی)، کیفی و یا ترکیب این دو باشند. داده‌های کمی یا عددی معمولاً از مشاهدات فرآیندهای فیزیکی بدست می‌آیند. هر مشاهده شامل n متغیر اندازه‌گیری شده می‌باشد که در یک بردار n بعدی X_k - $X_k=[x_{k1},x_{k2},...,x_{kn}]^T, X_k\in R^n,$ قرار می‌گیرد. مجموعه‌ای از N مشاهده به صورت $X=\{x_k|k=1,2,...,N\}$ نوشته می‌شود که بیانگر یک ماتریس $N\times n$ به صورت رابطه (۱) می‌باشد:

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nn} \end{bmatrix} \qquad (۱)$$

در تعاریف علم تشخیص الگو[6]، به ردیف‌های ماتریس X، الگوها[7] یا اشیا[8] و به ستون‌های این ماتریس ویژگی‌ها[9] یا مشخصات[10] گفته می‌شود. ماتریس X را ماتریس الگو یا به صورت ساده‌تر ماتریس داده نام‌گذاری کرده‌اند. خروجی مدل خوشه بندی، تعدادی خوشه است. خوشه به گروهی از اشیاء اطلاق می‌شود که نسبت به مابقی اشیاء در گروه‌های دیگر مشابهت بیشتری با یکدیگر دارند. داده‌ها می‌توانند خوشه‌هایی با اشکال هندسی، اندازه‌ها و چگالی‌های متفاوت ایجاد نمایند (Han and Kamber, 2006). منظور از واژه مشابهت میزان تشابه از لحاظ ریاضی است. در واقع تشابه بین دو شی، میزان فاصله بین این دو شی می‌باشد. در بیشتر موارد مقدار فاصله اقلیدسی بین دو شیء به عنوان معیار مشابهت استفاده می‌شود (Van der Heijden et al., 2004).

اولین گام پس از ایجاد ماتریس داده‌ها، نرمال‌سازی داده‌های این ماتریس می‌باشد. تفاوت مقیاس‌ها در متغیرهای مختلف، نتایج خوشه‌بندی را تحت تاثیر قرار خواهد داد. به این منظور باید همه داده‌ها را هم‌مقیاس کرده و آنها را نرمال نمود. در این تحقیق از روش نرمال‌سازی کمینه- بیشینه استفاده شد تا داده‌ها در محدوده صفر تا یک قرار گیرند (Johnson and Wichern, 1999). پس از ایجاد ماتریس نرمال شده داده‌ها، ماتریس مشابهت[11] از ماتریس نرمال شده داده‌ها ساخته می‌شود. درایه‌های ماتریس مشابهت $N\times C$ بعدی (N برابر با تعداد اشیاء و C برابر با تعداد خوشه‌ها) بیانگر میزان عضویت هر شیء به هر خوشه می‌باشد که این مقدار در مجموعه کلاسیک، مانند روش استفاده شده در این تحقیق، صفر یا یک و در مجموعه‌های فازی یک عدد بین صفر و یک خواهد بود.

تابع هدف روش K-Means برای خوشه‌بندی مجموعه اشیاء X به تعداد C خوشه، به‌صورت رابطه (۲) بیان می‌شود که در فرایند خوشه‌بندی این تابع هدف حداقل می‌شود:

$$J(X;V)=\sum_{i=1}^{C}\sum_{k\in i}\left\|x_k^{(i)}-v_i\right\|^2 \qquad (۲)$$

که در این رابطه $\left\|x_k^{(i)}-v_i\right\|^2=(x_k^{(i)}-v_i)^T.(x_k^{(i)}-v_i)$، $x_k^{(i)}$ فاصله شی kام از خوشهٔ i از مبدا مختصات می‌باشد و v_i فاصله مرکز جرم خوشهٔ i از مبدا مختصات به شمار می‌رود. مرکز خوشه‌ها از رابطه (۳) به صورت زیر به دست می‌آید:

$$v_i=\frac{\sum_{k=1}^{N_i}x_k}{N_i}, \qquad x_k\in A_i \qquad (۳)$$

که در آن N_i تعداد اعضای مجموعه A_i و A_i مجموعه اعضای خوشه iام می‌باشد.

روش K-Means به دلیل سهولت اجرا کاربرد گسترده‌ای در مسائل خوشه‌بندی یافته است. میزان پیچیدگی این روش تابع تعداد اشیاء مجموعه X می‌باشد. مشکل اصلی این روش، حساس بودن آن به چگونگی انتخاب اولیه مراکز خوشه‌ها است. در صورتی که مراکز خوشه‌های ابتدایی نامناسب انتخاب شوند تابع هدف به جای نقطه حداقل سراسری به سمت یک حداقل موضعی همگرا می‌شود. برای

رفع این مشکل در این تحقیق مدل خوشه‌بندی برای یک تعداد خوشه خاص به دفعات متفاوت اجرا گردید و در هر بار اجرای مدل مراکز خوشه‌ها به صورت تصادفی در فضای مختصات انتخاب گردیدند تا اطمینان حاصل شود تابع هدف به حداقل سراسری رسیده است. روند تغییرات مقادیر شاخص صحت‌سنجی خوشه‌بندی در برابر افزایش تعداد خوشه‌ها نشان دهنده آن است که مقادیر تابع هدف می‌تواند به مقدار حداقل سراسری خود همگرا شود یا خیر. در صورتی که تغییر محسوسی در مقادیر شاخص صحت‌سنجی از یک تعداد خوشه به بعد رخ ندهد، این اطمینان حاصل می‌شود که تابع هدف خوشه‌بندی به مقدار حداقل سراسری نزدیک شده است و نتایج خوشه‌بندی قابل اعتماد می‌باشد (Theodoridis and Koutroumbas, 2003).

۲-۲- شاخص صحت‌سنجی خوشه‌بندی K-Means

برای تعیین تعداد بهینه خوشه‌ها از شاخص صحت‌سنجی دیویس- بولدین (DB) از جعبه‌افزار Pattern recognition نرم‌افزار Matlab استفاده شده است. از نگاه این شاخص، خوشه‌بندی خوب زمانی رخ می‌دهد که خوشه‌های ایجاد شده کروی شکل، متراکم و قابل تفکیک از هم باشند. با فرض تفکیک n داده به g خوشه، مقدار کلی شاخص DB برای هر بار عمل خوشه‌بندی از رابطه (۴) بدست می‌آید. مقدار این شاخص برای دو خوشه i و k از رابطه (۶) بدست می‌آید.

$$I_{DB} = \frac{1}{g}\sum_{i=1}^{g} R_j \qquad (۴)$$

$$R_j = \max_{k=1,2,...,g;k\neq j} R_{jk} \qquad (۵)$$

$$R_{jk} = \frac{\sigma_j + \sigma_k}{\left\|\mu_j - \mu_k\right\|},\quad j,k = 1,2,...,g;\quad k \neq j \qquad (۶)$$

در این رابطه μ_j برابر میانگین فواصل تمامی اشیا واقع شده در خوشه j و σ_j میزان پراکندگی درونی خوشه j می‌باشد که از رابطه (۷) قابل محاسبه است. مقدار R_{jk} زمانی به حداقل مقدار خود می‌رسد که فاصله بین دو خوشه از یکدیگر زیاد و پراکندگی درونی خوشه‌ها به کمترین مقدار خود برسد.

$$\sigma_j = \sqrt{\frac{1}{n_j}\sum_{x_i \in C_j}\left\|x_i - \mu_j\right\|^2} \qquad (۷)$$

در این رابطه C_j مجموعه داده‌های متعلق به خوشه j و n_j تعداد این داده‌ها می‌باشد.

شاخص DB (دیویس و بولدین) برابر با میانگین مقادیر R_j تعریف می‌شود که R_j برابر حداکثر فاصله R_{jk} بدست آمده از مقایسه خوشه j با سایر خوشه‌ها می‌باشد (Davies and Bouldin, 1979). جهت تعیین محدوده اجرای مدل در روش خوشه‌بندی توصیه شده است مدل خوشه‌بندی را به تعداد m مرتبه که $2 \le m \le \sqrt{N}$، که N برابر تعداد داده‌ها یا همان ردیف‌های ماتریس داده‌ها است، اجرا کرده و در هر بار ماتریس مقادیر عضویت و معیارهای صحت‌سنجی محاسبه شود (Kim, et al., 2004).

تعداد بهینه خوشه‌ها در نقطه‌ای که شاخص DB حداقل شود یا اختلاف شیب بین خط قبل و بعد از آن نقطه تفاوت معنی‌داری داشته باشد، به‌دست می‌آید (Theodoridis and Koutroumbas, 2003).

۲-۳- شبکه آبیاری مورد مطالعه در تحقیق

هدف اصلی از این تحقیق معرفی و تشریح روشی جدید جهت خوشه‌بندی مکانی شبکه‌های آبیاری به منظور استخراج مناطق همگن در سطح شبکه از لحاظ خصوصیات در نظر گرفته می‌باشد. برای نمایش کاربرد این روش بر روی یک شبکه آبیاری واقعی، شبکه آبیاری قزوین انتخاب گردید. کانال‌های انتقال شبکه قزوین شامل یک کانال اصلی، ۱۲ کانال درجه دو و کانال‌های فرعی خروجی از کانال‌های درجه دو می‌باشد که آب را به واحدهای درجه ۳ می‌رساند. کانال‌های درجه دو و کانال‌های فرعی خروجی از آن‌ها بر اساس تغییر در ظرفیت طراحی آن در طول مسیر به تعدادی بازه تقسیم شده که بر این اساس ۱۶۲ بازه کانال به دست آمده است. هر بازه به قسمتی از کانال اطلاق می‌شود که دارای ظرفیت یکسان می‌باشد. جهت انجام خوشه‌بندی، داده‌های فنی و فیزیکی بازه‌ها از پایگاه داده‌های شبکه آبیاری قزوین در اختیار محققین قرار گرفت. خصوصیات در نظر گرفته شده برای هر بازه کانال که تشکیل دهنده ستون‌های ماتریس داده‌ها می‌باشد، در این تحقیق عبارتند از : طول و ظرفیت بازه، تعداد آبگیر و تعداد سازه‌های انتقال موجود در بازه و سطح آب‌خور تحت پوشش آن بازه از کانال، مقادیرحداکثر و نیز ضریب تغییرات خصوصیات در نظر گرفته شده که در جدول ۱ نشان داده شده است. بنابراین ماتریس ورودی مدل خوشه‌بندی در این تحقیق، یک ماتریس ۱۶۲×۵ بعدی می‌باشد که اشیاء (ردیف‌های) آن بیانگر هر کدام از بازه‌های کانال و ویژگی‌ها (ستون‌های) آن بیانگر خصوصیات فیزیکی بازه‌ها می‌باشد. خروجی خوشه‌بندی در این حالت بیانگر خوشه‌هایی است که بازه‌های قرار گرفته در آن دارای خصوصیات فیزیکی مشابه با هم می‌باشند.

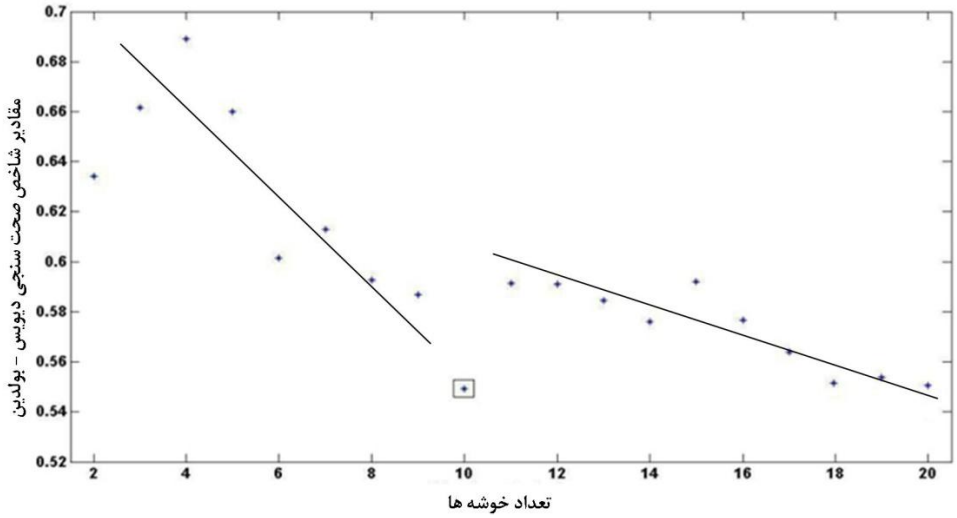
جدول ۱- حدود تغییرات خصوصیات فیزیکی کانال‌های انتقال شبکه آبیاری قزوین

	طول (m)	ظرفیت (m^3/s)	سطح آب خور تحت پوشش (m^2)	تعداد آبگیر	تعداد سازه‌های انتقال
مقدار حداکثر	۶۶۹۰/۱	۷/۳۵	۴۹۱۲۸۳۱۷/۱	۱۳	۵۶
مقدار حداقل	۱۲۰	۰/۱	۳۰۵۳/۸۳	۱	۰
ضریب تغییرات	۰/۶۴	۰/۸۷	۰/۷۹	۰/۶۶	۰/۹۶

۳- نتایج و تحلیل نتایج

جهت انجام خوشه‌بندی کلاسیک K-Means، از جعبه‌افزار تشخیص الگو در نرم‌افزار Matlab استفاده گردید. پس از نرمال‌سازی ماتریس داده‌ها، عمل خوشه‌بندی به تعداد ۲ تا ۲۰ خوشه با روش خوشه‌بندی کلاسیک K-Means انجام شد و برای هر بار اجرای این الگوریتم، مقدار شاخص صحت‌سنجی DB محاسبه گردید. بر اساس توصیه ذکر شده در قسمت ۲-۲ جهت تعیین تعداد بهینه خوشه‌ها، مدل خوشه‌بندی می‌توانست از تعداد ۲ تا ۱۳ خوشه اجرا گردد، اما برای اطمینان از تعیین دقیق‌تر تعداد بهینه خوشه‌ها تعداد ۲۰ خوشه به عنوان حد بالایی خوشه‌بندی انتخاب گردید. تغییرات شاخص صحت‌سنجی DB در مقابل تعداد خوشه‌ها در شکل ۱ نشان داده شده است. همان‌طور که در این شکل قابل مشاهده است، در عمل از تعداد ۱۴ و ۱۵ خوشه به بعد تغییرات به وجود آمده در مقادیر شاخص صحت‌سنجی قابل ملاحظه نمی‌باشد. بر اساس شکل ۱ در تعداد ۱۰ خوشه مقدار شاخص DB، که با یک مربع متمایز شده است، به حداقل مقدار خود (DB = ۰/۵۴۸) رسیده است. همچنین همان‌طور که توسط خطوط رسم شده در شکل ۱ نشان داده شده، تعداد ۱۰ خوشه محل تغییر شیب دو خط ترسیم شده می‌باشد. بنابراین تعداد بهینه خوشه‌ها برای داده‌های این تحقیق بر اساس شاخص DB برابر ۱۰ خوشه می‌باشد.

جدول ۲ چگونگی توزیع ۱۶۲ بازه کانال‌های انتقال در ۱۰ خوشه ایجاد شده و تعداد اعضای هر خوشه را نشان می‌دهد. همان‌طور که از این جدول قابل مشاهده است، خوشه هفتم (C7) با ۳۰ بازه دارای بیشترین عضو است در حالیکه خوشه سوم و دهم هر کدام با ۵ عضو کمترین تعداد بازه را به خود اختصاص داده‌اند. توزیع بازه‌های کانال در خوشه‌های ایجاد شده در جدول ۲ به تصویر کشیده شده است. وضعیت پراکنش مکانی بازه‌های هر کانال درجه دوم شبکه آبیاری قزوین (L1, L2,…,M) به هر کدام از خوشه‌های به دست آمده را می‌توان از این جدول بدست آورد. جهت استخراج اطلاعات بیشتر از خوشه‌های بدست آمده، جدول ۳ تنظیم گردید که در آن دامنه تغییرات این عوامل در هر کدام از خوشه‌های حاصل شده آمده است. از مقایسه جدول ۳ با جدول ۲ می‌توان روابطی که گویای چگونگی قرارگیری بازه‌هایی با خصوصیات مشترک در خوشه‌ها می‌باشد، استخراج نمود. برخی از نتایج حاصله در ادامه ذکر شده است. بازه‌های برخی کانال‌ها در بین اکثر خوشه‌ها پخش شده‌اند و نمی‌توان رابطه خاصی بین بازه‌های این کانال‌ها متصور شد. برای



شکل ۱- مقادیر شاخص صحت‌سنجی DB در برابر تعداد خوشه‌ها

نمونه کانال L7، ۱۸ عضو دارد که پراکندگی زیادی بین خوشه‌های مختلف دارند. در مقابل درصد قابل توجهی از بازه‌های برخی کانال‌ها در یک یا دو خوشه خاص قرار گرفته‌اند. به عنوان مثال کانال L8 شامل ۲۴ بازه می‌باشد که ۸تای آن مربوط به خوشه C5 است. مورد بعدی کانال L3 است که ۴۱ بازه کانال را شامل می‌شود و تقریباً یک چهارم این بازه‌ها یعنی ۱۰ تای آن در خوشه C7 جای گرفته که خود ۳۰ عضو دارد. علاوه بر آن تعداد ۱۳ بازه دیگر کانال L3 به دو خوشه C2 و C5 متعلق است که مطابق جدول ۳ این دو خوشه از لحاظ سه خصوصیت ظرفیت، طول و تعداد سازه‌های انتقال به یکدیگر مشابه هستند. از آن‌جایی که بازه‌های کانال قرار گرفته در یک خوشه دارای خصوصیات فیزیکی مشابه می‌باشند، زمانی که این بازه‌های مشابه در طول یک کانال انتقال قرار گیرند، می‌توان مدیریت یکسانی در ارزیابی و نگهداری و بهره‌برداری از این بازه‌های مشابه اتخاذ نمود.

نکته دیگری که در تقسیم‌بندی خوشه‌ها مشاهده می‌شود آن است که عامل ظرفیت نقش پررنگ‌تری در ایجاد خوشه‌ها نسبت به سایر عوامل داشته است. این حالت به وضوح در خوشه C3 مشاهده می‌گردد. خوشه مذکور دارای ۵ بازه می‌باشد که بیشترین مقادیر ظرفیت را دارا هستند در حالیکه ۴ خصوصیت دیگر بازه‌های این خوشه پراکندگی زیادی دارد. بعد از خوشه C3 به ترتیب خوشه‌های C1، C10 و C8 ظرفیت‌های متوسط و حداقل را به خود اختصاص داده‌اند. سایر خوشه‌ها در یک محدوده ظرفیتی مشابه هستند که در آن‌ها سایر پارامترها سبب جدائی این خوشه‌ها از یکدیگر شده‌اند. به عنوان مثال خوشه‌های C8، C7 و C4 به ترتیب بازه‌هایی با طول‌های کم، متوسط و زیاد را به خود اختصاص داده‌اند.

خصوصیت تعداد سازه‌های انتقال نیز تنها در خوشه C8 به عنوان عامـــل مشترک میان بازه‌ها محسوب می‌شود. دو خوشه C2 وC5 از

جدول ۲- چگونگی توزیع بازه‌ها در خوشه‌های ایجاد شده

خوشه	تعداد اعضا	L1	L2	L3	L4	L4A	L5	L6	L7	L8	L9	L10	L20	کانال‌های M
C1	۱۶	۰	۲	۴	۰	۰	۱	۳	۴	۲	۰	۰	۰	۰
C2	۲۱	۳	۳	۷	۱	۰	۱	۱	۰	۲	۲	۱	۰	۰
C3	۵	۰	۰	۴	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰
C4	۱۵	۰	۱	۳	۰	۱	۰	۲	۱	۳	۰	۱	۲	۱
C5	۲۵	۱	۰	۶	۰	۰	۲	۱	۴	۸	۱	۱	۰	۱
C6	۱۷	۴	۱	۳	۲	۲	۱	۱	۰	۰	۰	۰	۰	۳
C7	۳۰	۲	۵	۱۰	۰	۰	۰	۳	۴	۳	۰	۰	۰	۳
C8	۱۶	۱	۲	۲	۰	۰	۱	۲	۴	۲	۱	۰	۱	۰
C9	۱۲	۲	۲	۲	۰	۱	۰	۱	۱	۲	۰	۰	۰	۱
C10	۵	۰	۱	۰	۱	۱	۰	۱	۰	۱	۰	۰	۰	۰

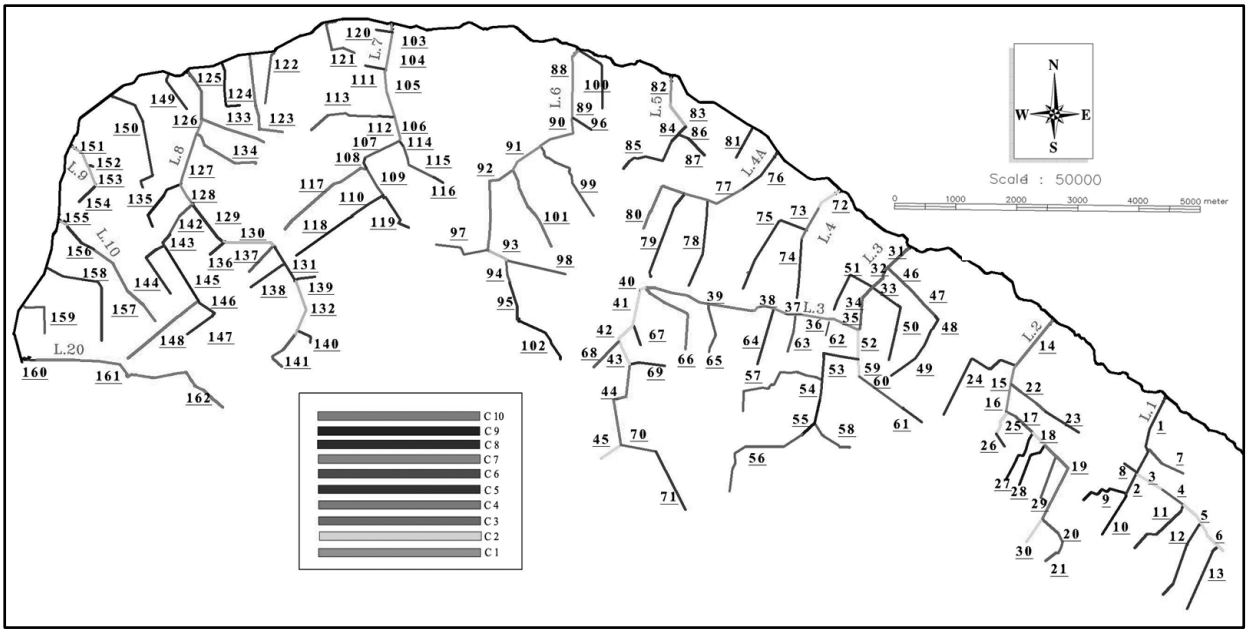
جدول ۳- دامنه تغییرات عوامل به تفکیک خوشه‌های بدست آمده

ردیف	نام خوشه	تعداد بازه	طول بازه (m)		ظرفیت (m^3/s)		سطح آب‌خور (m^2)		تعداد آبگیر		تعداد سازه‌های انتقال	
			حداقل	حداکثر	حداقل	حداکثر	حداقل	حداکثر	حداقل	حداکثر	حداقل	حداکثر
۱	C 1	۱۶	۴۰۰	۲۷۹۰	۱.۴۳	۳.۸	۱۱۸۹۲۷۷	۱۵۱۰۶۱۴۹	۱	۵	۲	۱۲
۲	C 2	۲۱	۴۰۰	۱۸۸۰	۰.۳۴	۱.۸۲	۵۴۲۶۰۶	۱۰۲۱۷۱۶۸	۱	۲	۲	۱۳
۳	C 3	۵	۴۰۰	۲۵۹۰	۴.۳	۷.۳۵	۱۵۹۲۶۶۱	۲۷۳۶۰۴۶۳	۱	۵	۵	۱۶
۴	C 4	۱۵	۳۴۵۶	۶۱۹۰	۰.۱	۲	۱۵۷۵۰۷۱	۱۸۹۸۱۵۷۴	۱	۶	۳	۳۰
۵	C 5	۲۵	۳۲۵	۲۱۲۹	۰.۱	۱.۲۵	۹۴۴۲۰۷	۸۵۷۴۳۲۰	۲	۵	۰	۱۱
۶	C 6	۱۷	۲۲۵۴	۶۱۹۱	۰.۱۵	۱.۶۳	۲۳۳۵۸۵۵	۲۱۱۱۵۱۰۲	۵	۱۳	۱۰	۵۶
۷	C 7	۳۰	۱۵۳۵	۳۶۹۰	۰.۱	۱.۱۴	۳۴۳۵۴۱	۸۲۹۷۸۶۸	۱	۳	۲	۳۰
۸	C 8	۱۶	۱۲۰	۱۰۰۰	۰.۱	۰.۹۵	۲۸۰۹۶۲	۵۵۰۸۹۷۷	۱	۲	۰	۶
۹	C 9	۱۲	۲۴۶۰	۳۷۲۰	۰.۱	۱.۵۸	۳۰۵۴	۶۹۷۳۶۹۳	۳	۹	۵	۲۵
۱۰	C 10	۵	۱۲۰۰	۳۳۸۱	۱	۲.۸۵	۱۸۹۴۸۴۸۶	۴۹۱۲۸۳۴۲	۱		۳	۳۰

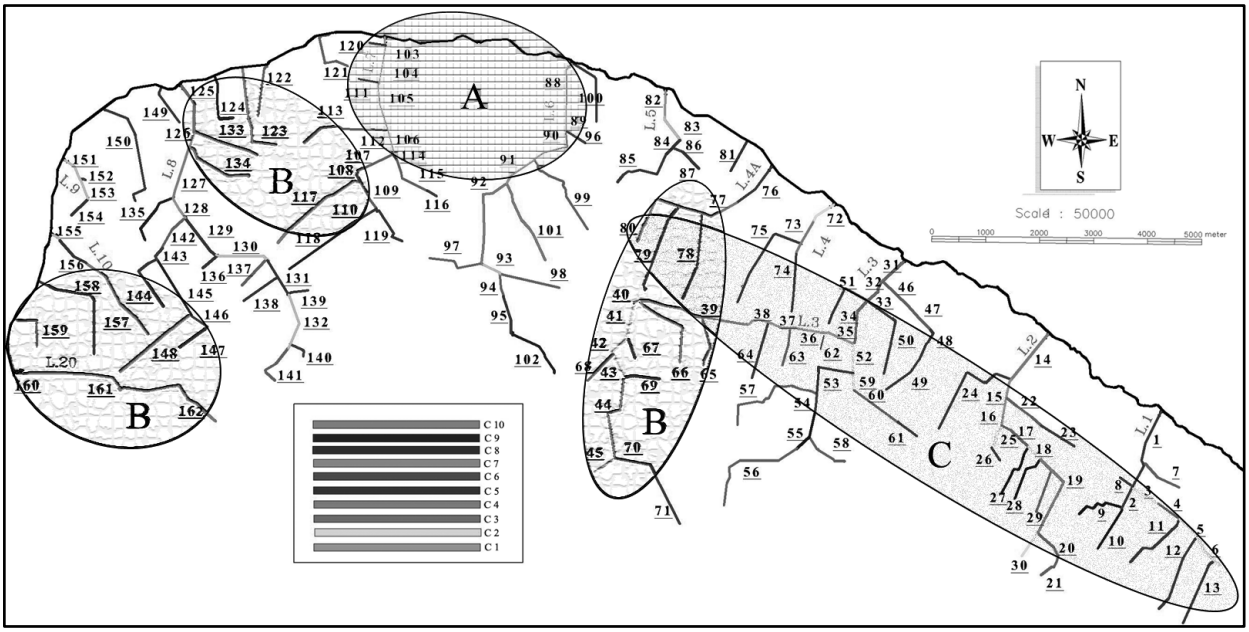
لحاظ سه خصوصیت ظرفیت، طول و تعداد سازه‌های انتقال به یکدیگر مشابه‌اند ولی خصوصیت تعداد آبگیر سبب جدایی این دو خوشه از یکدیگر شده است. براین اساس می‌توان نتیجه گرفت خوشه‌بندی تمامی خصوصیات را به طور ترکیبی و هم‌زمان در جداسازی خوشه‌ها از یکدیگر در نظر می‌گیرد و تنها به چند خصوصیت محدود جهت خوشه‌بندی اکتفا نکرده است.

شکل ۲ خروجی مدل خوشه‌بندی را روی شبکه آبیاری قزوین به تصویر می‌کشاند. از آن‌جایی که مقادیر عضویت حاصل شده در روش کلاسیک تنها صفر یا یک می‌باشد، می‌توان هر بازه کانال را با قطعیت به یکی از ده خوشه ایجاد شده، که در شکل ۲ با یک رنگ خاص ایجاد شده است، متعلق دانست. در این شکل همچنین، چگونگی توزیع مکانی خوشه‌های ایجاد شده به‌تصویر کشیده است. در این شکل هر خوشه با یک رنگ مشخص شده است و با توجه به نام و رنگ هر بازه می‌توان عضویت آن‌ها را به هر خوشه مشخص کرد. از آن‌جایی که ویژگی‌های در نظر گرفته شده برای خوشه‌بندی، ویژگی‌های فیزیکی هستند، این شکل نشان‌دهنده پهنه‌بندی فیزیکی بازه‌های کانال انتقال درجه ۲ شبکه قزوین می‌باشد. از نقشه پهنه‌بندی ارائه شده می‌توان برخی الگوهای موجود در پراکندگی مکانی بازه‌های کانال‌های انتقال را استخراج نمود. به عنوان مثال به نظر می‌رسد بازه‌های قرار گرفته در برخی از خوشه‌ها از لحاظ موقعیت مکانی در یک منطقه یا به صورت چند کلونی در سطح شبکه قرار گرفته است. این روند به خصوص در مورد خوشه‌های C1، C2 و C3 قابل رویت است.

در شکل ۳ خوشه C1 (رنگ نارنجی) به صورت محدوده A مشخص گردیده که در منطقه بالادست و میانی دو کانال L6 و L7 پراکنده شده است. خوشه C4 (رنگ صورتی) به صورت سه کلونی B در میان‌دست و پایین‌دست شبکه متمایز شده و خوشه C6 (رنگ قهوه‌ای) با یک کلونی کشیده C در محدوده بالادست شبکه نشان داده شده است. خروجی ناحیه‌بندی فیزیکی شبکه آبیاری قزوین چند کلونی، متشکل از بازه‌های کانال مشابه، در قسمت‌های بالادست، میانی و پایین‌دست این شبکه می‌باشد. این‌کار، مدیران شبکه را قادر می‌سازد با ایجاد دفاتر جداگانه بهره‌برداری و نگهداری مجزا در سه ناحیه بالادست، میانی و پائین‌دست شبکه، بتوانند بازدیدها و سرویس‌های مورد نیاز را به کلونی‌های مجاور خود بدهند. از آنجایی که هر کلونی مجموعه‌ای از بازه‌های کانال با خصوصیات فیزیکی تقریباً مشابه به همدیگر است، غالباً مشکلات مشابهی برای بازه‌های واقع‌شده در هر کلونی ایجاد می‌شود. هر کدام از این دفاتر بهره-برداری و نگهداری اقدام به سرویس‌رسانی به ناحیه مجاور خود نموده که سبب هدفمند شدن تخصیص اعتبارات، امکانات و نیروی ماهر در محدوده شبکه خواهد شد.



شکل ۲- پهنه بندی مکانی کانال‌های انتقال با روش خوشه بندی K-Means



شکل ۳- توزیع مکانی کلونی شکل خوشه‌های C1، C4 و C6

۴- خلاصه و جمع‌بندی

در این تحقیق پهنه‌بندی مکانی شبکه آبیاری قزوین، بر اساس مشخصات فیزیکی بازه‌های کانال‌های انتقال درجه دوم این شبکه و با بهره‌گیری از روش خوشه‌بندی کلاسیک K-Means انجام گردید. براساس نتایج شاخص صحت‌سنجی خوشه‌بندی مورد استفاده در این تحقیق، کانال‌های انتقال درجه دوم شبکه آبیاری قزوین به ۱۰ خوشه دسته‌بندی شدند. بازه‌های کانال قرار گرفته در هر خوشه از لحاظ خصوصیات فیزیکی و فنی مشابه هم‌دیگر هستند. مهم‌ترین کاربرد خوشه‌بندی در این تحقیق، خلاصه‌سازی شبکه انتقال قزوین مشتمل بر ۱۶۲ بازه کانال به ۱۰ خوشه یا به عبارت دیگر به ۱۰ منطقه همگن فیزیکی در سطح این شبکه می‌باشد. خلاصه‌سازی به خصوص در بحث مدیریتی سبب افزایش قدرت تصمیم‌گیری مدیران شبکه در زمان حداقل خواهد شد. مناطق همگن فیزیکی استخراج شده در سطح شبکه در صورتی که مانند برخی از خوشه‌های ایجاد شده در این تحقیق به صورت کلونی در سطح شبکه قرار گیرند، امور مربوط به بهره‌برداری و بازرسی‌ها را برای مدیران شبکه آسان خواهند نمود. دلیل این امر آن است که بازرسی و نظارت بر عملکرد سازه‌ها و کانال‌ها در یک زمان برای دسته‌ای از کانال‌های مشابه صورت می‌پذیرد. روند فوق مدیران بهره‌برداری شبکه را قادر می‌سازد علاوه بر انجام بازرسی‌ها، تخصیص اعتبارات و امکانات را جهت انجام عملیات بهره‌برداری و نگهداری از حالت موردی خارج ساخته و تصمیم‌گیری‌های خود را کلی‌تر و برای دسته‌ای از بازه‌ها در نظر گیرند. این کار سبب صرفه‌جوئی قابل ملاحظه‌ای در زمان و هزینه و به دنبال آن، تسهیل عملیات بهره‌برداری و نگهداری خواهد شد.

یکی دیگر از کاربردهای پهنه‌بندی مکانی شبکه‌ها، استفاده از نتایج پهنه‌بندی در تسهیل امور نگهداری کانال‌های شبکه می‌باشد. به عنوان مثال چون در این تحقیق از خصوصیات فیزیکی و فنی بازه‌های کانال‌ها جهت خوشه‌بندی استفاده شده است و بازه‌های مشابه از لحاظ خصوصیات فنی در یک خوشه قرار گرفته‌اند، می‌توان بازه‌هائی را که با مشکلات سازه‌ای مانند خرابی سازه‌های انتقال یا آبگیرها هستند را شناسائی نمود و مشکلات مشابه را برای سایر بازه‌های مشابه با بازه مذکور پیش‌بینی کرد. در مجموع روش خوشه‌بندی ابزاری مفید و موثر در شناسائی مناطق همگن در شبکه‌های آبیاری برای تسهیل امر مدیریت و بهره‌برداری شبکه‌ها می‌باشد.

۵- تشکر

نویسندگان مقاله مراتب سپاس و تشکر خود را از دفتر بهره‌برداری و نگهداری شبکه‌های آبیاری سازمان مدیریت منابع آب ایران و نیز شرکت بهره‌برداری شبکه آبیاری قزوین، به جهت همکاری صمیمانه ایشان در انجام این تحقیق اعلام می‌دارند.

پی‌نوشت‌ها

1- Unsupervised learning task
2- Identifiers
3- Goulburn
4- Shi-Jin
5- Data mining
6- Pattern Recognition
7- Patterns
8- Objects
9- Features
10- Attributes
11- Similarity matrix

۶- مراجع

حیدریان، س. ا.، فرداد، ح.، منعم، م. ج.، لیاقت، ع.، قاهری، ع. و تشنه‌لب، م. (۱۳۸۲)، "به کارگیری رویکرد فازی در ارزیابی سیستم‌های آبیاری"، مجله تحقیقات مهندسی کشاورزی، ۱۷(۴): صص ۴۷-۶۳.

خلخالی، م.، منعم، م. ج.، و ابراهیمی ک. (۱۳۸۷)، "تدوین مدل پشتیبانی تصمیم برای ارزیابی و بهبود عملکرد شبکه‌های آبیاری و زهکشی"، مجله تحقیقات مهندسی کشاورزی، ۹ (۱): صص ۱۲۵-۱۴۰.

منعم، م. ج.، علیرضائی، م. ر. و صالحی طالشی، ا. (۱۳۸۱)، "ارزیابی عملکرد بهره‌برداری ازشبکه‌های آبیاری به روش تحلیل پوششی داده‌ها"، مجله علوم و فنون کشاورزی و منابع طبیعی، ۶ (۴): صص ۲۵-۱۱.

Bruscoli, P., Bresci, E. and Preti, F. (2001), "Diagnostic analysis of an irrigation system in the andes region," *Agriculture Engineering International: CIGR Journal*, 3(1), pp. 12-26.

Burt, C. (2001), *Rapid Appraisal Process (RAP) and Benchmarking Explanation and Tools*, FAO, Bangkok, 50p.

Davies, D. L. and Bouldin, D.W. (1979), "A cluster separation measure", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1(4), pp. 224-227.

Han, J. and Kamber, M. (2006), *Data Mining: Concepts and Techniques*, Elsevier Inc, San Francisco, 743p.

Johnson, R.A. and Wichern, D.W. (1999), *Applied multivariate statistical analysis*, John wiley & sons. New York, 550p.

Kim, D.W., Lee, K.H., and Lee, D. (2004), "On cluster validity index for estimation of the optimal number of fuzzy clusters," *Pattern Recognition*, 37 (4), pp. 2009-2025.

Malano, H. and Burton, M. (2001), *Guidelines for Benchmarking Performance in the Irrigation and Drainage Sector*, FAO, IPTRID, Rome, 145p.

Malano, H. and Gao, G. (1992), "Ranking and classification of irrigation system performance using fuzzy set theory: case study in Australia and China," *Irrigation and Drainage Systems*, 6 (2), pp. 129-148.

Oad, R. and Mc Cornick, P.G. (1989), "Methodology for assessing the performance of irrigation agriculture," *ICID Bulletin*, 38 (1), pp. 42-53.

Theodoridis, S. and Koutroumbas, K. (2003), *Pattern Recognition*, Elsevier Press, USA, 837p.

Valente, J.O. and Pedrycz, W. (2007), *Advances in Fuzzy Clustering and Its Applications*. John Wiley & Sons Ltd, England, 434p.

Van der Heijden, F., Duin, R. P. W., de Ridder, D., and Tax, D. M. J. (2004), *Classification, Parameter Estimation and State Estimation*. John wiley & sons Ltd, England, 423p.