

ارائه‌ی یک دسته‌بند داده با استفاده از ترکیب دسته‌بندی کننده‌ی ملهم از سیستم‌ایمنی مصنوعی و Fuzzy-knn

الهام مفیدی روچی^۱، امیر مسعود افتخاری مقدم^۲

^۱ دانشکده مهندسی کامپیوتر و فن‌آوری اطلاعات، دانشگاه آزاد اسلامی، واحد قزوین، قزوین، ایران
elham.mofidi@gmail.com

^۲ گروه کامپیوتر، دانشکده مهندسی کامپیوتر و فن‌آوری اطلاعات، دانشگاه آزاد اسلامی، واحد قزوین، قزوین، ایران
eftekhari@qiau.ac.ir

چکیده

یکی از مسایل اساسی و مهم در داده‌کاوی، ایجاد یک مدل کارآمد با حداکثر دقت، برای دسته‌بندی داده می‌باشد. در این مقاله، یک دسته‌بند دومرحله‌ای که از ترکیب دسته‌بندی کننده‌ی ملهم از سیستم‌ایمنی مصنوعی و الگوریتم Fuzzy-knn تشکیل شده، معرفی گردیده است. در مرحله‌ی اول، دسته‌بندی کننده، بر مبنای شبکه‌ی ایمنی مصنوعی و انتخاب کلونی عمل می‌کند و مجموعه‌ای بهینه از داده‌های آموزشی را به عنوان دسته‌بندی کننده ارائه می‌دهد و در مرحله‌ی دوم، با استفاده از الگوریتم Fuzzy-knn، داده‌های تست دسته‌بندی می‌شوند. در مرحله‌ی اول، از شیوه "وزن‌دهی فازی" (Fuzzy weighting) استفاده می‌شود و همچنین برای محاسبه‌ی میل ترکیبی از معیار HEOM (Heterogeneous Euclidian overlap metric) فازی استفاده می‌شود. برای بررسی کارایی مدل پیشنهادی از پنج مجموعه داده از UCI که برای کاربردهای دسته‌بندی مورد استفاده قرار می‌گیرند، استفاده شده است که عبارتند از Australian Credit و German Credit و Pima Indians Diabetes و Iris و Ionosphere. برای ارزیابی کارایی دسته‌بندی کننده، از شیوه 10-fold cross-validation استفاده شده است. نتایج بدست آمده، نشان می‌دهد که پس از اعمال HEOM فازی و fuzzy-knn، مدل پیشنهادی نتایج قابل قبولی را ارائه می‌دهد.

کلمات کلیدی

دسته‌بند داده، سیستم ایمنی مصنوعی، الگوریتم Fuzzy-knn، شبکه‌ی ایمنی مصنوعی، انتخاب کلونی، وزن‌دهی فازی، معیار HEOM فازی

Unsupervised AIC Immune Recognition System
Adaptive immune k-means Discrimination-based AIS
AIS و AIC with Swarm Learning و Simple AIS

الگوریتم aiNet با افزودن مفهوم شبکه ایمنی به الگوریتم CLONALG، به جهت خوشه بندی و کاهش ابعاد داده‌های گسسته ارائه شد [۵].

تیمیس در سال ۲۰۰۰، الگوریتم یادگیری بر مبنای شبکه‌ی ایمنی مصنوعی تحت نام RAIN را مطرح کرد که یک "سیستم ایمنی مصنوعی با منابع محدود" می‌باشد. مهمترین مشخصه آن، محدود بودن تعداد سلول های B می‌باشد [۶].

الگوریتم AIRS توسط واتکینز در سال ۲۰۰۱ مطرح شد که از روش دسته‌بندی نزدیکترین همسایگی استفاده می‌کند. این الگوریتم

۱- مقدمه

یکی از مسایل اساسی و مهم در داده‌کاوی، ایجاد یک مدل کارآمد و مناسب، برای دسته‌بندی داده می‌باشد، به گونه ای که بتواند داده‌های با ابعاد گوناگون را با حداکثر دقت و درستی طبقه‌بندی کند. اخیراً، تعداد زیادی از روش‌های دسته‌بندی در زمینه‌های مختلف، توسعه داده شده‌است که عبارتند از: الگوریتم ژنتیک، تئوری مجموعه راف، مدل‌های نزدیکترین همسایگی، ماشین بردار پشتیبان، تئوری فازی و درخت‌های تصمیم و مطالعات ما، یک شیوه جدید از محاسبات تکاملی تحت نام "سیستم ایمنی مصنوعی" را پیشنهاد می‌دهد [۳ و ۴]. در حوزه‌ی دسته‌بندی داده، الگوریتم‌های بر مبنای AIS مختلفی مطرح شده است که عبارتند از: Artificial Resource Limited AIS،