



ارائه الگوریتم جدید برای دنباله های ترتیبی جهت فشردن آنها در فرآیند کاوش

محبوبه راسخی^۱

چکیده:

داده کاوی فرآیند کشف الگوهای معتبر، جدید، مفید و قابل درک است. می توان گفت داده کاوی فرآیندی است که طی آن با استفاده از ابزارهای تحلیل داده به دنبال کشف الگوها و ارتباطات میان داده های موجود که ممکن است منجر به استخراج اطلاعات جدیدی از پایگاه داده گردند می باشد. دو مثال از الگوی ترتیبی عبارتند از: " ۸۰٪ مشتریانی که تلویزیون می خرند، در همان روز دوربین هم خواهند خرید"، " هر بار که سهام شرکت مایکروسافت ۵٪ کاهش یابد، سهام IBM هم در خلال ۳ روز ۴٪ کاهش خواهد یافت". الگوی اولی کمک می کند که مدیریت قفسه فروشگاه را به خوبی انجام بدهیم و الگوی دومی به شرکت کمک می کند که در بحران اقتصادی به درستی عمل کند. در این پژوهش ابتدا مروری را بر الگوریتم های استخراج الگو انجام دادیم و براساس ترتیب تاریخی ارائه الگوریتم ها آنها را مورد بررسی قرار دادیم. این الگوریتم ها خود در دو دسته براساس *Apriori* و براساس *FP-Growth* تقسیم می شوند. سپس دو الگوریتم پیشنهاد شده *SeqKrimp* و *GoKrimp* را با استفاده از اصل حداقل طول توصیف کمینه، یعنی الگوهای مفیدند که بیشتر پایگاه داده را فشرده می کنند، را معرفی نمودیم. پس از آن روش پیشنهادی خود را معرفی کرده و نام آن را *newKrimp* گذاشتیم و نتایج را بر روی پایگاه داده های استاندارد مشاهده نمودیم.

واژه های کلیدی: داده کاوی، کشف الگو، پایگاه داده ترکیبی، الگوریتم استخراج، زیر دنباله

^۱ کارشناسی ارشد مهندسی فناوری اطلاعات، موسسه غیرانتفاعی علامه جعفری واحد رفسنجان، رفسنجان، ایران،

m.rasekhi6068@gmail.com