



ID: P-207

کاوش قوانین وابستگی در توالی‌های ویروس آنفولانزا

۱۲۴

مریم بهرامی در اسله^{۱*}، اشکان سامی^۱، اسماعیل ابراهیمی^۲، محمدهادی صدرالدینی^۱

۱. دانشگاه شیراز، دانشکده مهندسی برق و کامپیوتر، بخش مهندسی و علوم کامپیوتر و فناوری اطلاعات

۲. دانشگاه شیراز، دانشکده کشاورزی، بخش زراعت و اصلاح نباتات

mbahrami@cse.shirazu.ac.ir

ویروس آنفولانزا از سال ۱۹۰۰ تاکنون، منجر به چهار اپیدمی مرگبار در سراسر جهان شده است: ۱۹۱۸-آنفولانزای اسپانیایی-H1N1، ۱۹۵۷-آنفولانزای آسیایی-H2N2، ۱۹۶۸-آنفولانزای هنگ کنگ-H3N2 و ۲۰۰۹-H1N1. علاوه بر شیوع سریع، نرخ بالای جهش اعم از تغییر آنتی‌ژن و انحراف آنتی‌ژن، کارایی واکسن را کاهش داده است، از این رو واکسن آنفولانزا نیاز به بازنگری سالانه دارد و به این دلیل استخراج سالانه الگوی ویروس در گردش، ضروری است. مشکل دیگر این است که دیتاست‌های آنفولانزا در تمام زیرگروه‌های آن، توزیع نامتوازی دارد. از CBA، طبقه‌بندی بر اساس قوانین وابستگی، که یکی از تکنیک‌های داده‌کاوی برای کشف قوانینی است که منجر به کلاس مشخصی می‌شوند، برای استخراج الگوی ویروس استفاده شد. برای حل مشکل عدم توازن داده‌های ویروس از زیرگروه‌های مختلف، آستانه‌های متفاوتی از support و confidence بر مبنای فرکانس هر زیرگروه در دیتاست استفاده شد. نهایتاً توانستیم قوانین با اطمینان بالا، تقریباً ۱۰٪، برای هر کدام از زیرگروه‌های ویروس استخراج کنیم، برای نمونه در توالی DNA آنتی‌ژن هم‌گلو تینین قانون $C_{1617}=A \rightarrow H=3$, Support=26%, Confidence=100% و قانون $C_{410}=G \rightarrow H=1$, Support=40%, Confidence=100% بدست آمد. قوانین بدست آمده، در سیستم‌های مبتنی بر قانون، نیز با توجه به طول زیاد توالی‌ها، در استخراج نواحی موثر در توالی به‌منظور کاهش حجم بالای داده، برای مطالعات بیشتر قابل استفاده است.

کلمات کلیدی: آنفولانزا، کاوش قوانین وابستگی، توالی.

