

استفاده از داده‌کاوی برای تحلیل رفتار بینندگان تلویزیونی

بهروز مینایی بیدگلی^{۱۹۷}، کیاوش شیدانی^{۱۹۸}

چکیده

در سالهای اخیر فرهنگ رسانه‌ای به پیشرفت‌هایی نایل گشته است. مطابق با آن تعامل کانال‌ها با بینندگان به شیوه‌های بنیادی و اساسی در حال تغییر است. کانال‌های تلویزیونی به منظور نظارت بر این گونه تغییرات نیازمند آرایه راه‌حل‌های مناسبی هستند. امروزه کانال‌های تلویزیونی از طریق تجزیه و تحلیل برنامه‌های دیده شده توسط بینندگانشان به افزایش ارزش بیننده دست یافته‌اند. ابزارها و فناوری‌های داده‌انبار، داده‌کاوی و دیگر تکنیک‌های مدیریت بینندگان تلویزیونی، روش‌هایی هستند که فرصت‌های جدیدی را برای این کانال‌ها فراهم خواهند کرد. در واقع دیدگاه کانال محوری جای خود را به بیننده محوری داده است. بنابراین، با جمع‌آوری داده‌های مربوط به بیننده و تصمیم‌گیری براساس الگوهای استخراج شده از روابط پنهان میان داده‌ها به وسیله ابزار داده‌کاوی، می‌توان به علائق بیننده‌محوری خود جامه عمل پوشاند. با به‌کارگیری ابزار داده‌کاوی ترجیح‌های بینندگان تلویزیونی را از میان انبوهی از آمار بینندگان کشف کرده و به وسیله الگوهای پنهان حاکم بر این داده‌ها می‌توانیم رفتار بینندگان را پیش‌بینی کرده و به این ترتیب نیاز بیننده محوری را که هدف اصلی شبکه‌های تلویزیونی است را جامه عمل پوشیم. برای نیل به این هدف از روش‌های مختلف داده‌کاوی از قبیل استفاده از خوشه‌بندی برای یافتن گروه‌های بینندگان، رگرسیون برای نرمال‌سازی آمار بینندگان تجمیع، رده‌بندی برای مدل کردن رفتار بیننده، و الگوریتم‌های یافتن قوانین انجمنی و الگوهای توالی برای ارائه سیستم پیشنهاددهنده و غیره استفاده می‌کنیم. در این مقاله ضمن بررسی روش‌های مختلف تحلیل داده‌های بینندگان سعی می‌شود الگوهای کلی رفتاری کشف گردد.

جمع‌آوری اطلاعات بینندگان (*Viewer's Profile*) به کمک دستگاه‌های *Set-up Box* انجام می‌پذیرد؛ این دستگاه‌ها در منازل کار جمع‌آوری برنامه‌های دیده شده را انجام می‌دهند. به این ترتیب اطلاعاتی نظیر گروه سنی، جنسیت، کانال، زمان تماشا و... گرد-آوری می‌شوند. دادگان این تحقیق داده‌های واقعی کشور آلمان در سال 2007 می‌باشد. این داده‌ها در دو دسته کلی اطلاعات برنامه-های پخش شده و اطلاعات دقیقه‌ای (هر 5 دقیقه) برای کانال‌های مختلف در مناطق مختلف جغرافیایی یا همان استانهای کشور آلمان به تفکیک گروه‌های سنی در اختیار است. حجم داده‌های برنامه‌ای در حدود 30000 ردیف در 24 ساعت بوده و حجم اطلاعات دقیقه‌ای در حدود 390000 ردیف در روز می‌باشد. اما مشکل بزرگ پیش رو، پیکربندی خاص این داده‌ها و پراکنده بودن آنها در فایل‌های متفاوت (بسته به استان و کانال) بوده که با پیاده‌سازی ابزاری برای بازیابی این اطلاعات و وارد نمودن آنها در یک پایگاه داده این مشکل مرتفع گردید. با ابزار *WEKA* رده‌بندی و خوشه‌بندی بر روی داده‌های آماده‌سازی شده صورت گرفته است. درخت تصمیم‌گیری حاصل دارای پایداری بسیار بالایی بوده که این امر ناشی از حجم نسبتاً بالای داده‌های در دسترس برای آموزش و آزمایش می‌باشد. توسط درخت تصمیم‌گیری قادر خواهیم بود که به پیش‌بینی میزان بیننده برای کانال خاص در مناطق جغرافیایی و رده‌های سنی مختلف دست یابیم. همچنین خوشه‌بندی داده‌ها الگوهای رفتاری بسیار گویا و قابل توجهی را در اختیار قرار داده است.

کلمات کلیدی

داده‌کاوی، تحلیل رفتار بینندگان تلویزیونی، الگوهای داده، پیشنهاد انتخاب، شخصی‌سازی

¹⁹⁷ استادیار دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران

¹⁹⁸ دانشجوی فوق لیسانس مهندسی فناوری اطلاعات دانشگاه علم و صنعت ایران