



یک روش ترکیبی خوشه بندی و الگوریتم رقابت استعماری برای پیش بینی

پیوند در شبکه های اجتماعی

حمید کوهساری^۱، دکتر ایمان عطارزاده^۲ و دکتر علی هارون آبادی^۳

^۱ گروه مهندسی کامپیوتر، دانشکده فنی، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، تهران،

hamidkouhsari1390@gmail.com

^۲ استادیار گروه مهندسی کامپیوتر، دانشکده فنی، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، تهران،

attarzadeh.iauctb@gmail.com

^۳ استادیار گروه مهندسی کامپیوتر، دانشکده فنی، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، تهران،

a.harounabadi@gmail.com

چکیده - امروزه، پیش بینی پیوند در شبکه های اجتماعی در حال توسعه و گسترش می باشد. به گونه ای که پیوندهای ارتباطی بین کاربران ها در شبکه های اجتماعی دائماً در حال افزایش، تحول و بهره برداری به وفور بوده و همچنین، روزانه کاربران بسیاری در اقصا نقاط جهان به این شبکه اجتماعی بزرگ افزوده می شوند. از این رو، کاربرد روزافزون این شبکه ها و هم حفظ حضور کاربران شبکه ها، امری مهم و چالش برانگیز است. از آنجا که ساختار این شبکه ها، براساس ساختار گراف شبکه می باشد، پیش بینی پیوندها را در این تحقیق مورد بررسی قرار می دهیم. با توجه به اینکه مسائل بهینه سازی، نوعی از مسائل $NP-Hard$ می باشد، برای بهینه سازی دقیق تر گراف و پیش بینی پیوند موثرتر از الگوریتم رقابت استعماری که یک مسئله بهینه سازی می باشد، استفاده می کنیم. در این پایان نامه، ابتدا یک روش پیش بینی پیوند در شبکه های اجتماعی با رویکرد انتخاب ویژگی براساس الگوریتم رقابت استعماری پیشنهاد می گردد. در ابتدا داده ها، با الگوریتم $k-means$ خوشه بندی می شوند. سپس، در رویکرد پیشنهادی از انتخاب ویژگی استفاده می شود به این ترتیب که الگوریتم رقابت استعماری بت تابع برازندگی پیرسون، بهترین ویژگی ها را در بین ویژگی های موجود انتخاب می کند و سپس مجدداً عملیات خوشه بندی بر روی داده ها با ویژگی های جدید صورت می پذیرد. تا ویژگی های موثر در خوشه بندی توسط این الگوریتم استخراج شود. سپس با استفاده از درخت تصمیم روش پیشنهادی، ورود کاربر جدید در ساختار اراده شده، مشخص می گردد که کاربر جدید به کدام خوشه تعلق دارد. در نهایت از الگوریتم پیش بینی پیوند $Friendlink$ ، برای پیش بینی پیوند براساس ساختار الگوریتم پیشنهادی استفاده می شود. در ارزیابی روش پیشنهادی از دیتاست $Pokec$ برای شبکه اجتماعی استفاده می کنیم و از مقادیر دقت، صحت، بازخوانی و $F-measure$ در ارزیابی استفاده می شود. نتایج بدست آمده نشان می دهد که روش پیشنهادی نسبت به سایر رویکردهای مورد بررسی، کارایی و دقت بالاتری را دارد.

کلید واژه- انتخاب ویژگی، خوشه بندی، همبستگی پیرسون، الگوریتم رقابت استعماری، پیش بینی پیوند.

۱- مقدمه

کند [۱]. شبکه های اجتماعی از یک گروه از مردم که علایق، منافع سابقه و وضعیت کاری همچنین فعالیت هایشان را به اشتراک گذاشته اند، ساخته شده است. در این گونه شبکه ها، مردم می توانند با یکدیگر با روش های مختلفی ارتباط برقرار کنند. آنها می توانند از لحاظ اجتماعی، به آپلود فایل هایی مانند تصاویر، فیلم ها و فایل های دیداری در پروفایل خود بپردازند و آنها را به اشتراک گذارند. می توان گفت شبکه های اجتماعی از نودهایی که بازیگران در شبکه هستند، تشکیل شده است. این نود ممکن است یک

در حال حاضر در سراسر جهان، جنبش شبکه های اجتماعی شکل گرفته است، یعنی بسیاری از وبسایت ها امکانات و زیرساخت های شبکه اجتماعی را در خود جای داده اند. طبق آمارها هر کاربر اینترنتی زمان کمی را صرف بازدید از یک وب سایت می کند، در حالی که وجود امکانات شبکه های اجتماعی باعث ماندگاری بازدید کنندگان و کاربران در این نوع وبسایت ها می شود و در نهایت ارتباطی پویا و دایمی بین مخاطب و خدمات دهنده برقرار می