

شناسایی اسپم ایمیل ها با استفاده از درخت پوشای بیزین

صادق خوارزمی¹ و علی فرهمند نژاد²

چکیده

ما در این مقاله روشی ساده برای دسته بندی ایمیل ها به دو دسته اسپم و غیر اسپم بر اساس شبکه های بیزین با استفاده از درخت پوشا ارائه کردیم. تا کنون روش های بسیاری در مورد شناسایی اسپم ایمیل ها در مقالات ارائه شده است. در درخت پوشای بیزین بجز گره ریشه تمامی گره ها حد اکثر دو گره والد دارد. ما برای بدست آوردن زیر مجموعه ای از خصوصیت ها که هر کدام نشان دهنده یک کلمه می باشد از روش BST استفاده می کنیم که برای انتخاب کلمات مرتبط بکار می رود. در این پژوهش کارایی روش با استفاده از تکنیک های F-measures و Accuracy اندازه گرفته شده است که نتایج آن در این مقاله نشان داده شده است. نتایج این کار نشان می دهد که دسته بندی ایمیل ها بر اساس درخت پوشای بیزین دقت قابل رقابتی در عمل خواهد داشت

کلمات کلیدی

اسپم ایمیل , شبکه های بیزین , درخت پوشای بیزین , داده کاوی , اندازه گیری کارایی

Spam Email Detection Using Bayesian Spanning Tree

Sadegh Kharazmi; Ali Farahmand Nezhad

ABSTRACT

In this paper we proposed a simple approach for email classifying as spam or legitimate based on Bayesian network using spanning tree. Up to now many approaches about spam email detection exhausted in presented papers. In Bayesian spanning tree except root node, each node has at most two parent nodes. Also for obtaining a subset of attributes that each of them show a word, we used BST (Bayesian Spanning Tree) method to relevant word selections. We measure performance based on F-measures and accuracy. Our works show that emails classifying based on Bayesian Spanning Tree has good comparable accuracy in experience

KEYWORDS

Spam email, Bayesian network, Bayesian Spanning Tree, Data mining, Performance measures

kharazmi@pnu.ac.ir
afarahmand@pnu.ac.ir

¹ دانشجوی کارشناسی ارشد مهندسی نرم افزار - دانشگاه پیام نور
² دانشجوی کارشناسی ارشد مهندسی نرم افزار - دانشگاه پیام نور