IDMC'07
20-21 Nov.2007
Amir Kabir University

استفاده از کلاسه کننده های متخصص برای داده کاوی داده های جریانی

زهرا میرزامومن[۱]، ستار هاشمي[۲]، محمدرضا کنگاوري[۳]

چکیده

در این مقاله، نتایج اولیه استفاده از مدل مجموعه کلاسه کننده هاي متخصص (EOE) در داده کـاوي داده هـاي جریـاني گزارش شده است. EOE از k کلاسه کننده دودویي مجزا تشکیل شده است که هر یک مسوول تشخیص نمونه هـاي یـک کـلاس خاص از نمونه هاي سایر کلاس ها مي باشند. براي کلاس بندي یک نمونه جدید، تمام این k کلاسه کننده اجرا مي شوند و راي کلاسه کننده اي که بیشترین اطمینان را داشته باشد انتخاب مي شود. بنابراین این مدل کاملا با مدل هاي مجموعـه اي موجـود بـراي داده کاوي داده هاي جریاني که اغلب آنها از کلاسه کننده هاي چندکلاسه ساخته شده اند، یعني در آنها هر یک از اعضا دقیقا کلاس یک نمونه ورودي را از بین همه کلاس هاي موجود انتخاب مي کند، متفاوت است. ما در مورد فواید، مشکلات، راه حـل هـا و پیچیـدگي زماني این مدل در داده کاوي داده هاي جریاني تحقیق و بررسي کرده ایم. همچنین روی چهار مجموعه داده معـروف، دقـت کـلاس بندی و زمان اجرای این مدل را با چهار مدل موفق در داده کاوی داده های جریانی مقایسه کرده ایم. تحلیل هـاي تئـوري و شـواهد تجربي نشان مي دهند که این مدل مي تواند یادگیري و بروزرساني سریعتر و دقت کلاس بندي بـالاتري را نسـبت بـه بسـیاري از الگوریتم هاي معروف داده کاوي داده جریاني موجود ارائه دهد. انتظار مي رود که این نتایج مورد توجه و علاقه محققان در این زمینه قرار بگیرد، چرا که یک روش ساده ولي عالي و کارامد در مقایسه با مدل هاي داده کاوي موجود براي داده هاي جریاني پیشنهاد مـي دهد.

کلمات کلیدی

داده کاوي داده جریاني، کلاسه کننده هاي چندتایی

Using Expert Classifiers for Data Stream Classification

Zahra Mirzamomen; Sattar Hashemi; Mohammad Reza Kangavari

ABSTRACT

We report the preliminary results of using ensemble of expert classifiers (EOE) for data stream classification. EOE learns k individual binary classifiers, each one to distinguish the instances of a single class from the instances of all other classes. To classify a new instance, the K classifiers are run and the one that returns the highest confidence is chosen. Thus EOE is different from existing data stream classification schemes whose majority use multi-class classifiers, each one to discriminate among all the classes. We research EOE's advantages, challenges, solutions and time complexity for data stream classification. Also, we compare the classification accuracy and execution time of EOE on four famous datasets with four successful and state of the art data stream classification schemes. Theoretical analysis and empirical evidence reveal that ensemble of expert classifiers can offer faster training and updating and higher classification accuracy than many existing popular data stream classification algorithms. We expect these results to be of interest to researchers and practitioners because they suggest a simple but very elegant and effective alternative to existing classification schemes for data streams.

KEYWORDS

Mining data streams, Ensemble classifiers

[1] دانشکده مهندسي کامپیوتر، دانشگاه علم و صنعت ایران، پست الکترونیکي: z_mirzamomen@comp.iust.ac.ir
[2] دانشکده مهندسي کامپیوتر، دانشگاه علم و صنعت ایران، پست الکترونیکي: s_hashemi@iust.ac.ir
[3] دانشکده مهندسي کامپیوتر، دانشگاه علم و صنعت ایران، پست الکترونیکي: kangavari@iust.ac.ir