

Classifier Recommendation Using Data Complexity Measures

در این مقاله برای معرفی روش دسته بندی مناسب از معیارهای بررسی میزان پیچیدگی دیتاست استفاده شده است بصورت کلی ۴ دسته کلی از الگوریتم های کلاسیک مورد استفاده قرار میگیرد که عبارتند از:

- Artificial Neural Networks (ANN)
- Support Vector Machine (SVM)
- k-Nearest Neighbor (KNN) classifier
- Decision Trees (DT)

در این مقاله از ۱۴۱ دیتاست گوناگون استفاده شده است. روند کار را بصورت زیر میتوان خلاصه نمود
چهار دیتاست جدید میسازیم که هر کدام ۱۴۱ سطر دارند و ۲۲ ویژگی. در واقع هر سطر بیانگر یکی از دیتاستهاست و هر ویژگی یکی از معیارهای بررسی پیچیدگی است. بعنوان برچسب هر سطر دقت هر دسته بند قرار میگیرد. مثلا در جدول اول که مربوط به ANN است دقت دسته بندی دیتاست ها به کمک این دسته بند بعنوان برچسب قرار میگیرد و تا درخت تصمیم گیری این فرآیند تکرار میشود

در مرحله بعدی به کمک ۳ روش رگرسیون برچسب هایی که در مرحله قبل به دست آمده اند را پیش بینی میکنیم. این سه روش عبارتند از

- Random Forest
- Support Vector Regressors (SVR)
- Weighted k-Nearest Neighbor (DWNN)

نهایتا برای دیتاستی جدید، ابتدا ۲۲ ویژگی محاسبه میشود و بعد پیش بینی میشود که با استفاده از هر کدام از الگوریتمهای دسته بند انتظار میرود که به چه دقتی دست یابیم و هر روشی که بیشترین دقت را پیشنهاد کرد بعنوان روش منتخب بازگردانده میشود

معمولا روش Random Forest نتایج بهتری را بازمیگرداند و به ترتیب بعد از آن SVR و DWNN