

ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТРИЧНИХ ВІДСТАНЕЙ ПРИ ВИРІШЕННІ ЗАДАЧ КЛАСИФІКАЦІЇ

Годовиченко М.А.

Науковий керівник — проф. каф. «Інформаційних систем»,

докт. техн. наук. Антощук С.Г.

В процесі вирішення задач класифікації одними з найпоширеніших є методи, які засновані на функціях відстані. Найбільш часто використовуваний метод — метод K -найближчих сусідів. Ефективність цього методу безпосередньо залежить від обраного параметра K та обраної метрики. Нерівномірність розподілу ознак заважає аналітичному пошуку параметра K , а особливості навчальної і контрольної вибірок не дають можливості ефективного вибору метрик без наявності апріорної інформації [1].

Найчастіше, модифікації методу K -найближчих сусідів намагаються визначити та урахувати апріорну інформацію щодо вибірок (розподіл значень тощо) чи створюються для вирішення конкретної задачі класифікації. Такі методи важкі у реалізації та потребують знання апріорної інформації [2].

Автором даної роботи запропонована метрика інформативності, яка враховує геометричну близькість об'єктів в n -мірному просторі ознак, а також інформативність об'єктів навчальної вибірки, яка вираховується шляхом оцінки близькості об'єкту навчальної вибірки до інших об'єктів свого класу і оцінки віддаленості від об'єктів іншого класу. Дана метрика використовує Евклідову метрику як міру відстані між об'єктами та на її основі формує міру інформативності.

Дана метрика дозволяє істотно понизити вплив параметра K на ефективність методу K -найближчих сусідів і підвищує ефективність самого методу.

1. Ту Дж., Гонсалес Р. Принципы распознавания образов. – М.: Мир, 1978. – 411 с.
2. Патрик Э. Основы теории распознавания образов. – М.:Советское радио, 1980. – 864 с.