Algoritam k-najbližih susjeda i njegove unaprijeđene inačice za prepoznavanje uzoraka

Radoš, Marin

Undergraduate thesis / Završni rad

2015

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:200:212400

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-28**

Repository / Repozitorij:

Faculty of Electrical Engineering, Computer Science and Information Technology Osijek

zir.nsk.hr

Image not found or type unknown

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

ELEKTROTEHNIČKI FAKULTET

Sveučilišni studij

Algoritam k-najbližih susjeda i njegove unaprijeđene inačice za prepoznavanje uzoraka

Završni rad

Marin Radoš

Osijek, 2015.

Sadržaj

1. UVOD ... 1
1.1. Zadatak završnog rada ... 1
2. KLASIFIKACIJA UZORAKA ... 2
2.1. Neke primjene klasifikacije ... 3
2.1.1. Primjena u biologiji ... 3
2.1.2. Primjena u medicini ... 3
2.1.3. Primjena u računalstvu ... 4
2.2. Česti postupci klasifikacije ... 4
2.2.1. Naivni Bayesov klasifikator ... 4
2.2.2. Umjetna neuronska mreža ... 5
2.2.3. Metoda potpornih vektora ... 5
2.3. Klasifikator k-najbližih susjeda ... 6
2.3.1. Određivanje najbližih susjeda ... 8
2.3.2. Određivanje klase pomoću većine ... 8
2.3.3. Određivanje klase pomoću dodjeljivanja težine putem udaljenosti ... 9
2.4. Unaprijeđene inačice k-najbližih susjeda klasifikatora ... 10
2.4.1. Unaprjeđenje funkcije udaljenosti ... 10
2.4.2. Unaprjeđenje dodavanjem težine atributima ... 11
2.4.3. Unaprjeđenje normalizacijom podataka ... 12
3. USPOREDBA KVALITETE KLASIFIKACIJE UNAPRIJEĐENIH KLASIFIKATORA K-NAJBLIŽIH SUSJEDA ... 13
3.1. Prikaz eksperimenta ... 13
3.1.1. Skupovi podataka korišteni u eksperimentu ... 14
3.1.2. Računalna okolina u kojoj se provodi testiranje ... 15
3.2. Prikaz izgrađene testne okoline ... 15
3.2.1. Prikaz važnih dijelova koda ... 16
3.2.2. Prikaz rezultata u testnoj okolini ... 17
3.3. Prikaz, usporedba i analiza rezultata ... 18
3.3.1. Mjerenja za algoritam k-najbližih susjeda s Euklidskom udaljenosti ... 19
3.3.2. Mjerenja za algoritam k-najbližih susjeda s Manhattan udaljenosti ... 20
3.3.3. Mjerenja za algoritam k-najbližih susjeda s dodijeljenom težinom ... 21
3.3.4. Mjerenja za algoritam k-najbližih susjeda s normalizacijom podataka ... 22
3.3.5. Grafička usporedba rezultata ... 23
4. ZAKLJUČAK ... 27
LITERATURA ...
SAŽETAK ...
ŽIVOTOPIS ...
PRILOZI ...

ALGORITAM K-NAJBLIŽIH SUSJEDA I NJEGOVE UNAPRIJEĐENE INAČICE ZA PREPOZNAVANJE UZORAKA

Sažetak. U ovom radu prikazan je problem klasifikacije uzoraka pomoću algoritma k-najbližih susjeda, te unaprjeđenje tog algoritma raznim načinima. Algoritam k-najbližih susjeda uspoređen je s njegovim unaprijeđenim inačicama na temelju eksperimentalnih rezultata dobivenih mjerenjem točnosti klasifikacije u različitim uvjetima, s različitim skupovima podataka i brojem susjeda k. Određeni su najveći nedostatci standardnog algoritma k-najbližih susjeda kroz funkciju udaljenosti, dodjeljivanje jednake važnosti svim atributima, te raspon veličina atributa. Ti nedostatci su riješeni u obliku unaprjeđenja algoritma drugačijom funkcijom udaljenosti, u ovom radu Manhattan, dodavanjem težine važnijim atributima, te normalizacijom skupa podataka. Kroz sva tri unaprjeđenja vidljivo je povećanje točnosti klasifikacije kroz dobivene eksperimentalne rezultate.

Ključne riječi: euklidska udaljenost, klasifikacija, najbliži susjedi, normalizacija

K-NEARAEST NEIGHBORS ALGORITHM AND ITS IMPROVED VERSIONS FOR PATTERN RECOGNITION

Abstract. In this paper, the problem of classification was shown using the k-nearest neighbors algorithm and its improvements through various ways. K-nearest neighbors algorithm was compared with its improved variants based on the experimental results obtained by measuring classification accuracy in different conditions, with different data sets and the number of neighbors. Biggest disadvantages of the standard k-nearest neighbors algorithm were defined through a distance function, assigning equal importance to all the attributes, and size range of the attributes. These shortcomings are addressed in the form of improving the algorithm with different distance function, in this paper, Manhattan, adding weight to important attributes, and normalization of the data set. Through all three improvements, an increase in the classification accuracy is shown through the experimental results.

Keywords: euclidean distance, classification, nearest neighbors, normalization