Tuesday, 23 May 2023 4:33 PM

It is obvi�ous that the running times of Alpha-investing, Fast-OSFS, OSFS and SAOLA algorith
are shorter than that of OFS�A3M, OFS-Gapknn and RHOFS algorithms. This is mainly becau
the first four algorithms mostly perform statistical comparison methods with low time complexity to calculate the relevance between features

(Run time)

1- By exploring the class separability in the boundary region of rough hypercuboid appro integrated feature evaluation criterion is proposed by examining not only the explicit contained in the positive region but also the useful implicit patterns derived from the region
2- onsider the explicit patterns of classification while ignoring the use of implicit pattern the discriminating ability of a feature or a subset of features. The ignorance of implicit these methods may cause the absence of useful information which may lead to degra prediction performance.
3- Don't need pre-knowlwdge of dataset
4- Handles scale and variety
5- Get better with more dataset present. But it will also get slow as it will more data as r further analysis
6- Handles slowness by distributed computing on spark
7- Based on the information provided in the context, it does not seem like the RHDOFS a requires prior knowledge of the dataset. The algorithm is designed to handle streamin dynamic environment, where the full feature space is unknown in advance and featur streams. The algorithm uses an incremental iterative approach to select relevant and redundant features in an online fashion. Therefore, it is designed to work without pri of the dataset.
8- It performs better or comparable to methods like soala

ms
se

ach, a novel
patterns
boundary

s to measure
patterns in
dation in the

dd for

lgorithm
g data in a
es arrive in
non-
or knowledge

9- the proposed RHOFS method is inferior to Fast-OSFS and OSFS algorithms, but outperf
investing, SAOLA and OFS-A3M algorithms in terms of the average cardinality of featu
and the average rank.

10- Compared with OFS-Gapknn, RHOFS can achieve more compact subsets on most of th
although the average number of features of RHOFS is larger.

Working on the Rough Hypercuboid Approach based Online Feature Selection (RHOFS) i
implementing the methodology described earlier to perform dynamic feature selection
learning tasks. The process starts with initializing a seed set of features and gradually ad
removing features based on their contribution to the hypercuboids constructed around
instances. The RHOFS algorithm continuously updates the feature set as new data arrive
maintain a concise and relevant feature representation.

However, it's important to note that RHOFS does have its complexity and drawbacks. Or
arises from the evaluation of feature contributions to the hypercuboids, which can be co
intensive, especially when dealing with high-dimensional datasets. Additionally, the onli
RHOFS requires constant monitoring and updating, which can add computational overhe

Another drawback of RHOFS is the sensitivity to the initial seed set. Depending on the in
the algorithm may converge to different feature subsets, leading to potential variations
This sensitivity makes the initialization process critical and may require domain knowled
preprocessing steps to ensure a suitable seed set.

Furthermore, RHOFS assumes that the hypercuboids accurately represent the data distr
data contains complex patterns or non-linear relationships, the hypercuboids might not
nuances effectively, potentially leading to suboptimal feature selection.

Lastly, RHOFS may face challenges when dealing with highly imbalanced datasets or whe
instances introduce concept drift. The algorithm's ability to adapt to evolving data may l
the selected feature set might not adequately represent the changing characteristics of

Considering these complexities and drawbacks, it is important to carefully assess the sui
for specific applications and datasets. Alternative feature selection methods and comple
techniques, such as dimensionality reduction or ensemble approaches, can be considere
limitations and enhance the performance of RHOFS in practical scenarios.

forms Alpha-
re subset

e datasets,

nvolves
in machine
lding or
incoming
es, aiming to

ne complexity
omputationally
ne nature of
ead.

nitial selection,
in performance.
lge or additional

ribution. If the
capture these

en the incoming
be hindered, and
the dataset.

itability of RHOFS
ementary
ed to address