

实验数据结果

日期: September 7, 2020

目录

按照老师的要求，又新找了一些数据集进行了实验，下面是实验的结果：

数据集名称	特征维度
kdda	20216830
avazu-app	1000000
news20	62061
sector	55197
smallNORB	18432

实验结果：

KDDA dataset(过去的):

method	k	b	c	acc	feature dimension
origin	\	\	\	88.438	20216830
pminhash	32	3	1	86.781	
\	32	2	1	86.781	
\	128	3	1	86.798	
\	256	3	1	87.066	
	256	6	1	87.884	
\	512	3	1	87.494	4096

avazu-app dataset:

method	k	b	c	acc	feature dimension
origin	\	\	\	99.334	1000000
pmh	128	3	1	100.000	
pmh	256	3	1	99.999	4096

精度太高，看不出来区别，就没有进行下去。

news20 dataset:

method	k	b	c	acc	feature dimension
origin				85.374	62061
pmh	128	6	1	71.475	8192
pmh	256	6	1	79.539	16384
pmh	512	6	1	82.87	32768
pmh	1024	6	1	84.222	65536
pmh	2048	6	1	85.249	131072
pmh	4096	6	1	85.274	262144
pmh	8192	6	1	86.125	524288
pmh	8192	3	1	83.897	65536

sector dataset:

method	k	b	c	acc	feature dimension
origin				93.701	55197
pmh	128	6	1	84.783	8192
pmh	256	6	1	88.618	16384
pmh	512	6	1	90.61	32768
pmh	1024	6	1	91.394	65536
pmh	2048	6	1	92.017	131072
pmh	4096	6	1	92.266	262144
pmh	8192	3	1	91.175	65536

smallNORB dataset:

method	k	b	c	acc	feature dimension
origin				78.436	18432
pmh	128	6	1	45.3415	8192
pmh	256	6	1	56.004	16384
pmh	512	6	1	64.818	32768