FOCUS-RRK FEATURE SELECTION TECHNIQUE FOR ENHANCING THE ACCURACY IN SOCIAL NETWORK DATA

Rajkumar Ramasamy*

RVS Nagar, Kajamalai Post, Tiruchirapalli, Tamilnadu, India.

*Corresponding Author
Rajkumar Ramasamy
RVS Nagar, Kajamalai
Post, Tiruchirapalli,
Tamilnadu, India.

ABSTRACT

Feature Selection is an effective technique in reducing the dimensionality of features in many applications where Online Social Network's datasets involve hundreds or thousands of features. The raw data are usually with inferior qualities such as redundancy and unreliability. Mining raw data for a given task has proved to be deteriorating in efficiency. There are many algorithms in data mining which can be used in dealing with redundant or missing data. Before the mining process begins those instances with missing value are considered insignificant to the result and thus can be eliminated from the dataset. The purpose of this research paper is twofold: Identifying and removing the redundant features with their latest solutions for Consistency Measure in an Online Social Network.

KEYWORDS: Online Social Network, Redundant features, Consistency.

1. INTRODUCTION

Online Social Network represents an emerging area which also brings many challenges and research Opportunities besides numerous socializing facilities. This research paper focuses on the Consistency Measure. By definition, data mining is a process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other repositories. Large parts of those data are with inferior qualities such as redundancy and unreliability. Data mining implements various useful information by looking at a small fraction of a large amount of data.

2. ONLINE SOCIAL NETWORK

Online Social Network like Facebook, Twitter, LinkedIn etc., became the popular interaction, recreation and Socializing facility on the net. Users pick more engaging sites, anyplace they may note acquainted faces like friends, relatives or colleagues.

2.1 Social media vs. Social networking

It's fascinating to notice that though individuals use Social networking and Social media interchangeably, they are not exactly the identical. Social networking will share some of the common Social media terminology. However, there's a small difference between those types of Online Communication. Social networking is an outlet to make online connection with other people, while Social media provides a venue for broadcasting facts. LinkedIn is an example of a Social networking site, whereas YouTube and Flicker are Social media websites.

Various existing algorithms and techniques have lot of drawbacks with lesser accuracy over the one of a kind datasets. consequently, in the present work, the use of Consistency measuring for removing feature selection for improves overall data quality for Online social network.

3. FEATURE SELECTION

As part of the data preprocessing procedure feature selection uses certain algorithms preserve only useful features and remove redundant and noisy information, at the same time enhancing efficiency without significantly reducing the accuracy of the classifier.

3.1 Feature Subset Selection

Figure 3.1 The Basic Structure of Feature Selection

The figure shows the basic structure of feature subset selection can be analyzed the practice of identifying and removing as lot of inappropriate and unnecessary features as achievable. This is for the reason that, redundant features do not redound to receiving a better analysis.

4. REDUNDANT FEATURES

To eliminate the redundant features from the dataset, the redundant features is removed via constructing the minimal spanning tree and separate the tree having the edge distance more than its neighbor to form the separate clusters, from the clusters features which might be strongly associated with the target features are selected to form the subset of features. Many feature subset selection algorithms, a few can efficaciously do away with irrelevant features however fail to address redundant features yet some of others can eliminate the irrelevant whilst taking care of the redundant feature.

Relief method

Relief is ineffective at removing redundant features as two predictive but notably correlated features are in all likelihood each to be pretty weighted. Relief cannot eliminate the redundant features. Relief-F- extends Relief, permitting this method to work with noisy and incomplete data sets and to deal with multiclass problems.

Koller and sahmi method

Koller and sahmi method eliminate any feature that doesn't contribute any additional information to the relaxation of the features.

Correlation based Feature Selection algorithm

The later removes redundant features from relevant ones via choosing representatives from classifier results and thus produces the final subset using CFS algorithm. The redundant feature elimination is a piece of sophisticated. Redundant features are assembled and representative feature may be taken out of the classifier results. Hence, notions of feature redundancy are commonly in terms of feature correlation and feature target concept correlation. As a result, most effective a totally small numbers of discriminative features are selected.

5. SEARCH PROCESS

The search process typically specializes in three aspects: where to begin, how to produce the primarily subsequent candidate subset and when to stop. Based on these elements, the exhaustive search generates every feasible combination of subsets. Search techniques are very crucial. A good search method can lessen the computational cost while enhancing the accuracy at the same time.

6. Exhaustive Search

Consistent with exceptional looking nature, the exhaustive search uses the algorithm to generate every possible combination of feature subset and compute Consistency rate for each of them. A threshold is setup at the start in step with the Consistency rate calculated for the first selected feature subset which is also set to be the current best subset. As the algorithm proceeds, the current best subset may be replaced by using the one with higher Consistency rate and smaller size of feature set wherein case the threshold is also replaced. The exhaustive search can start with either a set with one feature and carry out by adding features in to set or with a complete feature set and subtract features form it. It's evident that exhaustive search is time consuming and computational expensive as it calculates every combination many of which may be redundant and could be avoided if a learning process is involved.

An example for the exhaustive search is FOCUS RRK. Under is the exhaustive search algorithm in which Concal() is a function to calculate the Consistency rate for a given dataset D. S is the full feature set for the D.
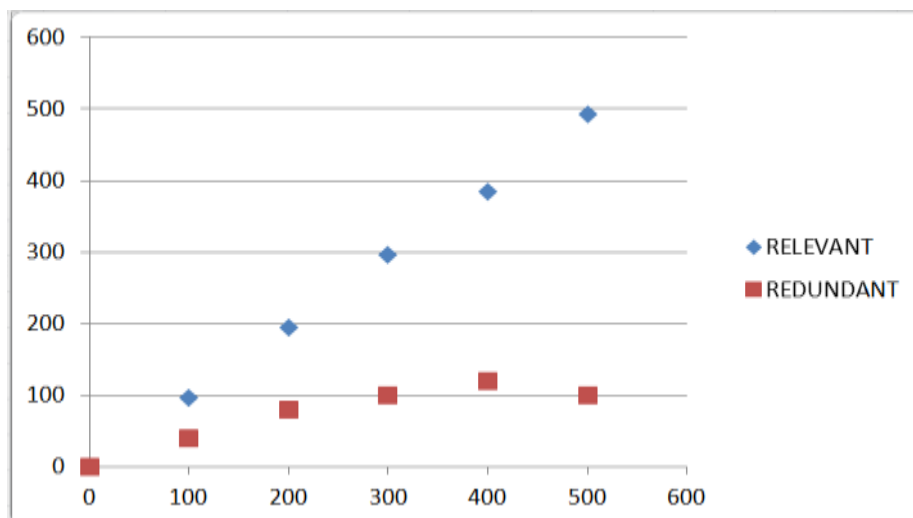
Algorithm

Input: Data D, Feature set fs

Output: Consistent Feature Subset.

1. BestSet = $\lceil$
2. BestConsistencyRate= $-\infty$
3. For i= 1 to |fs|
4. Begin
5. $\forall$fs' where |fs'| = i and fs' $\in$ fs
6. $CCONR$ = ConCal (fs',D)
7. if $CCONR$ $\geq$ BestConsistency Rate
8. if |BestSet| $<$ |fs|
9. BestSet = fs'
10. BestConsistencyRate = $CCONR$
11. End
12. Return BestSet

7. DATA SET

	Userid	name	gender	dob	address	mobile	email	pwd	id	join_date	snw
DELETE	1	sharmi	Female	14.Jan.1993	aarathangi	8344526894	sharmi@gmail.com	sharmi	BPKfUZ	27:03:2016	FaceBook
DELETE	101	rams	Female	2.Nov.1992	madurai	9655446994	rams@gmail.com	rams	GBJL7m	27:03:2016	LinkedIn
DELETE	201	Anjali	Female	4.Mar.1982	London	7654390823	anjali@gmail.com	anjali	fiZqN9	31:03:2016	Twitter
DELETE	301	simbu	Male	4.Jan.1992	puthupatti	9977886655	simbu@gmail.com	simbu	3nVuAS	15:04:2016	FaceBook
DELETE	401	simbu	Male	4.Jan.1992	puthupatti	9977886655	simbu1@gmail.com	simbu1	qUf7ty	15:04:2016	Twitter
DELETE	501	simbu	Male	4.Jan.1992	puthupatti	9977886655	simbu2@gmail.com	simbu2	nFMdyG	15:04:2016	LinkedIn
DELETE	601	keerthi	Female	2.Jun.1992	puthukotti	9876543210	keerthi@gmail.com	keerthi	nFKRtN	27:03:2016	LinkedIn
DELETE	701	nivetha	Female	25.Aug.1993	thanjavur	8973946796	nivetha@gmail.com	nivetha	KWj85c	28:03:2016	FaceBook
DELETE	801	mounika	Female	7.Jan.1992	madathahalli	9750060023	mounika1@gmail.com	mounika1	Az4kJH	16:04:2016	Twitter
DELETE	901	manju	Female	1.May.1993	manarkudi	9988776645	manju@gmail.com	manju	gQMjVx	13:04:2016	Twitter
DELETE	1001	sharmi	Female	15.Jan.1992	madathahalli	7750724707	sharmi1@gmail.com	sharmi1	k19dZK	14:04:2016	Twitter
DELETE	1101	sharmi	Female	16.Jan.1991	salem	8887722334	sharmi2@gmail.com	sharmi2	1JAPXn	14:04:2016	LinkedIn
DELETE	1201	keerthi	Female	3.Jun.1992	ooty	9987655432	keerthi1@gmail.com	keerthi1	tovrGl	14:04:2016	FaceBook
DELETE	1301	rams	Female	3.Nov.1993	bommidi	8833447721	rams1@gmail.com	rams1	q7jR2O	14:04:2016	FaceBook
DELETE	1401	Roshan	Female	12.Sep.1993	nammakal	9090808077	Roshan1@gmail.com	Roshan1	3ZIML6	14:04:2016	FaceBook

7.1 *FOCUS RRK Algorithm Result*

The *FOCUS RRK* algorithm obtains features relevant to the target concept by eliminating redundant features.

CONCLUSION

We present that greater concept on feature subset selection, using *FOCUS RRK* algorithm for calculate measuring for consistency on social network dataset. The process of *FOCUS RRK* algorithm to identifying and remove the redundant features and measuring for consistent rate.

There are many exhaustive search strategies in exercise which includes $FOCUS\ RRK$ is most time efficient, close to optimal. $FOCUS\ RRK$ Algorithm is used while there may be no known manner to locate an optimal solution in which case the goal is to expand an easy process with provable better running time and an improved solution. Also $FOCUS\ RRK$ algorithm removing the redundant features and targeted data result can be achieved.

REFERENCES

1. Vipin kumar and Sonajharia minz. Feature Selection: A literature Review., June 2014.

2. Velavan.P and Subashini.S. Correlation Based Feature Selection., April 2014; 3.

3. Ghazi Fuad Khamis and Ramadas Naik.T., Identifying and removing irrelevant and redundant features in high dimension data using feature subset., 2: 2015.

4. Harshali. D and Gangurde., Feature Selection using Clustering Approach for Big Data. ITCCE., 2014.

5. Jiliang Tang and uan Liu. Feature Selection for Social Media Data. ACM., Oct 2014; 8.