ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 1, January 2018

COST-BENEFIT ANALYSIS OF PRUNING BASED DATA MINING IN DISTRIBUTED DEMOGRAPHIC DATABASES

S.S.G.KRISHNAN
1Assistant Professor, Department Of Computer Science And Engineering, Sriram Engineering College,, India.

ssgkrishnan@gmail.com

Abstract

A general awareness about the over-whelming and geometric growth of the nodes in distributed demograhic databases is the need of the day. There can be an achievement of tremendous advantages on the part of the participating nodes if they simple fine-tune them by quickly reacting to the adverse effects of the scenario. All that is needed is any one or combination of the three aspects, namely natural pruning, intelligent pruning and adaptive behaviour of the participating nodes. The direct impact of the contribution of the nodes is a manifold, including a balance between demographic burst and intelligent / adaptive behaviour of the participating nodes.

Key Terms

Demographic Databases, Natural Pruning, Intelligent Pruning, Adaptive Behaviour, Participating Nodes.

Predictive analytics

Predictive analytics encompasses a variety of statistical techniques from predictivemodelling, machine learning, and datamining that analyze current and historical facts to make predictions about future or otherwise unknown events.

In business, predictive models exploit patterns found in historical and transactional data to identify risks and opportunities. Models capture relationships among many factors to allow assessment of risk or potential associated with a particular set of conditions, guiding decisionmaking for candidate transactions.

The defining functional effect of these technical approaches is that predictive analytics provides a predictive score (probability) for each individual (customer, employee, healthcare patient, product SKU, vehicle, component, machine, or other organizational unit) in order to determine, inform, or influence organizational processes that pertain across large numbers of individuals, such as in marketing, credit risk assessment, fraud detection, manufacturing, healthcare, and government operations including law enforcement.

Predictive analytics is used in actuarialscience, marketing, financial services, insurance, telecommunications, retail, travel, mobility, healthcare, childprotection, pharmaceuticals, capacityplanning and other fields.

One of the best-known applications is creditscoring, which is used throughout financialservices. Scoring models process a customer's credit history, loan application, customer data, etc., in order to rank-order individuals by their likelihood of making future credit payments on time.

Definition

Predictive analytics is an area of statistics that deals with extracting information from data and using it to predict trends and behavior patterns. Often the unknown event of interest is in the future, but predictive analytics can be applied to any type of unknown whether it be in the past, present or future. For example, identifying suspects after a crime has been committed, or credit card fraud as it occurs. The core of

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 1, January 2018

predictive analytics relies on capturing relationships between explanatoryvariables and the predicted variables from past occurrences, and exploiting them to predict the unknown outcome. It is important to note, however, that the accuracy and usability of results will depend greatly on the level of data analysis and the quality of assumptions. [4] discussed about a method, End-to-end inference to diagnose and repair the data-forwarding failures, our optimization goal to minimize the faults at minimum expected cost of correcting all faulty nodes that cannot properly deliver data. First checking the nodes that has the least checking cost does not minimize the expected costin fault localization. We construct a potential function for identifying the candidate nodes, one of which should be first checked by an optimal strategy. We proposes efficient inferring approach to the node to be checked in large-scale networks.

Predictive analytics is often defined as predicting at a more detailed level of granularity, i.e., generating predictive scores (probabilities) for each individual organizational element. This distinguishes it from forecasting. For example, "Predictive analytics—Technology that learns from experience (data) to predict the future behavior of individuals in order to drive better decisions." In future industrial systems, the value of predictive analytics will be to predict and prevent potential issues to achieve near-zero break-down and further be integrated into prescriptiveanalytics for decision optimization. Furthermore, the converted data can be used for closed-loop product life cycle improvement which is the vision of the Industrial Internet Consortium.

Predictive Analytics Process

Predictive Analytics Process

1. **Define Project** : Define the projectoutcomes, deliverable, scope of the effort, business objectives, identify the data sets that are going to be used.
2. **Data Collection** : Data mining forpredictive analytics prepares data from multiple sources for analysis. This provides a complete view of customer interactions.
3. **Data Analysis** : Data Analysis is theprocess of inspecting, cleaning and modelling data with the objective of discovering useful information, arriving at conclusion
4. **Statistics** : Statistical Analysisenables to validate the assumptions, hypothesis and test them using standard statistical models.
5. **Modelling** : Predictive modellingprovides the ability to automatically create accurate predictive models about future. There are also options to choose the best solution with multi-modal evaluation.
6. **Deployment** : Predictive modeldeployment provides the option to deploy the analytical results into everyday decision making process to get results, reports and output by automating the decisions based on the modelling.
7. **Model Monitoring** : Models aremanaged and monitored to review the model performance to ensure that it is providing the results expected.

Types

Generally, the term predictive analytics is used to mean predictive modeling, "scoring" data with predictive models, and forecasting. However, people are increasingly using the term to refer to

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 1, January 2018

related analytical disciplines, such as descriptive modeling and decision modeling or optimization. These disciplines also involve rigorous data analysis, and are widely used in business for segmentation and decision making, but have different purposes and the statistical techniques underlying them vary.

Demographic data:

Demographic data refers to data that is statistically socio-economic in nature such as population, race, income, education and employment, which represent specific geographic locations and are often associated with time. For example, when referring to population demographic data, we have characteristics such as areapopulation, population growth or birthrate, ethnicity, density and distribution. With regard to employment, we have employment and unemployment rates, which can be related further to gender and ethnicity.

Techopedia explains *Demographic Data*

Demographic data is often gathered by census organizations, both government and private, which may use the data for research, marketing, and environmental and human development. Data such as population and employment and all their related data fields such as density, ethnicity and gender can be used by the government to plan for infrastructure development such as roads, hospitals and law enforcement.

Demographic data is also used by marketers and business entities for targeted advertising and product distribution. For example, in areas with a large population density of Latinos, fast food restaurants often offer Mexican-themed foods as opposed to Mediterranean ones. Branches of famous American fast food companies in foreign countries often tailor their menu according to the local taste. All of these are due to demographic research using demographic data.

India is the second most populated country in the world with nearly a fifth of the world'spopulation. According to the 2017 revision of the World Population Prospects[1], the population stood at 1,324,171,354.

During 1975–2019 the population doubled to 1.2 billion. The Indian population reached the billion mark in 1998. India is projected to be the world's most populous country by 2022, surpassing the population of China. It is expected to become the first political entity in history to be home to more than 1.5
billion people by 2030, and its population is set to reach 1.7 billion by 2050.[5][6] Its population growth rate is 1.2%, ranking 94th in the world in 2013.

India has more than 50% of its population below the age of 25 and more than 65% below the age of 35. It is expected that, in 2020, the average age of an Indian will be 29 years, compared to 37 for China and 48 for Japan; and, by 2030, India's dependencyratio should be just over 0.4.

Population within the age group of 0–6

Population between age 0–6 by state/union territory[71]

State /UT Code	State/ UT	Total	Male	Female
1	Jammu and Kashmir	2,008,670	1,080,662	927,982
2	Himachal Pradesh	763,864	400,681	363,183

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 1, January 2018

3	Punjab	2,941,570	1,593,262	1,348,308
4	Chandigarh	117,953	63,187	54,766
5	Uttarakhand	1,328,844	704,769	624,075

Population between age 0–6 by state/union territory[71]

State /UT Code	State/ UT	Total	Male	Female
6	Haryana	3,297,724	1,802,047	1,495,677
7	Delhi	1,970,510	1,055,735	914,775
8	Rajasthan	10,504,916	5,580,212	4,924,004
9	Uttar Pradesh	29,728,235	15,653,175	14,075,060
10	Bihar	18,582,229	9,615,280	8,966,949
11	Sikkim	61,077	31,418	29,659
12	Arunachal Pradesh	202,759	103,430	99,330
13	Nagaland	285,981	147,111	138,870
14	Manipur	353,23	182,68	170,55

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 1, January 2018

Population between age 0–6 by state/union territory[71]

State /UT Code	State/ UT	Total	Male	Female
		7	4	3
15	Mizoram	165,536	83,965	81,571
16	Tripura	444,055	227,354	216,701
17	Meghalaya	555,822	282,189	273,633
18	Assam	4,511,307	2,305,088	2,206,219
19	West Bengal	10,112,599	5,187,264	4,925,335
20	Jharkhand	5,237,582	2,695,921	2,541,661
21	Odisha	5,035,650	2,603,208	2,432,442
22	Chhattisgarh	3,584,028	1,824,987	1,759,041

Population between age 0–6 by state/union territory[71]

State /UT Code	State/ UT	Total	Male	Female
23	Madhya Pradesh	10,548,295	5,516,957	5,031,338
24	Gujarat	7,564,464	3,974,286	3,519,890
25	Daman and Diu	25,880	13,556	12,314
26	Dadra and Nagar Haveli	49,196	25,575	23,621
27	Maharashtra	12,848,375	6,822,262	6,026,113
28	Andhra Pradesh	8,642,686	4,448,330	4,194,356
29	Karnataka	6,855,801	3,527,844	3,327,957
30	Goa	139,495	72,669	66,826

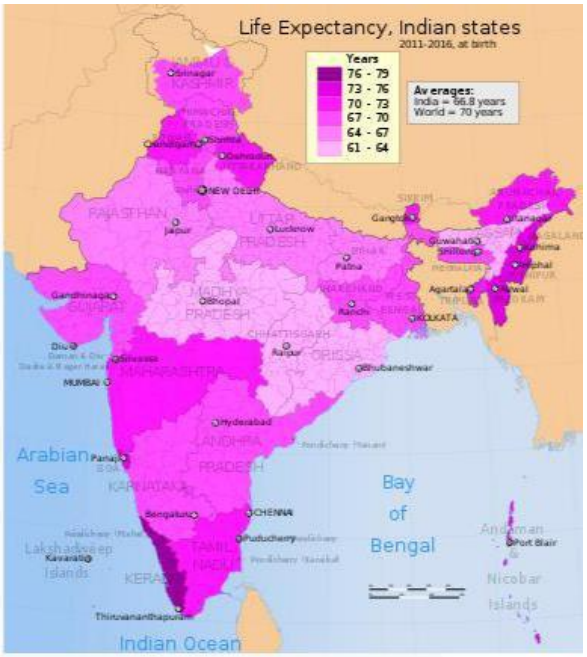
ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 1, January 2018

Population between age 0–6 by state/union territory[71]

State /UT Code	State/ UT	Total	Male	Female
31	Lakshadweep	7,088	3,715	3,373
32	Kerala	3,322,247	1,695,889	1,626,358
33	Tamil Nadu	6,894,821	3,542,351	3,352,470
34	Puducherry	127,610	64,932	62,678
35	Andaman and Nicobar Islands	39,497	20,094	19,403
–	**Total (India)**	**158,789,287**	**82,952,135**	**75,837,152**

Population above the age of 7



Life expectancy map of India, 2011–2016.

Population above the age of 7 by state/union territory[71]

State/UT Code	State /UT	Total	Male	Female
1	Jammu and Kashmir	–	–	–
2	Himachal	–	–	–

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 1, January 2018

Population above the age of 7 by state/union territory[71]

State/UT Code	State /UT	Total	Male	Female
	Pradesh			
3	Punjab	–	–	–
4	Chandigarh	–	–	–
5	Uttarakhand	–	–	–
6	Haryana	22,055,357	11,703,083	10,352,274
7	Delhi	14,782,725	7,920,675	6,862,050
8	Rajasthan	58,116,096	30,039,874	28,076,222
9	Uttar Pradesh	169,853,242	88,943,240	80,910,002
10	Bihar	85,222,408	44,570,067	40,652,341

Population above the age of 7 by state/union territory[71]

State/UT Code	State /UT	Total	Male	Female
11	Sikkim	546,611	290,243	256,368
12	Arunachal Pradesh	1,179,852	616,802	563,050
13	Nagaland	1,694,621	878,596	816,025
14	Manipur	2,368,519	1,187,080	1,181,439
15	Mizoram	925,478	468,374	457,104
16	Tripura	3,226,977	1,644,513	1,582,464
17	Meghalaya	2,408,185	1,210,479	1,197,706
18	Assam	26,657,965	13,649,839	13,008,126

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 1, January 2018

Population above the age of 7 by state/union territory[71]

State/UT Code	State /UT	Total	Male	Female
19	West Bengal	81,235,137	41,740,125	39,495,012
20	Jharkhand	27,728,656	14,235,767	13,492,889
21	Odisha	36,911,708	18,598,470	18,313,238
22	Chhattisgarh	21,956,168	11,002,928	10,953,240
23	Madhya Pradesh	62,049,270	32,095,963	29,953,307
24	Gujarat	52,889,452	27,507,996	25,381,456
25	Daman and Diu	217,031	136,544	80,487
26	Dadra and Nagar Haveli	293,657	167,603	126,054

Population above the age of 7 by state/union territory[71]

State/UT Code	State /UT	Total	Male	Female
27	Maharashtra	99,524,597	51,539,135	47,985,462
28	Andhra Pradesh	76,022,847	38,061,551	37,961,296
29	Karnataka	54,274,903	27,529,898	26,745,005
30	Goa	1,318,228	668,042	650,186
31	Lakshadweep	57,341	29,391	27,950
32	Kerala	–	–	–
33	Tamil Nadu	65,244,137	32,616,520	32,627,617
34	Puducherry	1,116,854	545,553	571,301
35	Andaman and	340,447	182,23	158,21

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

Population above the age of 7 by state/union territory[71]

State/UT Code	State /UT	Total	Male	Female
	Nicobar Islands		6	1
–	**Total (India)**	**1,051,404,135**	**540,772,113**	**510,632,022**

Population projections

India is projected to overtake China as the world's most populous nation by 2030. India's population growth has raised concerns that it would lead to widespread unemployment and political instability. Note that these projections make assumptions about future fertility and death rates which may not turn out to be correct in the event. Fertility rates also vary from region to region, with some higher than the national average and some lower.

Source:

- **2020**: 1,326,093,000
- **2030**: 1,460,743,000
- **2040**: 1,571,715,000
- **2050**: 1,656,554,000

2020 estimate

In millions (example: 361 = 361,000,000)

Source:

Year	Under 15	15–64	65+	Total
2000	361	604	45	1010
2005	368	673	51	1093
2010	370	747	58	1175
2015	372	819	65	1256
2020	373	882	76	1332

Natural Pruning:

A cross-section of the geo-spatial nodes do not yield offspring nodes and hence they are naturally pruned.

Intelligent and Adaptive Pruning:

A quick survey of demopgraphics or population data above can cause an awareness among the entire human race to intelligently curtail yielding of their offspring node(s).

Adaptive Pruning:

Though embarrassing, it is a fact that our population is growing in a geometric progression. Hence it is the sole responsibility of all the participant nodes to adapt themselves to contribute towards the overall increase in the rate of pruning of the involved nodes to promote certain advantages as follows:-

Advantages of Natural Pruning / Intelligent and Adaptive Pruning across arbitrary nodes in Demographic and distributed databases:

a) Population balancing

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 1, January 2018

b) Improved access to resources

c) Better life-style

d) Enhanced allocation and sharing of global and native resources

e) Better employment opportunities

f) Improved literacy rate

g) Free / affordable medical assistance to local / regional community

h) Reduced environmental and atmospheric pollution

i) Improved hygiene

j) Global Longevity of the entire human race

k) Improved Understanding and harmony among citizens

l) Ability to balance between birthrates and workforce requirements in organizations

ACKNOWLEDGEMENTS

The author is immensely happy to have mingled with the Sriram Engineering College family and is highly grateful to the Senior and middle level management of the Sriram Group of Institutions.

REFERENCES

[1] The World-wide Web or www

[2]https://en.wikipedia.org/wiki/Demographics_of_India

[3] M. T. Özsu and P. Valduriez, *Principles ofDistributed Databases* (3rdedition) (2011)Springer, ISBN 978-1-4419-8833-1

[4] Christo Ananth, Mary Varsha Peter, Priya.M., Rajalakshmi.R., Muthu Bharathi.R., Pramila.E., "Network Fault Correction in Overlay Network through Optimality", International Journal of Advanced Research Trends in Engineering and Technology (IJARTET), Volume 2, Issue 8, August 2015, pp: 19-22

[5] Optimized Prune Based Data Mining (Opbdm) For Distributed Databases: An Adaptive Approach, G.Parthasarathy, Dr.D.C.Tomar, Research Scholar, Dept. Of Computer Science And Engineering, Sathyabama University, Chennai, India.
Senior Professor, Dept. Of Information Technology, Jerusalem College Of Engineering, Chennai, India. Journal of Theoretical and Applied Information Technology30th September 2014. Vol. 67 No.3 ISSN: 1992-8645