

A methodology for e-Readiness Computation using Divide, Predict and Conquer Approach

Gajendra Sharma, Deep Raj Sharma

Department of Computer Science and Engineering, Kathmandu University, Nepal
Corresponding author: gajendra.sharma@ku.edu.np

Received: Dec. 15, 2015 Revised: May 21, 2016 Accepted: July 15, 2016

Abstract: There exist some problems which seems partially similar to Divide and Conquer approach but can't be solved using this approach. Such type of problems can be divisible into sub problems of different size but the same mechanism may not be useful to solve all of the sub problems. While using the same mechanism to solve all of the sub problems they can be classified into two kinds, first having a solution, second having no solution. The proposed approach in this paper analyses the features of sub-problems that can be solved and finds a solution for the unsolvable problems using those features. Further, we advocate this method is suitable for computation of e-Readiness in developing countries like Nepal, who are bounded to invest a very low budget in Technological Researches.

Keywords: e-Readiness, Divide Predict and Conquer, Linear Regression, Clustering.

1. Introduction

Let us assume a problem P , such that it is decomposable into a set of micro sized sub-problems. Let us denote P as the set of subproblems as follows:

$$P = \{P_1, P_2, \dots, P_n\} \quad (i)$$

Let us consider another term called feature vector, $F = (f_1, f_2, \dots, f_n)$ which is a tuple, where an arbitrary f_i indicates an element of the features. This feature vector is broad-casted to each and every element of P . We have a subset of P , P_τ such that P_τ consists of subproblems which have a direct solution. Direct solution indicates that we have all the values for the features of an element e , which belongs to P_τ are present. The other term, we call P_Ω is also a subset of P such that $P_\Omega \cap P_\tau = \emptyset$ and P_Ω has no direct solution that is for the any element e , which belongs in P_Ω , we lack some features for it.

1.1 The Three Steps

Divide: A single problem P is decomposed into a set of small sub-problems, $\{P_1, P_2, \dots, P_n\}$. The elements of P_τ have a direct solution so they are solved.

Prediction: On the basis of solution obtained for elements of P_τ , a prediction is done to get the solution for the elements of P_Ω . Prediction uses statistical and mathematical techniques like Linear regression, Curve fitting, Interpolation etc.

Conquer: We achieve solutions for all the subproblems after using the Prediction part. The conquer part assembles solutions of all the subproblems providing a solution to our main problem P. Conquer part can be done using some mathematical model, which we call Hypothesis.

This paper is organized as follows. In Section 2, we discuss about the calculation of e-Readiness using Divide, Predict and Conquer approach. In Section 3, A literature review has been described in the field of calculation of e-Readiness which mainly focuses on CSPP e-Readiness method as well as APEC method. Section 4 deals with the proposed model. Section 5 gives a recursive algorithm to compute e-readiness. Section 6 describes the methodology, we proposed to be used. Finally, in Section 7 we have some discussion about e-Readiness computation.

2. Calculating e-Readiness of Nepal

Let us consider the problem of calculating e-readiness in the context of Nepal. When we visualize the problem with Divide, Predict and Conquer approach we have,

$P =$ Calculation of e-Readiness which is a set of subproblems of Calculation of e-Readiness of seven States at the present scenario. So, we have $P = \{P_1, P_2, \dots, P_7\}$. Similarly, a P_i where i belongs to $\{1, 2, 3, \dots, 7\}$, and is the state number, can be further divided into subproblems $P_i = \{P_{i1}, P_{i2}, \dots, P_{id}\}$, where d is the number of districts under i^{th} state. This process can be repeated recursively to a base case, where base case can be a VDC (Village Development Council), DDC (District Development Council) or even it can be more microscopic like a Municipality.

$P_i \subseteq P$, for every element of P_τ , we can calculate e-Readiness using the framework proposed in [2]. We assume the places indicated in Fig. 1 in numbers, as the elements of P. Further, $P_\Omega = P - P_\tau = P_\tau^c$ (Complement of P_τ).

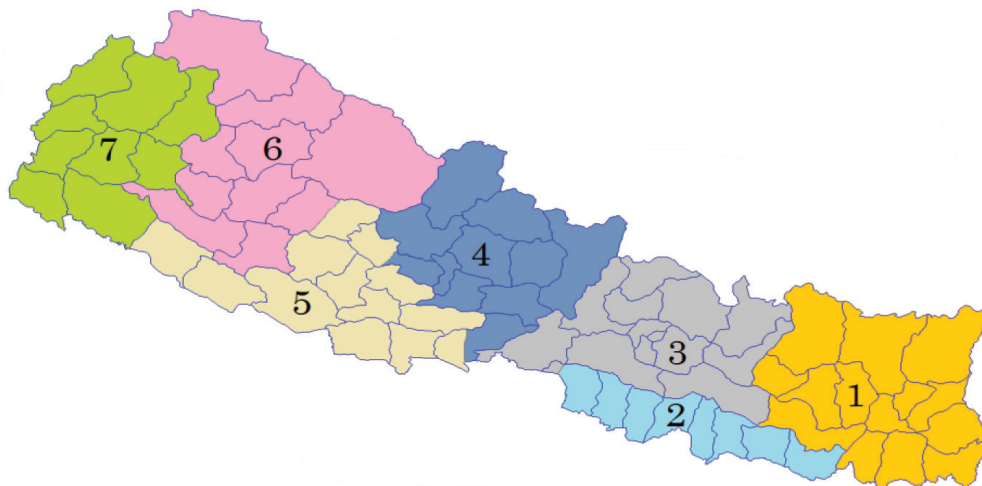


Fig. 1: Map of Nepal [6]

For the elements of P_Ω , we can use prediction models to calculate e-Readiness. Thus, we obtain e-Readiness for every element of P. Finally, as a hypothesis, we use weighted sum average of

e-Readiness obtained for different districts. Tung X. Bui et al. [9], has used- 'Knowledgeable citizens', 'Access to skilled Workforce', 'Macro economy', 'Digital infrastructures', 'Industry Competitiveness', 'Culture', 'Ability to Invest and Cost of living and Pricing' as the factors for computing e-Readiness. These eight factors has in overall 52 measures. For instance, the Knowledgeable citizens factor has 6 measures- 'Adult literacy rate', 'Secondary enrollment', 'Tertiary enrollment', '8th grade Achievement in Science', 'MGMT Education Available in first class business schools' and 'Flexibility of people to adapt to new challenges'. Our features sets are the vector of these 52 measures suggested by Tung X. Bui [9].

2.1 Motivation

It is essentially important to have e-Readiness index for the better development of modern businesses in Country. In the context of Nepal, the country lacks fresh e-Readiness index. In its absence, lots of businesses in country are taking place without actual projection of the need of goods. Beside business e- Readiness is an important factor for government too, for the development of an electronic society. Our main objective is the construction of a model to compute e-Readiness of Nepal.

2.2 Statement of the Problem

Nepal do not have a fresh e-Readiness index. Using divide predict and conquer approach such that it can calculate the latest approximate e-Readiness value with a very little investment is the major issue which this paper deals about.

3. Related Works

The number of tools that are ready-to-use to assess a nation's e-readiness is limited, as relatively few organizations have presented their assessment methods for use by others. However, there is a wide range of reports and other resources that can be re-worked into "assessment tools", including third-party reports, position papers and survey results. This list of e-readiness assessment tools starts with those formally presented as assessment methodologies with ready- to-use questionnaires, and then goes on to look at assessment methods that can be garnered from other resources. Some of the tools included measure access to and use of ICT as opposed to "e- readiness" perse, although most of the tools included here look at Internet and other ICT use within a broader economic, socio- political context. Many third-party assessments rank countries according to their measure of e-readiness, providing indicators for competitiveness by comparing countries using various measures. Some of these benchmarking reports have been included in this discussion of e-readiness reports, as the measures utilized in these reports are useful indicators for e-readiness, and contribute to a broader understanding of the field.

3.1 Ready to Use Tools

There are few Ready-to-use Tools freely available on the web. In this discussion, three of such tools have been identified:

The Computer System Policy Project's (CSPP)'s Readiness Guide for Living in the Networked World. A Guide for Developing Countries, Asian Pacific Economic Cooperation's (APEC)'s E-Commerce Readiness Assessment. Harvard University's Center for International Development

(CID)’s Readiness for the Networked World. These tools produce scores or ratings from which communities or countries are supposed to identify where they fit according to each tool’s definition of e-readiness [3].

CSPP readiness guide for Living in the networked world was published in 1998, and is available at www.cspp.org. CSPP is a “public policy advocacy group, comprised of the Chairman and Chief Executive Officers” of US information technology companies. This self-assessment tool is designed to help individuals and communities determine how prepared they are to participate in the “Networked World.” The guide measures the prevalence and integration of ICTs in homes, schools, businesses, health care facilities, and government offices, with additional focus on competition among access providers, speed of access, and government policy. They have divided the measurement into five Category: Infrastructure, Access, Applications and services, Economy; and “Enablers” (policy, privacy, security, ubiquity). The CSPP Readiness Guide provides a series of 23 questions, for community members to ask about the community itself. For each question, the users choose from a set of answers, which represent four progressive “stages” of development. The 23 questions are divided into the five categories listed above. For example, “chosed the level of access to network services provided in homes in your community. a) 25% of homes have a computer / access device.” The assessment produces a rating that indicates which of four progressive stages of development the community is at for each of the five categories listed above. “An over- all ‘score’ for the community can be estimated by simply averaging the scores across the criteria” [4].

3.2 APEC e-Readiness Assessment

The Asian Pacific Economic Cooperation (APEC) Electronic Commerce Steering Group developed this guide to help governments develop their own focused policies, adapted to their specific environment, for the healthy development of e-commerce.” Six categories are measured for “readiness for e-commerce:” Participants are asked 100 multiple-choice grouped into the six categories listed above. The possible answers indicate progressive levels of e-readiness for a country. No overall scoring occurs. The product of the assessment is the answers to the 100 questions. Countries are supposed to work on area with less than optimal answers, since they are “impediments” to the deployment of e-commerce” [1].

4. Model

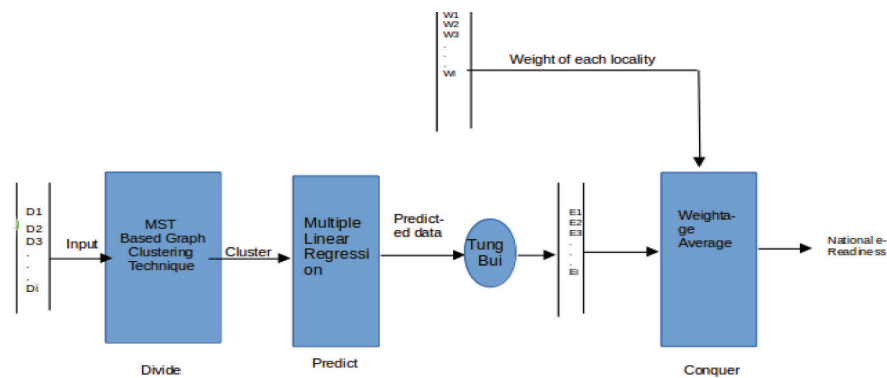


Fig. 2: Proposed Model

4.1 MST based Graph Clustering for Divide

An Euclidean minimum spanning tree (EMST) is a spanning tree of a set of n points in a metric space (E_n), where the length of an edge is the Euclidean distance between a pair of points in the point set. We first Construct MST by constructing the Graph of elements in P or P_i or even an element from decomposable problems of P_i . The Graph is constructed on the basis of Euclidean distance of features between the elements of the P or P_i . After the Construction of Graph we use some threshold to make the Clusters of the designed Graph as described by Bhaskar Adepu et al. [2]. The Clusters are the decomposition of the Problem. In our Scenario, we call it as Divide. This will create two set of subproblems, one kind of subproblems have a direct solution, other kind of subproblems needed to be sent through prediction part to get the desired approximate solution.

4.2 Multiple Linear Regression for Prediction

Multiple linear Regression can be used when we have more than one variable for prediction of a particular value. We use the following equation for prediction of the missing value [9]:

$$\theta = (X^T X)^{-1} X^T y \quad (1)$$

where X^T is the transpose of matrix X and X is a matrix developed using the features we used for formulation of Clusters.

A^{-1} is the inverse of matrix A .

y is a singular matrix developed with the set of features which is missing from a particular locality. From Normal equation-1, we calculate θ . After calculation of θ , we use the equation:

$$M_D = M_R \theta \quad (2)$$

where M_D is the missing feature from the features vector and M_R are the rest of the features from the same features vector.

4.3 Weightage average mean

Formally, the weighted mean of a non-empty set of data $\{x_1, x_2, \dots, x_n\}$ with non-negative weights is:

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \dots + w_n x_n}{w_1 + w_2 + \dots + w_n} \quad (3)$$

Using equation(3), we will get the final e-Readiness value.

5. Methodology

The block diagram of our methodology for e-Readiness calculation is shown in the Fig. 3 below. First two parts consists of Data Collection and Data Preprocessing. In the third part, we insert the data developed from preprocessing to our Model. Finally, we validate. If the validation is good enough, we publish the result otherwise we come to the conclusion that this sort of data is not applicable for our model. This is one of the drawback of our model. Removal of this drawback is our future work.

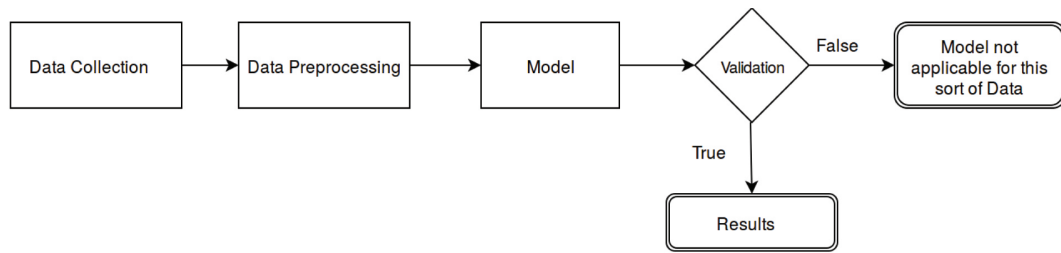


Fig. 3: Methodology

5.1 Data Collection

Survey is done for the collection of Data. We try to gather all sort of data for different parameters which are necessary for the e-Readiness Calculation. The data are gathered on the basis of proposed States of our Country.

5.2 Data Preprocessing

Today's real-world data are highly susceptible to noisy and inconsistent due to their typically huge size (often several gigabytes or more) and their likely origin from multiple, heterogeneous sources. Low-quality data will lead to low-quality results. There are several data preprocessing techniques. Data cleaning can be applied to remove noise and correct inconsistencies in data. Data integration merges data from multiple sources into a coherent data store. Data reduction can reduce data size by, for instance, aggregating, eliminating redundant features, or clustering. Data transformations (e.g., normalization) may be applied, where data are scaled to fall within a smaller range like 0.0 to 1.0 [7].

5.3 Model

We insert the data, after preprocessing, to the Model we developed. Our model predicts all the incomplete parameters and calculates the National e-Readiness. Before calculating the national e-Readiness, our model calculates the e-Readiness of all the districts then e-Readiness of all the states Recursively.

5.4 Validation

We validate our result using some predefined hypothesis. Predefined Hypothesis roughly analyses the expected e-Readiness value of different Districts and States. If the result follows the Hypothesis generated by us, we make a conclusion that our result is correct. In the other cases, we make a Conclusion that the data gathered do not favors our Model.

6. Algorithm

The algorithm is inherited from the divide and conquer approach. The divide-and-conquer strategy solves a problem by:

1. Breaking it into subproblems that are themselves smaller instances of the same type of problem

2. Recursively solving these subproblems
3. Appropriately combining their answers

We have inserted Predict part here. So, the skeleton of our algorithm becomes:

1. Break the problem into subproblems of two types, recursively: direct solvable and indirectly solvable.
2. Solve the solvable problem and Use prediction for indirectly solvable problems
3. Conquer the Result.

For calculation of e-Readiness, the algorithm is presented below:

Algorithm for National e-Readiness:

***Input:** Set of all available features of all the districts of all seven States.*

Algorithm:

For each State do Recursively:

Divide: Generate dissimilarity matrix from the features which are available for all the districts.

Using the dissimilarity matrix, generate a Graph and find it's minimum spanning Tree.

Use MST based Clustering Technique to divide the Problem into subproblem.

Predict: For each missing feature in nodes of each cluster do:

predict the missing feature using Linear Regression

Conquer: Merge the Output of all the Clusters using Weightage Mean.

***Intermediate Output:** e-Readiness of a State*

Merge the output of all the State using Weightage Mean.

***Final Output:** National e-Readiness.*

7. Conclusion

e-Readiness is one of the important factor for the digitizing Society. The calculation of e-Readiness deals with many kinds of Social aspects related with Economical matters, Cultural matters, Literacy rate, Poverty etc. We need to go through a huge mass of data and need to apply different mechanisms to get a correct result. When we talk about the important aspects of e-Readiness, mainly to be settled in any kind of society in this current modern world, we need to know exactly, how much the society is dipped into electronic acceptance. Then only, we can know how to make an impression in business strategies as well as other social related factors. The key factor for any nation for the implementation of e-Governance is the rate of acceptance of electronics in society. Thus, Government can predict how much at a particular moment is good to invest on e-Governance. This paper simply gives a methodology to calculate e-Readiness in the context of Nepal. Nepal, being a country with full Geographical diversity, it is really a difficult task to reach at every place of the country and do particular computations based on any kind of research. In the case of e-Readiness, this paper has given a new concept to calculate e- Readiness using some prediction. An algorithm is proposed which can be applied to calculate not the exact but an approximate figure for e-Readiness Parameter. For the future work, we will check the accuracy of the model proposed here.

References

- [1] APEC, 'e-commerce Readiness Assessment', The Asian Pacific Economic Cooperation (APEC,2000) <http://www.apecsec.org.sg>
- [2] Bhaskar A and Bejjanki KK (2011), A Novel Approach for Minimum Spanning Tree Based Clustering Approach, *International Journal of Computer Science and Engineering Systems*, **5**: 69-73.
- [3] Bridges.org (2001). Comparison of E-Readiness Assessment Models. <http://www.internetpolicy.net/readiness/ereadinesscomparison.pdf>, accessed on May 22, 2016
- [4] Bridges.org (2005b) eReady For What? E-Readiness in Developing Countries: Current Status and Prospects toward the Millennium Development Goals, http://www.infodev.org/infodev-files/resource/Infodev_Documents_3.pdf, accessed on May 21, 2016.
- [5] https://class.coursera.org/ml-005/lecture?lecture_player=flash, accessed on May 22, 2016.
- [6] <http://www.karma99.com/2015/04/proposed-federal-structures-of-nepal.html> Accessed on May 22, 2016.
- [7] Jiawei H, Micheline K and Jian P: *Data Mining Concepts and Techniques*, Third Edition; pp 77.
- [8] Kapoor JN and Saxena HC (2005), *Mathematical Statistics*, Published by S. Chand & Company Ltd, 2013 ISBN 10: 8121912466 / ISBN 13: 9788121912464.
- [9] Tung X (2003), A Framework for Measuring National E-Readiness. *International Journal of Electronic Business* **1**: 3-22.