

Application of Stochastic Gradient Descent Algorithm in Evaluating the Performance Contribution of Employees

R. M. V. S. Ratnayake¹, K. G. U Perera², G. S. K. Wickramanayaka³,
C. S. Gunasekara⁴, K. C. J. K. Samarawickrama⁵

¹(Senior Lecturer, Dept. of Textile & Clothing Technology, University of Moratuwa)

²⁻⁵(Dept. of Textile & Clothing Technology, University of Moratuwa)

Abstract: This paper presents the applicability of stochastic gradient descent algorithm as an approach for evaluation of performance contribution. Under that, performance evaluation is established in corporate, department and team levels with the intention of creating an employee incentive scheme.

Keywords: employee performance contribution, machine learning, stochastic gradient descent algorithm

I. Introduction

In development of an incentive scheme to any industry, main underlining consideration is how to identify contribution of departments and employee teams towards the overall company performance. It would not have a direct contribution relationship, especially in the case of indirect employees [1,2]. In addition, it is possible to have department or employee level KPIs (Key Performance Indicators) which are having a negative relationship towards the company's performance. Effective evaluation of performance contribution is challenging in such situations. There are no identifiable researches done on application of data driven approach like machine learning to overcome the above issues. Therefore, the purpose of this paper is to present the adaptability of stochastic gradient descent algorithm in evaluating the performance contribution.

II. Machine Learning

Machine learning is a class of algorithms which is data-driven. Main difference of these algorithms is the use of available data to decide a good solution for a problem. Except for using coded definitions, machine learning algorithms are used to learn by examples [3]. This quality can be illustrated with a human face recognition machine learning algorithm capable of identifying the human faces after it is trained with several samples [4,5].

Machine learning algorithm can be mainly either a supervised learning problem or an unsupervised learning problem. Training data of supervised learning problems are having examples of inputs corresponded with the target vectors/ output. Training data of an unsupervised learning problem does not consist such labelled inputs corresponded with the targets. Therefore unsupervised learning problems are used to identify groups of similar examples in the data set [3,6].

In case of evaluation of performance contribution, historic data of the performance measures of departments and employee teams can be used as training data. Corresponded overall performance of the company measured using suitable performance measures can be used as the targets for the training data set. It enables to consider the requirement as a supervised machine learning problem.

2.1 Supervised learning in machine learning

Andrew [7] has illustrated a supervised learning machine learning algorithm as in the Fig.1. Input values X are also called as features. There can be multiple X values simultaneously which represent different features. Y denotes the output which is the target variable of the problem. Since it is a supervised learning problem, pairs of (X,Y) becomes training examples to the problem.

Andrew [7] describes hypothesis 'H' as the predictor of the output Y which uses X values as input. The goal of a supervised learning problem is to find good predictor for the corresponding values of Y. If the target variable Y is continuous, it is further defined as a regression machine learning problem. In contrast, if the target variable is discrete, it is named as a classification problem.

In the case of incentive scheme development, relevant performance measures of company, departments and employee teams will become target variables and inputs. Therefore regression approach has to be taken in developing relationships for this type of a problem.

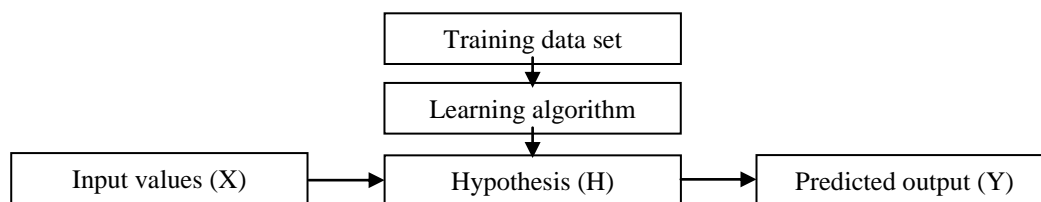


Fig1. Illustration of a supervised learning problem solving

2.2 Linear regression

Andrew [7] defines a simplest version of a hypothesis related with a supervised machine learning problem as;

$$H(x) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots$$

$$H(x) = \sum_{i=0}^n \theta_i X_i$$

All the θ_i s of the above hypothesis is named as parameters or as the weights. To find reasonable values for θ_i s, training data should be used and try to make H(x) closer to output (Y) values. To minimize the gap between H(x) and Y, a cost function can be introduced under ordinary least squares method as;

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (H_{\theta}(X^i) - Y^i)^2 ; \text{ Where } m \text{ is the number of training sets}$$

Given below gradient descent algorithm tries to reduce the value for cost function (J(θ)) by starting with an initial guess for θ_j and repeatedly minimizing the cost function.

$$\theta_j := \theta_j + \alpha \frac{\partial}{\partial \theta_j} J(\theta) ; \text{ Where } j = 0, \dots, n, \alpha = \text{learning rate}$$

Andrew [7] further simplifies the partial derivative part of the above gradient descent algorithm and proves a relationship as; $\frac{\partial}{\partial \theta_j} J(\theta) = (H_{\theta}(X^i) - Y^i) X_j^i$ for a single training example. Therefore gradient descent function related to a single training example can be expressed as;

$$\theta_j := \theta_j + \alpha (Y^i - H_{\theta}(X^i)) X_j^i$$

The above expression is named as least square means update rule in machine learning. Andrew (2012) expands the above gradient descent function for a single training data example and develops following algorithm suits to a complete training data set.

Repeat until convergence {

$$\theta_j := \theta_j + \alpha \sum_{i=1}^m (Y^i - H_{\theta}(X^i))^2 X_j^i \quad (\text{for every } j).$$

}

The above algorithm process the entire training data set on every step and results an optimum solution. Therefore it is named as batch gradient descent.

Andrew [7] argues that batch gradient descent always tries to converge in to a global optimum point related with an entire training data set. Therefore globally optimal performance weights can be expected if it is applied to evaluate the ‘contribution of departments to the overall company performance’ or ‘contribution of employee teams to a department’s performance’. Main constraint can be highlighted in adopting the above algorithm it is too slow since it proceeds through entire data set before attaining a value to any θ_j .

Andrew [7] presented stochastic gradient descent algorithm can be considered as an alternate to the batch gradient descent algorithm.

Loop {
For i=1 to m, {

$$\theta_j := \theta_j + \alpha (Y^i - H_{\theta}(X^i)) X_j^i \quad (\text{for every } j).$$

}
}

Advantage of the stochastic algorithm over batch gradient descent is, it updates the parameters with respect to each training example and identifies the gradient of error for each training example. Therefore speed of the algorithm is faster and more priority is given to the final training example processes. It is important because an incentive scheme requires identifying the performance relationship for the final period rather than a global optimum solution.

III. Methodology

In evaluating the performance contribution of employees, it is important to establish a clear link between employees' performance with the overall company performance. Therefore, it is important to identify each department's performance contribution for the company's performance of the same period. Similarly, employee performance contribution for the overall departmental performance should be evaluated. It can be represented as Fig.2.

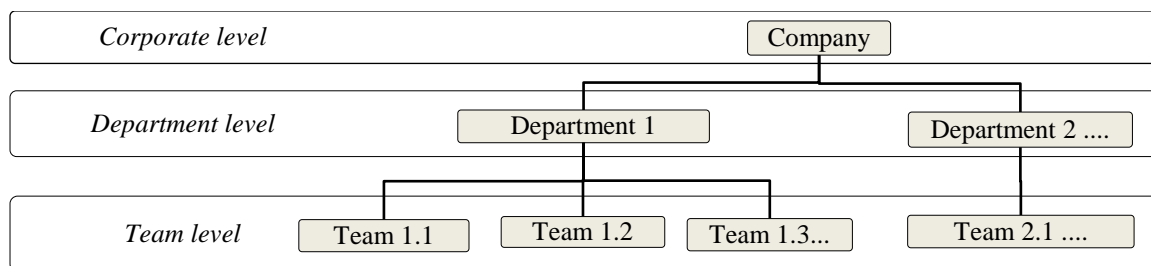


Figure 2. Performance contribution flow through organizational hierarchy

3.1 Evaluation of performance contribution of departments towards company level

Departmental performance contribution towards factory performance for the period can be evaluated as;

$$KPI_F = \theta_1 KPI_{Dept1} + \theta_2 KPI_{Dept2} + \theta_3 KPI_{Dept3} + \dots$$

Where

KPI_F : Weighted values for KPI representing factory

KPI_{Depti} : Weighted values for KPI representing department i, $i=1,2,\dots,n$

Each department's performance contribution for the factory performance can be found using a regression approach. With a stochastic gradient descent algorithm, above relationship can be solved using training examples (historic data of KPIs). With the values for $\theta_1, \dots, \theta_n$ each department's performance contribution can be found as a ratio (e.g.: for department 1 = $\frac{\theta_1}{\sum_{i=1}^n \theta_i}$). More localized optimal solution can be attained under this method which gives accurate results to the current incentive calculation period's performance contribution.

3.2 Evaluation of performance contribution of teams towards departmental level

As an example, assume performance of a single team is measured through two KPIs. Each KPI's current contribution towards the respective department's performance can be evaluated as following equation.

$$KPI_{Dept1} = \theta'_1 KPI_1 + \theta'_2 KPI_2$$

By using weighted KPI values to stochastic gradient descent algorithm, above relationship's θ'_1 and θ'_2 can be solved for the current period. The ratios of $\frac{\theta'_1}{\sum_{i=1}^2 \theta'_i}$ and $\frac{\theta'_2}{\sum_{i=1}^2 \theta'_i}$ represents the selected employee team's (team 1) contribution of two KPIs towards department 1's performance for current period.

3.3 Weighting the KPI values to use in gradient descent

Stochastic gradient descent algorithm has been used in two instances in order to find the weights for performance contributions. Given below are the linear regression solved through the gradient descent.

$$KPI_F = \theta_1 KPI_{Dept1} + \theta_2 KPI_{Dept2} + \theta_3 KPI_{Dept3} + \dots$$

$$KPI_{Dept3} = \theta'_1 KPI_1 + \theta'_2 KPI_2$$

However it is required to weight the KPI training examples before they are used in gradient descent. It is because different values of KPIs may have different scales and it would have a high impact on the θ or θ' values. e.g.: values for KPI_{Dept1} may give in thousands and values for KPI_{Dept2} may give in decimals. Therefore it is important to bring different scales to a single scale before it is being used. As a solution, it is suggested to weigh each KPI value from its average value of historic data.

3.4 Consideration on KPI co-relation to company performance

Some key performance indicators are having positive co-relation with the overall company performance and some may have a negative co-relation. As an example, standard hours achieved can be considered as a positively co-related KPI with the company performance because, greater the standard hour's achievement greater will be the opportunity to attain higher company performance. In contrast, a KPI like defect

rate is having a negative co-relation with overall company performance because, greater the defect rate, lesser will be the opportunity to attain a higher company performance.

Attention has been paid in categorizing the KPIs based on their co-relation because it can affect the performance contribution in two ways. If negatively co-related KPI data have been used in the gradient descent solving, respective gradient values will result minus values which creates problems in proceeding. In the case of incentive scheme, it would result misleading results if a KPI is having a negative co-relation with company performance. Therefore as a solution, it has been suggested to convert the values of negatively co-related KPIs to positively co-related values (1 - negative co-related KPI %) before they are processed to contribution evaluation.

IV. Conclusion

The paper has presented an application of stochastic gradient descent algorithm in evaluating the performance contribution of employees at different levels of organizational hierarchy. Under that, a logical relationship has been established between performance at corporate level, department level and team level. This approach can be used in evaluating the performance of employees and allocating the performance based incentives to them.

References

- [1] Weygandt, J.J., Kimmel, P.D. and Kieso, D. E., *Managerial Accounting: Tools for Business Decision Making*. (NJ: John Wiley & Sons, Inc., 2010) 5th Edition
- [2] Jackson, S., Sawyers, R. and Jenkins, G., *Managerial Accounting: A Focus on Ethical Decision Making*. (OH: South-Western Cengage Learning, 2009), 54.
- [3] Alpaydin, E., *Introduction to Machine Learning* (MA: Massachusetts Institute of Technology, 2004), 1-37.
- [4] Wei, J., Jian-qi, Z. and Xiang, Z., Face recognition method based on support vector machine and particle swarm optimization, *Expert Systems with Applications*, 38(4), 2011, 4390–4393.
- [5] Hua, G. et al., Introduction to the Special Section on Real-world Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33, 2011, 1921-1924.
- [6] Dua, S. and Du, X., *Data Mining and Machine Learning in Cyber security* (NW: Taylor & Francis Group, 2011).
- [7] Andrew Ng, (2012) *Supervised learning*, CS229 Lecture notes Stanford University, California. <http://cs229.stanford.edu/materials.html> (Accessed: 2013)