Benchmarking Organizational Performance Dynamics Using Memory Cells Recurrent Neural Networks

Robert Nsinga^{a*}, Stephen Karungaru^b

^{a, b}Tokushima University
2-1 Minamijosanjima, Tokushima 770-8506, Japan

*email: robert.nsinga@ceable.com

Abstract

Competitiveness in the workforce is being challenged on the basis of hard skills, acquired through lifelong systemic and measured learning, that a computer can be taught to reproduce with unmatched speed and accuracy. Cloud computing, for example, brought with it online accounting, sales, and even contract negotiation, etc.

The primary objective is to train a neural model based on soft skills as a complementary set of attributes that include empathy, trustworthiness extraversion, deference, ethicality, neuroticism, etc., to reclaim the competitive edge of the human workforce. The second objective is to present teams (as opposed to individuals) as a better, more optimized unit of organizational charting.

Keywords: neural memory, analytics, prediction.

1. Introduction

Many experts are already asking whether IoT (Internet of Things) will bring with it the “quantified employee” ⁽¹⁾, and if or when the 4th industrial revolution will make its mark⁽²⁾. Computers claim supremacy in speed and accuracy, but judgement and reflex remain a computing complex problem^(1,2). This study attempts to design intuitive memory from a set of vectors to consider in predictive models.

Recent advances in cognitive science argue about the different states of change that the human brain undergo. Some of these changes are shown to affect, in good or bad ways, the operational mechanisms with regards to how humans behave, work, or relate with one another.

This and many other researches have led to what is known as brain training ⁽³⁾. A 2015 census has demonstrated that 50% more engagement has occurred amongst employees in the last 20 years. It claims 75% of the time spent at work is in communicating, exchanging views and meetings ⁽⁴⁾.

With the advent of the internet, growing volumes in data, and powerful computing commodity, tough economic conditions have put pressure on competitive differentiation. Predictive analytics are part of a growing toolset used to optimize and automate decisions in order to stay competitive. Predictive models use known results to predict values for new or different data. This probabilistic exercise is based on estimated significance from a set of input variables.

Recent talent acquisition trends favor teams over individuals. Yearly appraisals, for instance, have lost value due to digital workspaces, or co-offices. This new type of workspace abstracts conventional biases of diversity and inclusion by focusing on accountability, data, and transparency. Once the sole responsibility of HR, hiring now runs across multiple teams, hiring managers inside an organizational structure and outside via consulting professionals, crowd-sourcing platforms, or bots ^(1,4).

2. Literature Review

Kozlowski et al., 2003 contributed an interesting differentiation in the field of industrial and organizational psychology that guides this paper's definition of a team ⁽⁵⁾. They concluded that organizational charts do not clearly support the quantization of performance, processes, and effectiveness. Primarily because these aspects are of a qualitative nature, and may be difficult to contextualize. They were indirectly referring to soft skills.

A combination of both qualitative and quantitative assessments was later conducted privately by Google with a mix of high- and low-performing teams. The results confirmed a similar discovery made by Dr. Edmondson, who is credited for coining the term “psychological safety”⁽⁶⁾. In her study she attributed overall lack of performance inside an organization to a behavioral trait acquired through experience (childhood, education, etc.) in which individuals, regardless of their responsibilities, carry a certain stigma. Psychologists refer to this trait as “impression management”, which studies show a large majority of today's working class unconsciously share.

This stigma is characterized by the absence of a rationale to contribute; which impacts the climate of openness in an organization and therefore its performance. Climate in this context may refer to organizational culture. Dr. Edmondson offered solutions to treat the stigma by prescribing clear guidelines. Google used insights from her research and applied their own internal assessments to derive similar implementations of what it means to be an effective team. Using standard, proprietary methods like the Berkeley Big Five Inventory (BFI), or the Toronto Empathy Questionnaire⁽⁷⁾ with other assessments including demographic variables like tenure, seniority, and location⁽⁸⁾, Google inferred statistical models on years of collected data with a wide array of variables⁽⁹⁾.

Their methods are proprietary, however this paper offers a proof of concept using synthesized data to train a deeply layered model for prediction with the assumption that there is uncertainty in the problem being solved by the team, and there is a need for interdependence of the members. Studies have shown that committing to a goal can help improve performance, either individual or group-wide⁽¹⁰⁾, which this paper also assumes a priori for simplicity.

Deeply layered machine learning models apply a method referred to as feature extraction by preserving the spatial relationship between inputs. By systematically convolving through nodes, the saliency of a feature is determined as it is fed to other nodes deeper in the network and via backpropagation to adjust the biases⁽¹¹⁾ which in turn supports the prediction accuracy of a new or different input. Assuming a relationship through correlation or causation between inputs and outputs is a heuristic approach to finding the function

$$f(x) = y \quad (1)$$

called *universal approximator* (the algorithm), such that the mapping of x to y , regardless of the activation function applied, but considering the multilayer feedforward architecture, approximates a known source of truth⁽¹²⁾.

Using the notions of psychological safety and collective intelligence⁽¹³⁾ to approximate empathy and openness as a function is this paper's main contribution. Dr. Edmondson suggested psychological safety as another dimension at par with other dimensions like accountability and transparency. A deep layered, multi-input and multi-output model is designed as a representation of the function used to test the hypothesis.

3. Proposed Methods

3.1 Dataset Composition & Structure

Combining sociologist surveys methods with machine learning, this paper uses multiple sources of research to synthesize the data. Focus is made on (a) an individual's features, (b) the features they exert in a team, (c) the workforce collective features, and (d) the problem features, quantified for their “solvability” boundary and added once the model has built a memory of the available workforce. Results are highlighted and compared using two different optimizations scenarios:

(a) Individual features

Based on factor analysis of various cognitive abilities and professional achievements. Actual data can be collected from monitored activity or via mobile sensors^(4, 5, 8, 10).

(b) Individual behavior in a team

Abilities exerted only in a team environment^(6, 7, 9, 10). They may be influenced by group norms.

(c) Collective features

Modeling metrics collectively shared as a sum of the collective intelligence of individuals^(6, 7, 9, 10, 11).

(d) Problem domain

A range of problem definitions are synthesized to represent the unpredictability of a solution^(6,9,12). Classified as either of two possible states: easy or hard, and two outcomes: solved or not.

Table 1. Problem Domain.

Feature	Schema	Range	Average
experience	int.	1~12	2
cert_density	float32	.1~1.9	0.9
education	int.	1~4	2
skillset	int.	1~10	4
func_scope	int.	1~5	3
sen_relattiv	int.	1~4	3

3.2 Training the model

Let $\varphi(\cdot)$ be a non-constant, bounded and increasing function, there exists an integer N , real constants $v_i, b_i \in \mathbb{R}$ and real vectors $w_i \in \mathbb{R}^m$, where $i = 1, \dots, N$ such that

$$F(x) = \sum_{i=1}^N v_i \varphi(w_i^T x + b_i) \quad (2)$$

is an approximate realization of equation 1⁽¹⁵⁾. Two optimization techniques are used as “ $v_i \varphi(\cdot)$ ”, namely the stochastic approximation of minima and maxima with SGD (Stochastic Gradient Descent) and its variant RMSProp (Root Mean Square Propagation, Tieleman et al., 2012) on identical learning rates ^(15,16).

With a sample size of 1,000 the training is performed in batches of 10 in 50 epochs. The result is evaluated on a subset of 100 samples. Running two separate supervised learnings, one with SGD, and another with RMSProp, the kernels, or “ w_i^T ” in equation 2, also called weights, are optimized using backward propagation.

3.3 Network Graph

To retain knowledge of the available workforce, this paper proposes expanding the main input dimensions using an embedding layer, then feeding the results into an LSTM algorithm ⁽¹⁶⁾ to build a “gut feeling”.

Auxiliary input from Table 1 is then concatenated with the LSTM output from Table 2, and the mashup is fed into a 3-layered ReLu network of 64 nodes each.

The final sigmoid function “ $w_i^T x + b_i$ ” layer outputs the prediction: given a problem X_2 , individual X_1 displays the optimal set of behaviors $Y_{1,2}$ in his current or assigned group/team configuration. Refer to figure 1.

Loss and accuracy are monitored using binary cross entropy and root mean squared computations. This model

Table 2. Combined individual/group features.

Feature	Schema	Range	Average
feats/tenure	int	1~12	2
certifications	int	1~5	3
skills overlap	int	1~10	3
recommendations	int	1~10	3
neuroticism	float32	.3~.8	0.4
familiarity	float32	[.0~1.9]	0.9
deference	int	1~10	6
ett_ratio	int	.1~1.9	1.0
openness	int	20~85	60

uses ReLu ⁽¹⁶⁾ to enforce feature detection and classification.

3.4 Error and Loss

As the model is being supervised, the proposed method iteratively reviews the error rates and overall training loss using mean squared error (MSE) and binary cross entropy (BCE) respectively ⁽¹⁶⁻¹⁹⁾. The former squares the difference between the estimator (see equation 2) and the evaluation data, or hypothesis, as per the equation

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (3)$$

for any vector $\hat{Y}$ of n predictions, and the vector of synthetic values Y , as proposed in this paper.

Binary cross entropy is computed as per the formula

$$H_{y'}(y) := - \sum_i y'_i \log(y_i) \quad (4)$$

Fig. 1. Network Diagram.

Fig. 2. SGD Output

Fig. 3. RMSProp Output

for training set y accuracy is tested against evaluation data y' to provide a measure of “encoding length”, which is the number of bits expected based on the hypothesis (evaluation data). Distribution is hereby hypothesized as well.

4. Experiments

The model was designed in Python, version 3.6.3 with Keras version 2.0.9 using the TensorFlow library (version 1.4.0), running on a Windows 10 Enterprise, version 1703 with 4GB of memory, operating on x64-based processor (Intel® Core™ i7-2700K at 3.80GHz).

Training samples and validation samples were identical, totaling 1,000 and 100 respectively. Input was fed in batches of 10 samples iterated over 50 epochs. The network graph is also identical in both experiments. Figures 2 and 3 show the readings of the results, which took respectively 108.211 and 116.913 seconds to complete.

4.1 SGD Optimized Learning

The input is configured at 12 vectors, or features, generated from synthetic data, given certain heuristics (psychological, sociological, human capital statistics, etc.). The sample contains 1,000 subjects for training. For optimal results, the experiment considers a 10/10 ratio. The samples

are fed into the network graph in figure 1 using the Stochastic Gradient Descent (SGD) optimizer, using a technique that applies a parameter update for each training sample at high frequency rates.

4.2 RMSProp Optimized Learning

A similar design is repeated using an optimization technique that compensates the average gradient by dynamically increasing the coefficient used in SGD. It is similar in design to the Adagrad optimizer (Zeiler, 2012) but resolves the latter’s learning rates limitations. RMSProp stands for Root Mean Squared Propagation.

5. Results

The SGD optimized learning reached 0.754 accuracy by the epoch 10 out of 50 and remained constant until the end. Loss smoothly decreased from a little above 1.28 at epoch 1 to 1.17899 by epoch 50. The t-SNE visualization algorithm⁽¹⁸⁾ does not suggest a noticeable separation in feature distribution across the dataset.

On the second experiment using RMSProp optimized learning, accuracy started a smooth hiking from 0.750 by epoch 5 until it reached well over 0.97 by the end of learning (epoch 50) against a loss below 0.2. t-SNE visualization shows 2 distinct clusters suggesting a successful learning.

Fig. 4. t-SNE 3D graph of RMSProp Optimization

6. Conclusions

By suggesting more emphasis on collective contributions and deference to group norms, and less on hierarchy or organizational charting, this paper highlights the need to frame each problem solving endeavor as a learning opportunity. The usefulness and contribution of imperfection is encouraged as a necessary step in problem solving, without sacrificing accountability. The proposed methods can be used to support talent acquisition in sourcing good team members by focusing on soft skills.

In future revisions live data will be sourced from a mix of automated surveys and API fetches on social media networks like Reddit, Instagram, etc. which store valid time-series data about familiarity (follows, followers, friend requests) and common interests (shared likes), cited here as non-exhaustive examples.

References

- (1) Bersin et al., Deloitte Global Human Capital Trends, <http://dupress.deloitte.com>, 2017, retrieved Nov. 2017.
- (2) Klaus Schwab, The Fourth Industrial Revolution, WEF, 2016.
- (3) Bamidis PD, Fissler P, Papageorgiou SG, et al., Gains in cognition through combined cognitive and physical training: the role of training dosage and severity of neurocognitive disorder, *Front Aging Neurosci.* 2015; 7:152.
- (4) Harvard Business Review, Jan-Feb 2016 (pp. 74-79)
- (5) Kozlowski, S.W.J. & Bell, B.S., Work Groups and Teams in Organizations, *Industrial and Organizational Psychology* (vol. 12), pp. 333-375, 2003
- (6) Edmondson, Amy, Psychological safety and learning behavior in work teams, *Administrative science quarterly* vol. 44/2, pp. 350-383, 1999.
- (7) Spreng R. N., McKinnon M. C., Mar R. A., The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures, Levine B. *Journal of personality assessment.* 2009; 91(1): 62-71. doi:10.1080/00223890802484381.
- (8) Baek, P. and Reidy, S., Is there a correlation between diversity in the workforce and diversity in the leadership and business results? Reference website: <http://digitalcommons.ilr.cornell.edu/student/119>, (2016). retrieved Nov 2017.
- (9) Guide: Understand team effectiveness, website: <https://rework.withgoogle.com/print/guides/5721312655835136/> (Nov 2016), retrieved Oct 2017.
- (10) Locke, Edwin & Latham, Gary, Building a practically useful theory of goal setting and task motivation - A 35-year odyssey, *The American psychologist.* 57. 705-17 2002.
- (11) Gal, Yarin, and Zoubin Ghahramani, Dropout as a Bayesian approximation: Insights and applications, *Deep Learning Workshop, ICML*, 2015.
- (12) Cybenko, G., Approximations by superpositions of sigmoidal functions, *Mathematics of Control, Signals, and Systems*, 2(4), 303-314, 1989.
- (13) A. W. Wooley et al, Evidence for a Collective Intelligence Factor in the Performance of Human Groups, *Science* 2010 (pp. 686-688)
- (14) C. Spearman, "General Intelligence", objectively determined and measured. *Am. J. Psych.* 15, 201 (1904)
- (15) Abadi et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, tensorflow.org, 2015.
- (16) Hochreiter et al., Long short-term memory, *Neural Computation* 9.8, pp. 1735-1780, 1997.
- (17) F. Chollet, Keras: Deep Learning for Python, <http://github.com/fchollet/keras>, 1f0bd8c, 2015.
- (18) Hinton et van der Maaten, Visualizing High-Dimensional Data Using t-SNE, *JMLR* 9, pp. 2579-2605, 2008.
- (19) Glorot et al., Deep Sparse Rectifier Neural Networks, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 315-323, 2011.