Development and Evaluation of An Employee Performance Appraisal Insight Report Generator

Mary Jane Magno-Tan[1], Claudine Almonte [2], Julie Flor Angad [3], Patricia Matignas [4], Ana Rose Trajano[5]

[1] Faculty, College of Computer and Information Sciences, Polytechnic University of the Philippines
[2345] College of Computer and Information Sciences, Polytechnic University of the Philippines

---***---

Abstract - *Performance appraisal is based on qualitative parameters that can be observed and measured by approximations, which are imprecise data. These include factors such as technical, quality, interpersonal, communication, approach to work, quantity, leadership and managerial skills specific for an organization. Hence, performance appraisal report is best presented through human language or texts instead of quantitative or tabular data that do not provide personal insights of the evaluator. This research focuses on development of a software using fuzzy inference technique, that allows employee performance evaluators to rate each employee according to some set criteria. Using Natural Language Generation (NLG), the results of evaluation that represent the conclusions of the evaluators are presented in a form of insight report which narratively states the performance of each employee in terms of their specific strengths, weaknesses and evaluator's recommendation; providing a basis for identifying and correcting disparities in employee's performance and may also provide the basis for other personal actions like salary appraisal, training and career development, promotion or termination. The degree of correctness of the insight reports generated by the software was measured using through experimentation, yielding high result. The acceptability of the insight report generator among human resource experts was determined using a survey form; respondents were agreeable in the acceptability of the software in terms of usefulness, functionality and user-friendliness. It is recommended that the software be enhanced by integrating Natural Language Processing (NLP) and sentiment analysis that allows evaluators to answer open-ended questions.*

Key Words: performance appraisal, fuzzy inference technique, natural language generation, insight reporting

1. INTRODUCTION

In communicating complex results, quantitative information or tabular data alone do not provide personal insights; they require interpretation expertise to understand. Users struggle to find out what those numbers mean and what they should do about it. [1]. Hence, there is a need for reports written in paragraph form or what is called insight report.

With the use of Natural Language Generation (NLG), these quantitative information can be improved. There have been several successful applications of NLG technology such as generating weather forecast from numerical weather simulation data [2], providing information about restaurants in a given city, producing report describing the simulation options that an engineer has explored, and summarizing pollutant information for environmental officials [3].

In the same way, performance appraisal systems need a report generator that can save time to communicate complex results and information by putting reports into easy-to-understand everyday language [4].

Performance appraisal and reporting are significant for an organization to monitor each employee for the enhancement of future performance, to project easily the good and bad performers within the organization, to effectively measure and evaluate the employee performance, and to identify the shortfall when performance does not meet the standards set by the organization.

Most organizations use numerical values or linguistic labels like good, very good, outstanding, etc. in their performance appraisal system. However, these scores are merely imprecise approximations as they are based on judgment making ability of the reviewer. The use of fuzzy logic allows reviewers to express themselves linguistically and to make assessments that are subjective in nature. Employee evaluation is based on many parameters like technical, communication skills, leadership skills, etc. These parameters are very fuzzy and not just black and

white. It employs spectrum of colors, accepting that things can be partly true and partly false at the same time. Such human like approach is well implemented using fuzzy logic, which models human like decision making and common sense [5].

The researchers used Fuzzy Inference Technique (FIT) for performance appraisal and Natural Language Generation (NLG) in producing insight report on each employee; providing information on the specific areas of weakness and strength, and recommended areas for improvement.

The research seeks to answer how accurate is fuzzy inference technique and natural language generation in producing narrative report on employee performance appraisal. Will the system be accepted by human resource managers in terms of user-friendliness, functionality, and usefulness?

2. METHODOLOGY

2.1 Software Development

The researchers used Python programming language to develop the insight report generator named *E-Perform*, following an iterative development process. Below is the software detailed architecture.
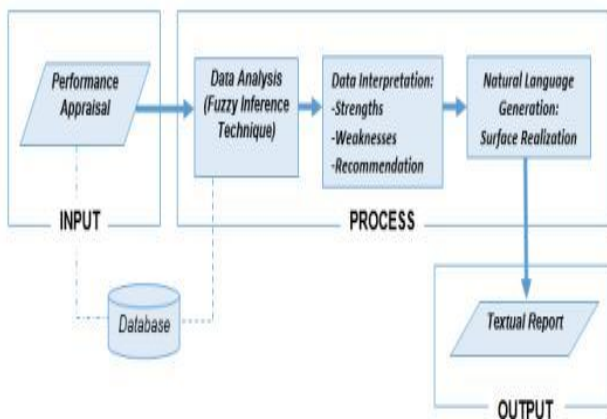

Figure 1. E-Perform Software General System Architectures

Figure 1 shows the general system architecture of E-Perform. It shows the flow of the process in generating the output. The input *Performance Appraisal* is the evaluation of the Appraiser to an employee in the form of rating (between 1-5). The user will enter the employee name, position and rating. Fuzzy inference technique is used in *Data Analysis* to analyze raw data which will be used in *Data Interpretation*, where keywords are used in *Natural Language Generation* using surface realization technique in generating textual reports containing strengths, weaknesses and recommendation.
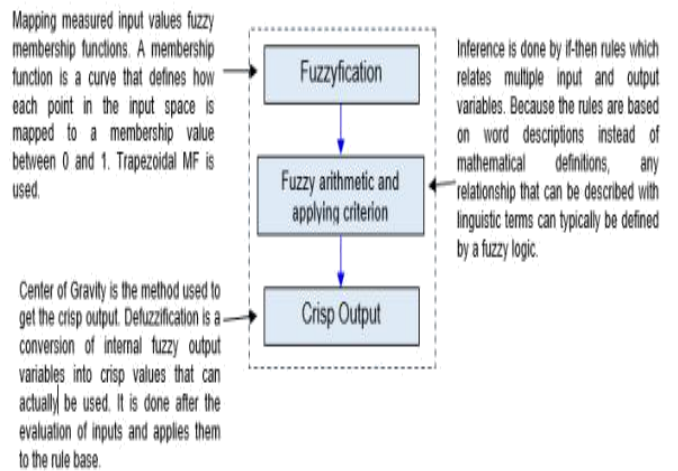

Figure 2. Fuzzy Architecture

Figure 2 shows *Fuzzy Architecture* of *Data Analysis* where the input (performance rating) will undergo fuzzyfication, to produce the crisp output which will serve as the input for *Data Interpretation.* The last process which is the *Natural Language Generation* has an architecture given below.
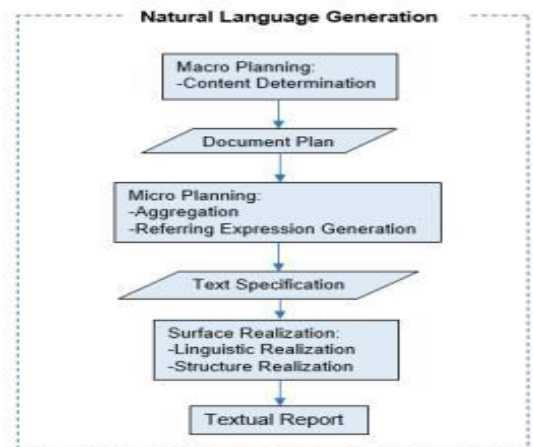

Figure 3. NLG Architecture

Figure 3 shows how a textual report is generated through natural language generation. The *Content Determination* is based from the *Data Interpretation* (strengths, weaknesses, recommendation), producing the document plan; it is followed by the *Aggregation and Referring Expression Generation* for text specification, and the *Surface Realization* (linguistic and structure) where words will form sentences conveying a thought.

Performance evaluation criteria or parameters considered in the E-Perform software are shown in Table 1.

Table 1. *Qualitative Performance Criteria/Parameters*

Performance Criteria	
Technical Skills	• Technical Proficiency • Analyzes Problems, • Employs tool of the job competently
Quality of Work	• Accuracy or Precision • Thoroughness or Neatness • Reliability or Dependability
Interpersonal Skills	• With Co-Workers • With Supervisors and Managers • Team participation
Communication Skills	• Written expression • Oral expression • Tact and diplomacy
Approach to Work	• Actively seeks ways to streamline • Flexible/Adaptable • Initiative and Motivation
Quantity of Work	• Priority Setting • Amount of work • Wok completed on schedule
Leadership Skills	• Support to company's goals and objectives • Supports responsible risk taking • Instills pride in performance, service, innovation and quality
Managerial Skills	• Trains and develop staff • Properly aligns responsibility, accountability and authority • Evaluates staff regularly

2.2 Software Evaluation

A survey questionnaire was used to assess the system's acceptability in terms of user-friendliness, usefulness and functionality. Purposive sampling was used to select ten (10) managers as respondents based on their availability, knowledge, and capacity.

Using the weighted mean as formula, final scores are calculated. Likert Scale was used to interpret the computed scores in the survey questionnaires intended for the managers. In the survey instrument, four (4) choices were provided for every question or statement. The choices represent the degree of opinion each respondent has on the given question. The range and interpretation of the four-point scale are shown in Table 2 below.

Table 2. *Likert Scale*

Scale	Range	Interpretation
4	3.26 – 4	Strongly Agree
3	2.51 – 3.25	Agree
2	1.76 – 2.50	Disagree
1	1 – 1.75	Strongly disagree

To measure the correctness of the insight reports generated by the software, ten (10) human resource experts evaluated the reports generated by the software to determine whether the outputs are correct or not. Each expert appraised ten (10) employees and evaluated the output of ten (10) insight reports generated. These experts were composed of managers, department heads and supervisors which are available, willing and has the knowledge and capacity to evaluate the system.

The tabular values derived from the evaluation were used to compute for the total sum of squares(SST) , treatment sum of squares(SSTR), and error sum of squares (SSE) which were then used to compute for the total mean square(MST), mean treatment of square(MSTR), mean square error(MSE). Using these obtained values, the observed value was computed.

Before getting the level of correctness, the critical value for the decision was derived and compared to the observed value to determine if there is no significant difference among the evaluations of the experts with regards to the correctness of reports generated by the software.

The observed value was then multiplied to 100 to get the level of correctness of E-Perform software in generating insight report on each employee performance.

3. RESULTS AND DISCUSSION

Table 3 shows that the three evaluation parameters of the acceptability of the system all got an agreeable verbal interpretation.

Table 3. *Evaluation on Acceptability of E-Perform in Terms of Usefulness, Functionality, User-Friendliness*

	$\bar{x}$	Verbal Interpretation
Usefulness	3.16	Agree
Functionality	3.16	Agree
User-friendliness	3.24	Agree

As presented, the software has an average of 3.16 in terms of usefulness, 3.16 in terms of functionality, and 3.24 in terms of user-friendliness. The results indicate that human

resource managers have high acceptance of the insight report generator; thus, natural language generation is highly accepted in communicating results of employee performance appraisal.

To determine the correctness rate of the software in generating the insight report on employee performance, 10 experts appraised 10 employees and evaluated the output of 10 insight reports generated by the software. Table 4 shows whether each expert agree on the software's generated report on each employee performance. The ten (10) columns represent the ten (10) human resource experts while the ten (10) rows represent the ten (10) insight reports generated for each employee.

Table 4. *Evaluation of Human Resource Experts on Insight Report Generated for Each Employee*

	Eval 1	Eval 2	Eval 3	Eval 4	Eval 5	Eval 6	Eval 7	Eval 8	Eval 9	Eval 10
Emp 1	✓	✓	✓	✓	X	✓	✓	✓	✓	X
Emp 2	✓	✓	✓	✓	X	X	✓	X	✓	✓
Emp 3	✓	✓	X	✓	✓	✓	X	✓	X	✓
Emp 4	X	✓	✓	✓	✓	✓	✓	✓	✓	✓
Emp 5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Emp 6	X	✓	✓	X	✓	✓	✓	✓	✓	✓
Emp 7	✓	✓	✓	✓	✓	X	✓	X	✓	✓
Emp 8	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Emp 9	✓	✓	✓	✓	✓	✓	✓	X	X	✓
Emp 10	X	✓	✓	X	X	✓	✓	✓	✓	X
Total	7	10	9	8	7	8	9	7	8	8
X	0.7	1	0.9	0.8	0.7	0.8	0.9	0.7	0.8	0.8

Legend: ✓ = Agree   X = Disagree

The observed value F computed was 0.609.

$$F = \frac{MSTR}{MSE} = \frac{0.098}{0.161} = 0.609$$

The critical value for the decision computed
df1 = c – 1 = 10 – 1 = 9
df2 = N – c = 100 – 10 = 90
The critical value based on f table α = 5% is $F_{9,90}$ = 1.986. Since the observed value 0.609 is lesser than the critical value 1.986, there is no significant difference among the evaluations of the experts with regards to the correctness of insight reports generated by the software.

Multiplying the observed value 0.609 to 100 to get the level of correctness, the result is 60.9% and therefore, the level of correctness is high.

Table 5 shows the interpretation of the level of correctness of the report generated by the software.

Table 5. *Result of Experts Evaluation on E-Perform Level of Correctness in Generating Insight Report*

Percentage	Level of Correctness
76-100	Very High
51-75	**High**
26-50	Low
0-25	Very Low

Using the formula for Mean Treatment of Square (MSTR) and Mean Square Error (MSE), the observed value 60.9% has a descriptive equivalent to *high* rating for the correctness rate of the software. Since the observed value 0.609 is lesser than the critical value 1.986, there is no significant difference among the evaluations of the experts with regards to the correctness of insight reports generated by the software.

Since the correctness rate and acceptability of the software is high, it can be safely concluded that the natural language generation techniques and the fuzzy logic algorithm used by the researchers to develop an employee performance appraisal insight report generator can properly and correctly generate narrative reports.

3. CONCLUSIONS

It is recommended that the software provide open-ended questions (in English language) that employee evaluators would answer with natural language (English) instead of fixed criteria set by the software for performance appraisal. This requires integration of Natural Language Processing (NLP) into the software, where the answers would be analyzed using sentiment analysis.

REFERENCES

[1]   Allen, Robbie (2014). Inroducing Wordsmith: Using Data to Reinvent How We Write. Retrieved September 30, 2014, from https://www.glassdoor.com/Overview/Working-at-Automated-Insights-EI IE891995.11,29.htm

[2]   Moore, J. (2013). *Natural Language Generation: An Introduction.* Retrieved September 13, 2014, from http://www.inf.ed.ac.uk/teaching/courses/nlg/lectures/2012/NLG2012Lect1.pdf

[3]   *Data-to-text Natural Language Generation and Evaluation.* (2013). Retrieved September 12, 2014, from http://staff.um.edu.mt/albert.gatt/hit-msra2012

[4]   Gardent, C. (2013). *Natural Language Generation.* Retrieved September 12, 2014, from http://www.loria.fr/~gardent/teaching/nlg-bkk06.pdf

[5]   NishaMacwan, Dr.PritiSrinivas Sajja(2013), Performance Appraisal using Fuzzy Evaluation Methodology. http/www.ijiet.com/Vol23/Issue23/IJIET141220 1309 50.pdf