Polibits
ISSN: 1870-9044
polibits@nlp.cic.ipn.mx
Instituto Politécnico Nacional
México

Castro-Hernández, Alberto; Swigger, Kathleen; Cemile Serçe, Fatma; Lopez, Victor
Classification of Group Potency Levels of Software Development Student Teams
Polibits, vol. 51, enero-junio, 2015, pp. 55-62
Instituto Politécnico Nacional
Distrito Federal, México

ISSN 2395-8618

Classification of Group Potency Levels of Software Development Student Teams

Alberto Castro-Hernández, Kathleen Swigger, Fatma Cemile Serçe, and Victor Lopez

Abstract—**This paper describes the use of an automatic classifier to model group potency levels within software development projects. A set of machine learning experiments that looked at different group characteristics and various collaboration measures extracted from a team's communication activities were used to predict overall group potency levels. These textual communication exchanges were collected from three software development projects involving students living in the US, Turkey and Panama. Based on the group potency literature, group-level measures such as skill diversity, cohesion, and collaboration were developed and then collected for each team. A regression analysis was originally performed on the continuous group potency values to test the relationships between the group-level measures and group potency levels. This method, however, proved to be ineffective. As a result, the group potency values were converted into binary labels and the relationships between the group-level measures and group potency were re-analyzed using machine learning classifiers. Results of this new analysis indicated an improvement in the accuracy of the model. Thus, we were able to successfully characterize teams as having either low or high potency levels. Such information can prove useful to both managers and leaders of teams in any setting.**

Index Terms—**Software development, group potency, machine learning.**

I. Introduction

BECAUSE of the rapid rise of globalization within industry, the use of virtual teams has dramatically increased in recent years [1]. Despite its known advantages, such as reducing costs and obtaining access to workers with different skills [2], managing remote teams remains challenging, largely because of the difficulty in using electronic media to communicate with members located at remote sites [3]. Nevertheless, this same electronic media now allows managers to keep track of the actions and interactions that occur within a work team. Moreover, the data obtained from these electronic media can be converted into useful information for not only the managers of global teams [4], [5] but also researchers who are looking at various elements of group dynamics. However, determining which information is most useful and how it can be used to predict different team characteristics remains a serious challenge for group-focused researchers.

The search for various team measures, as well as the techniques for modeling the interactions of these measures, has been met with various degrees of success over the past few years. For example, [6] characterized learning groups as graphs, with vertices representing students and edges representing the number of messages interchanged bidirectionally. Using this model, [6] identified *Milson's communication patterns* and count specific *graph theory elements*. The authors then used this new dataset to create decision trees that predicted five levels of performance. The authors' model was able to predict performance with 78.9% accuracy.

The work by [7] is another example of group-related modeling study. The authors of this research developed a tool, called TeCFlow, to analyze the interaction among employees within a company. Interaction rates were calculated by counting the number of messages exchanged between pairs of workers. This information was then displayed in a graphical format. The software was also able to detect collaboration among subgroups by looking at *communication density*. Once a subgroup was detected, the *Group betweenness centrality* measure allowed the user to find interesting events that might have occurred during a specific period. Based on an analysis of data from email exchanges that occurred within a company, the authors argue that they were able to predict groups' productivity as well as suggest ways to improve a group's performance.

A study by [8] proposed a very different type of model that was based on the premise that groups perform better if they use similar words. The authors of this study tried to predict a group cohesion measure (obtained from an Interaction Rating Questionnaire) by calculating percentages for the number of times a team used nine function words (i.e., auxiliary verbs, articles, common adverbs, personal pronouns, indefinite pronouns, prepositions, negations, conjunctions and quantifiers) in their communications. These percentages were then averaged and labeled as the group's Linguistic Style Matching (LSM) index. The authors also calculated percentages for the number of times "We," Future-oriented, and Achievement-oriented words were used by each team. The authors used these four variables to construct regression models to predict cohesion and group performance. Using

Manuscript received on December 29, 2014, accepted for publication on April 27, 2015, published on June 15, 2015.

Alberto Castro-Hernández and Kathleen Swigger are with Computer Science Department, University of North Texas, USA (e-mail: albertocastro-hernandez@my.unt.edu, kathy@cse.unt.edu).

Fatma Cemile Serçe is with Department of Information Systems Engineering, Atilim University, Turkey (e-mail: cemileserce@gmail.com).

Victor Lopez is with Facultad de Ingeniería de Sistemas Computacionales, Universidad Tecnológica de Panamá, Panama (e-mail: victor.lopez@utp.ac.pa).

ISSN 2395-8618

data collected from group chat communications on a small, collaborative task, the authors found that LSM was able to predict cohesiveness and, to some degree, performance.

Despite the intense research activity in group-oriented research, there remain many questions about which measures to model and which modeling techniques are most accurate. One purpose of this paper is to determine the accuracy of our proposed variables in predicting group potency values. A second purpose of the paper is to compare the prediction accuracy of several commonly used modeling techniques (i.e., regression and machine classifiers). A third purpose of the paper is to identify specific feature values that can promote high levels of the group cohesion construct within a distributed software teams.

II. Group potency

The estimation of group potency within teams has been the focus of much research over the past several years. The group potency construct is usually defined as "a collective belief regarding the team's ability to be successful" (as cited by [9]). The importance of the construct was, and continues to be, its strong relationship to group performance [10], [11], [12]. This strong relationship between group potency and performance has been found at both the individual and group levels, although the strength between the two variables has been shown to be higher at the group level [13].

Group potency levels for teams are generally obtained by aggregating individual scores gathered through questionnaires, or by asking the team, as a whole, to agree upon a single score. The latter procedure usually produces higher group potency scores but lower correlations with group performance, because other team members often persuade team members to inflate their group potency scores. Aggregated questionnaires also seem to be the preferred method for obtaining group potency levels for virtual teams [11].

Not surprisingly, a number of theoretical models have been proposed to predict or explain group potency. For example, [14] proposed a model that used the variables of group composition, charismatic leadership, and group size to cause group potency. In a similar study, charismatic leadership was found to be related to group potency [15], [16], because, as the author explained, the presence of a leader helps guide other members toward the team's goals. Other researchers have found that a team's skill level, knowledge, and performance can have a positive effect on group potency [12], [10], largely because such factors tend to increase a group's confidence levels and, thus, affect the member's perception of the group's abilities. In a similar manner, researchers have determined that there is a positive relationship between group potency and communication and cooperation [10], [16]. Both of these factors allow team members to learn about each other's skills and capabilities, hence increasing the group's overall collective confidence. Finally, group size was found to be related to group potency [12], because, as [14] argues, groups with insufficient (or too many) members feel less confident about the team's ability to complete the task.

The group potency construct has also been compared to another construct called group efficacy [17]. Group potency focuses on the team's ability to perform generally, while group efficacy focuses more on the team's ability to complete a specific task [11]. Because research results indicate a strong relationship between group efficacy and team cohesiveness [18], it seens plausible to believe that group potency may also have a positive relation with team cohesion.

What is apparent from this cursory review of the research is that there are many variables that seem related to group potency. What is not so obvious is knowing which variables can be used to model group potency, and which modeling technique performs best with a particular dataset. These two issues were investigated using a database consisting of electronic communications from three global software development projects. The goal of this research was to find an effective model that can successfully predict group potency levels in virtual software development teams.

III. Research methodology

A. Teams

This research involved three sets of teams that participated in three different virtual collaborative projects during 2012-2013. The first set of teams consisted of students from the University of North Texas (UNT in the US) and students from the Atilim University (AU in Turkey). Participants from the US were enrolled in a Human-Computer Interface course, while participants from Turkey were enrolled in a Software Development course. A total of 53 students participated in this collaborative project; 23 US students and 30 Turkish students. Ten teams were created; each team consisting of 4-6 students, with members from both universities.

The second set of teams was made up of students from different courses within UNT. About half of the participants were enrolled in a Human-Computer Interfaces (HCI) course, and the other half were enrolled in an Artificial Intelligence (AI) course. A total of 50 students participated in this project; 28 students from the HCI course and 22 from the AI course. Ten teams were created for this project; each team consisting of between 4-6 members from both courses.

The third set of teams were formed from students enrolled at UNT and the Universidad Tecnológica de Panamá (UTA at Panama). Participants from the US were enrolled in a Human-Computer Interfaces course, and participants from Panama were enrolled in two different database courses. A total of 64 students participated in the third project; 28 students from the US and 36 from Panama. Thirteen teams were formed for this project, each team containing between 3-5 members from both universities.

In summary, the characteristics and behaviors of 33 teams were analyzed in an effort to predict group potency levels.

ISSN 2395-8618

B. Software development projects

Separate, but similar, projects were created for each of the three sets of teams that participated in this study. The first project, involving US and Turkish students, occurred in November - December 2012 and lasted for 37 days. Each team was asked to complete a mobile application that could run on an Android phone. Sub teams in the US were responsible for developing the interface, while the Turkish teams implemented the mobile application.

The second collaborative project occurred in April–May 2013 and extended over a 35 day period. Student teams in the US were asked to develop an application that would use a reinforcement learning algorithm to decide where cars should park. The application was also suppose to include a display that would allow users to change the parameters to the algorithm. Sub teams in each course (AI and HCI) were asked to develop and test the application.

The third collaborative task occurred in November–December 2013 and lasted 37 days. Each US-Panama team was asked to re-design an existing website (i.e., the home page, the events page, and the contribution page) and implement a database that could support the various operations that were needed to maintain the pages. US sub team was in charge of developing the website, whereas Panamanian teams were responsible for designing and populating the database for the site.

C. Communication tools

A project-management web application based on the Redmine platform was used to collect the communication activities for each team. This application supports several collaborative tools including chat, forums, wikis, document sharing, etc. Additional programs were added to the Redmine tool that enabled the software to record and timestamp all interaction among team members and store them in a centralized database.

Students who participated in each project were asked to communicate with one another using only the Redmine project management tools. In addition, subjects were asked to use English to communicate with one another. Thus, both the Turkish and Panamanian participants were obviously using a second language to collaborate with the US students.

D. Measures

In order to determine which variables predict group potency levels for virtual teams, we developed three different predictor measures: *Team characteristics*, *Collaboration features*, and *Linguistic features*. Team characteristics represent group variables which are defined before the project's start, and they cannot be changed. Collaboration features describe variables which depend on team members' behavior during the project. Linguistic features are a detailed look into the messages' content exchanged.

The criterion variable of *Group potency level* was obtained by averaging a participant's responses to a group potency survey that was completed at the beginning of each project. This particular survey was developed by [14] and consists of eight questions in a five-point Likert scale that are designed to measure a subject's perceptions of their group's capabilities. The individual scores were then combined into a single *Group potency* score for each group.

1) Team characteristics: The *Team characteristics* variable was defined as *Team size*, *GPA average* (average of individual Grade Point Average) and *Team diversity*. The *Team size* measure was obtained by simply counting the number of members in a team. Both a team's *GPA average* and *Team diversity* scores were obtained by examining surveys completed by all subjects at the beginning of each project. A team's *GPA average* was computed by averaging the members' GPA's. The *Team diversity* score was operationalized as the inequality of GPAs among team members. Inequality was calculated by the Gini coefficent [19] of GPA values within a group. Gini coefficient has values from 0 (members' GPA are the same, or total GPA is distributed equally among team members) to 1 (total GPA comes from only one team member).

2) Collaboration feature characteristics: The *Collaboration feature* variable consisted of seven different measures: *Message average*, *Word average*, *Reply average*, *Message similarity*, *Word similarity*, *Reply Similarity*, and *Cohesion*. The *Message average* variable was computed by simply averaging the number of messages sent by group members. Similarly, the *Word average* for the group was computed by summing all the words in the teams' communications and then dividing the total by the number of members in the group. The *Reply Average* measure was defined as a reply to a message from a member who was different than the sender. The idea behind this measure is to try and capture the level of interaction among different members of a team.

Having collected the raw counts for a group's messages, words, and replies, we then calculated a similarity index for each of these measures. Thus, *Message similarity*, *Word similarity*, and *Reply similarity* were calculated as follows:

$$similarity_{ij} = 1 - \frac{abs(r_{ij} - r_{ji})}{r_{ij} + r_{ji}} \quad (1)$$

Where r_{ij} are the messages (words, replies) sent from member i to member j. A *Member's similarity* was then obtained by averaging all the paired similarity values, as shown in equation 2.

$$similarity_i = \frac{\sum_{j \in M, j \neq i} similarity_{ij}}{|M| - 1} \quad (2)$$

Where j are the teammates of i in team M. For a group-level measure, all team member's *Member's similarity* values were averaged (see equation 3).

$$group\_similarity = \frac{\sum_{i \in M} similarity_i}{|M|} \quad (3)$$

ISSN 2395-8618

These measures were based on the similarity measure proposed by [8]. The scores on each of these measures ranged between 0 and 1, with a 1 representing perfect similarity.

Previous research has found that cohesion is related to performance [20]. Researchers have also found a relationship between Group Potency and performance. Therefore, it seemed reasonable to assume that cohesion would be related to group potency. In order to test this relationship, a *Cohesion* measure was calculated by the LSM equation as proposed in [8]. It is important to mention that researchers who have used this measure have not found a relationship between LSM-based cohesion and performance in tasks that required virtual teams to communicate using emails [5]. Nevertheless, cohesion based on LSM has been used to show a positive relationship between cohesion and performance in chat communication settings [21], [22]. Since synchronous communication (e.g. chat) generates more messages among team members than asynchronous communication [23] (e.g. email), it is possible that group chats induce more language similarity among participants, causing an increase in group cohesiveness and, in turn, effecting group performance. Thus, we believe that in the chat setting described in this paper, the LSM cohesion measure was an appropriate measure to use to determine the relationships among cohesion, group potency and performance.

3) Linguistic features: The Linguistic Inquiry and Word Count (LIWC) tool [24], was used to identify linguistic clues that could help us understand the relationship between a team's language usage and group potency. LIWC is software that analyzes text on a word-by-word basis and calculates a percentage of words falling into one of 88 different categories. It can also be used to detect whether there are specific processes that high group potency teams use more or less than low group potency teams.

All communications from the three projects were analyzed using the LIWC software with the the English dictionary. In addition, the third project was analyzed using the Spanish dictionary, since there were some messages that contained Spanish words. The Spanish counts were then matched to the corresponding English category and included in the final percentages.

IV. EXPERIMENTS

A total of 167 students participated in the three virtual software development projects. These students were, in turn, organized into 33 different teams. From this dataset, we extracted 1588 communication activities: 1193 chat messages, 388 forum posts, and 7 Wiki pages.

The pre-project survey data yielded profile information (i.e., age, GPA, etc.) for 99.4% of our participants. The missing profile information was estimated using Multiple Imputation method for missing data [25].

Group potency questionnaires were obtained from 74.85% of the participants. Using incomplete data to aggregate to the group-level will cause an overestimation of the group potency and agreement values [26]. Thus, to remove this bias, we corrected group potency values by using the Systematic Nonresponse Parameters (SNP) [27] approach for missing data. Only one team reported insufficient data to estimate a group potency level, so this team was removed from the final dataset. The Group potency average for all groups was 3.63 (SD=0.7146). The agreement within-group members was calculated by the Interrater Agreement (IRA) measure [28]. The IRA average was 77.00%.

TABLE I
REGRESSION MODELS ON GROUP POTENCY

Features	Model	Correlation	RAE
Team+Collaboration	Linear	-0.2918	112.78%
	SMO	0.1221	**97.85%**
Linguistic	Linear	0.1262	177.08%
	SMO	0.1136	178.22%

A. Regression models and Results

In order to test the strength of our model, we designed two feature sets to predict group potency:

– Team characteristics + Collaboration measures
– Linguistic features

Because group potency scores are continuous values, these feature sets were tested using two regression models: 1) Linear regression, and 2) Support Vector Machine for regression (SMO) [29].

Results from our analysis results show a low correlation between our two regression models and group potency (see Table I). Neither Team-Collaboration or Linguistic features were able to predict group potency using either Linear or SVM regression. Table I also reports the Relative Absolute Error (RAE) for each feature and model. The RAE percentage is a measure of extent to which the scheme is an improvement over using the average to predict the outcome variable; a scheme is considered better than average when the RAE percentage is lower than 100%. In Table I, the RAE percentages are above 100 for all the measures, except for the SMO regression model with Team + Collaboration features, which is only slightly better than the average.

A closer inspection of the group potency values for each team revealed that two of the teams appeared to have abnormal averages for their group potency levels (see the two points on the left in Figure 1). We confirmed that these two teams were indeed outliers according to the Chauvenet's criterion [30]. Thus, we removed these two teams from our dataset and did a second analysis. The group potency mean for the 31 remaining teams was 3.74 (SD=0.5682).

The results from the second analysis are presented in Table II. As shown in the table, RAE percentages improved for all models, but the overall correlations between the predictor variables and group cohesion were still low. The SMO technique again produced the best predictive model, but the improvement over using averages was only 8.69%.

ISSN 2395-8618

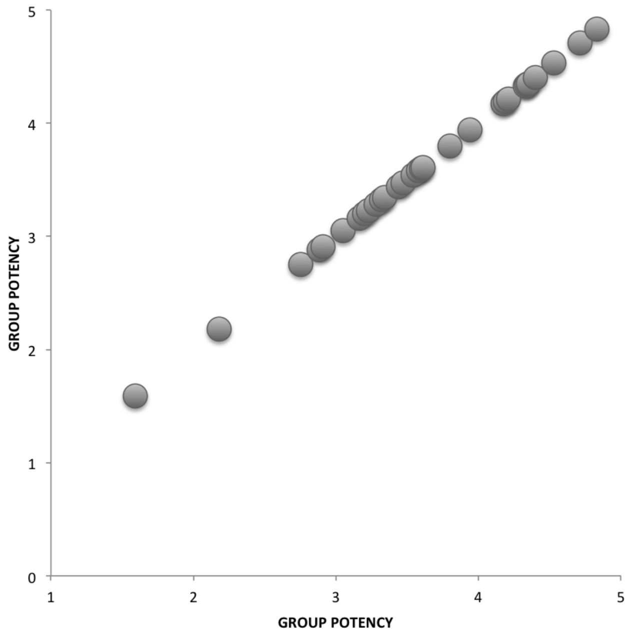

Fig. 1. Group potency values

TABLE II
REGRESSION MODELS (WITHOUT OUTLIERS) ON GROUP POTENCY

Features	Model	Correlation	RAE
Team+Collaboration	Linear	0.1108	98.20%
	SMO	0.3063	**91.31%**
Linguistic	Linear	-0.0486	153.09%
	SMO	0.081	130.34%

B. Binary classification

In order to improve upon our techniques for creating a model to predict group competency levels, several machine learning classifiers were used to test the predictive power of our variables. Since one of our objectives is to provide useful information about a group's internal state or status to managers or project leaders, we decided to convert the team data to a binary problem. Therefore, the group potency data was transformed to achieve better results for binary classification. All thirty-one groups were first ordered according to their group potency level scores. The top fifteen teams were labelled as the *High potency* group, and the bottom fifteen teams were labeled as the *Low potency* group. One team was removed from the analysis in order to maintain a balanced dataset. This new dataset was then analyzed using two common classifiers and two ensemble methods: 1) Support Vector Machine (SMO), 2) Naive Bayes (NB), 3) Bagging-REPTree (Bag), and 4) AdaBoost-DecisionStump (Ada).

Table III shows a comparison of the RAE percentages for the four different classifiers. As is normal, each classifier's accuracy rates are also reported. As anticipated, the RAE percentages for all the classifiers are much lower than in the previous experiments, indicating that our features were

TABLE III
CLASSIFICATION OF GROUP POTENCY LEVELS

Features	Classifier	Accuracy	RAE
Team+Collaboration	SMO	**70.00%**	**59.31%**
	NB	56.66%	88.10%
	Bag	43.33%	98.56%
	Ada	56.66%	94.96%
Linguistic	SMO	40.00%	118.63%
	NB	50.00%	102.66%
	Bag	53.33%	95.15%
	Ada	**70.00%**	**73.36%**

TABLE IV
ENSEMBLE METHODS ON GROUP POTENCY'S CLASSIFICATION

Features	Classifier	Accuracy	RAE
Team+Collaboration	Bag-SMO	**73.33%**	**67.22%**
	Ada-SMO	63.33%	80.88%

much more accurate at predicting group potency when using the binary classification methods as opposed to regression techniques. We also observed that the SMO classifier was a better model for predicting group potency using the combined Team-Collaboration features, whereas the Ada classifier was a better predictor when using only the Linguistic features. These same results are reflected in the accuracy rates reported in Table III.

Since the SMO classifier appeared to outperform the other classifying techniques, we then tried to improve the predictive capabilities of this classifier by adding some additional "boosting" power in the form of the AdaBoost and Bagging methods (with Team + Collaboration features). Results of these analyses are presented in Table IV. As reported in Table IV, only the Bag-SMO classifier showed an improvement in the accuracy rate of the classifier. However, the Bag-SMO classifier had a higher RAE percentage. A closer look at the outputs of the two classifiers showed that the SMO classifier was much better at predicating whether an instance was going to be *low potency* versus *high potency*. On the other hand, the Bag-SMO classifier was much better at identifying the exact potency level of an instance, with probabilities ranging from 0.6 to 1. Therefore, the Bag-SMO classifier tended to have higher RAE percentages.

Finally, we tested the three best machine learning classifiers (i.e., SMO, SMO-Bag, Ada-SMO) on a combined dataset of all three predictor measures. Results (see Table V) showed that the predictive powers of these classifiers were not as high as the previous experiment. Perhaps the performance of the classifiers was affected by the normalization of the data. That is, the classifiers may have had difficulty recognizing a feature that could satisfy a "Team+Collaboration or Linguistic" condition, since all of its features were collapsed into a single feature representation.

V. FEATURES RELATED TO HIGH-LEVEL GROUP POTENCY

The high accuracy levels of our binary classifiers led to a more detailed investigation of the specific features that might

ISSN 2395-8618

TABLE V
GROUP POTENCY'S CLASSIFICATION WITH ALL FEATURES

Features	Classifier	Accuracy	RAE
Team+Collaboration+Linguistic	SMO	53.33%	92.27%
	Bag-SMO	56.66%	91.61%
	Ada-SMO	**60.00%**	**80.09%**

have facilitated (or impeded) group potency levels in teams. Therefore, we looked at the output from the best classifiers (i.e., SMO with Team + Collaboration features, and Ada with Linguistic features) and examined the feature values that were used to predict the teams with high group potency levels. Table VI presents the features related to high group potency within teams. It should be noted that the features listed within parenthesis had a negative relationship with high group potency teams.

According to the results from the SMO classifier, negative *Message similarity* and negative *Word similarity* were related to high group potency. Since these two features were correlated with one another, as well as negatively related to high group potency, these results seem to suggest that a single person within the team may have been responsible for most of the communications. This is not an uncommon occurrence in virtual student team projects where it is often the case that a single leader emerges to help manage the task. As seen in other literature [31], individuals that emerge as a leader are often the people who produce the most communications in the teams. This is generally seen as a good thing, because a leader often causes the group to work more closely with one another. The negative relationship between *Reply similarity* and group potency seems to support our emergent leadership theory and shows that such a condition can help strengthen group potency within teams.

There were three variables that had a positive relationship with high group potency levels: *Team size*, *Word average*, and *GPA*. The positive relationship between *Size* and high group potency indicates that students believe that they are more likely to complete the task with more, rather than fewer, team members. The positive relationship between *Word average* and potency levels show that more participation provokes a higher perception of group potency within the team. Finally, the positive relationship of *GPA average* indicates the importance of the skill level of the participants to the potency construct.

The AdaBoost classifier, using the Linguistic features, produced 10 Decision Stump trees. The best performing features are listed in Table VI. The results of this analysis indicate that high group potency teams use fewer "I" words than low group potency teams. According to [32], the use of pronouns tends to show a person's focus. In this context, it appears that low group potency teams pay more attention to themselves (i.e., use of I) as opposed to high potency groups who focus on other group members (i.e., use of "You"). At the same time, low potency teams tend to communicate more about personal matters, such as health (i.e., the use of "Biological Process" words), than high potency groups. The "Verb" category was also negatively related to high potency levels. In a more detailed analysis of the corresponding subcategories within the Verb category, we found that low group potency teams used a much higher percentage of verbs related to the past and present subcategories than high group potency teams. On the other hand, high group potency teams used a higher percentage of verbs related to the future category than low group potency teams. It has been reported that future-oriented words can be be linked to performance indicators [5]. Thus, it is possible that high potency teams tend to use more future verbs because they are more focused on the project's tasks.

The literature on LIWC [32] also argues that the use of the other pronouns, such as "you," indicates that a speaker is more socially oriented. Thus, it appears to high group potency teams are more social than low group potency teams because of their more frequent use of "You" words. The positive relation between high group potency and prepositions suggests these high potency teams exchanged more complex information about a topic [33] than teams with low potency levels.

VI. CONCLUSIONS

In this study, we examined models for predicting group potency levels using aggregated variables that captured team characteristics, collaborative behaviors, and language use. At the same time, we explored a number of modeling techniques to determine which method would yield more accurate results when using typical global software development data to predict group potency. A series of virtual software development projects were developed to collect collaboration data through a distributed collaborative software system. These projects involved students from the US, Turkey, and Panama who worked together in distributed teams. Data obtained from groups' communication activities and surveys were used to predict the group potency construct.

Initial results involving the regression approach showed only a slight improvement over using the mean group potency score. Therefore, the group potency prediction task was converted to a binary classification problem, and several machine learning algorithms were tested and compared. The Bag-SMO classifier yielded the highest accuracy rates (i.e., 73.33%) using the Team + Collaboration feature dataset, while the SMO classifier had the lowest RAE percentage (i.e., 59.31%) on this same dataset. The AdaBoost (Decision Stump) classifier showed the highest accuracy rate (70%) using the Linguistic feature dataset. One explanation for the differences between the RAE percentages with continuous data versus the binary classifiers is that the transformation of the data into two groups allowed the differences among the different predictor variables to emerge. In a similar manner, the reason that the accuracy levels among the machine learning algorithms differed when using the two different features sets (i.e., Team + Collaboration versus Linguistic features) is that the Team + Collaboration features were much more correlated with one another than the Linguistic features. Thus, the results from the different

ISSN 2395-8618

TABLE VI
TOP FEATURES FROM BEST BINARY CLASSIFIERS FOR HIGH GROUP POTENCY TEAMS

Algorithm	Features
SMO with Team + Collaboration	(Message similarity), (Word similarity), Size, Word average, (Reply similarity) and GPA average
Ada with Linguistic	(I), (Biological processes), (Verbs), You, Prepositions

classifiers seemed to be affected by high (or low) variability in the two feature datasets.

The results from the machine learning models were then used to identify the particular values of the linguistic features that were used to predict group potency among team members. These results showed that high group potency teams sent fewer messages and seemed to be more diverse in their language use and message replies than low group potency teams. One explanation for these differences is that high potency teams may have had a leader (which we dubbed as "emergent") who, while dominating the conversation, was able to engender confidence among group members. Not surprising, the Collaborative measures of Size and GPA were also related to high potency group levels.

Results from our linguistic analysis indicated that high potency teams tended to be more focused on their team members (hence the use of "You" words) and communicated more about the task and future events than low group potency teams. In contrast, low potency teams talked more about themselves (hence the use of "I" words) and personal matters and tended to focus on the present and the past.

Although our initial attempt to predict group potency was not successful, we were able to obtain reasonable results by converting the task into a binary classification problem. Despite problem's conversion to binary classification reduce its outcome values, we believe that being able to predict low or high group-potency levels among global teams and provide this information to their corresponding leaders may result in proper interventions to reach a higher distributed team performance.

ACKNOWLEDGMENTS

The first author thanks Veronica Perez-Rosas for her help in the classifiers' selection for some experiments. Also, he gratefully acknowledges financial support from a CONACYT scholarship and from the Support for Graduate Studies Program of SEP. This material was also based upon work supported by the National Science Foundation under Grant No. 0705638.

REFERENCES

[1] J. D. Herbsleb and D. Moitra, "Global software development," *Software, IEEE*, vol. 18, no. 2, pp. 16–20, 2001.

[2] P. J. Agerfalk, B. Fitzgerald, H. H. Olsson, and E. Ó. Conchúir, "Benefits of global software development: the known and unknown," in *Making Globally Distributed Software Development a Success Story*. Springer, 2008, pp. 1–9.

[3] A. M. Townsend, S. M. DeMarie, and A. R. Hendrickson, "Virtual teams: Technology and the workplace of the future," *The Academy of Management Executive*, vol. 12, no. 3, pp. 17–29, 1998.

[4] R. P. Biuk-Aghai and S. J. Simoff, "An integrative framework for knowledge extraction in collaborative virtual environments," in *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work*, ser. GROUP'01. New York, NY, USA: ACM, 2001, pp. 61–70. [Online]. Available: http://doi.acm.org/10.1145/500286.500298

[5] S. A. Munson, K. Kervin, and L. P. Robert Jr, "Monitoring email to indicate project team performance and mutual attraction," in *Proceeding of CSCW'14, 17th ACM conference on Computer supported cooperative work & social computing*, 2014, pp. 542–549.

[6] G. Chen, C. Wang, and K. Ou, "Using group communication to monitor web-based group learning," *Journal of Computer Assisted Learning*, vol. 19, no. 4, pp. 401–415, 2003.

[7] P. A. Gloor and Y. Zhao, "Tecflow-a temporal communication flow visualizer for social networks analysis," in *CSCW'04 Workshop on Social Networks*, 2004.

[8] A. L. Gonzales, J. T. Hancock, and J. W. Pennebaker, "Language style matching as a predictor of social dynamics in small groups," *Communication Research*, vol. 37, no. 1, pp. 3–19, Feb. 2010.

[9] J. Mathieu, M. T. Maynard, T. Rapp, and L. Gilson, "Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future," *Journal of Management*, vol. 34, no. 3, pp. 410–476, Jun. 2008.

[10] A. De Jong, K. De Ruyter, and M. Wetzels, "Antecedents and consequences of group potency: A study of self-managing service teams," *Management Science*, vol. 51, no. 11, pp. 1610–1625, 2005.

[11] A. M. Hardin, M. A. Fuller, and J. S. Valacich, "Measuring group efficacy in virtual teams new questions in an old debate," *Small Group Research*, vol. 37, no. 1, pp. 65–85, Feb. 2006. [Online]. Available: http://sgr.sagepub.com/content/37/1/65

[12] A. E. Akgün, H. Keskin, J. Byrne, and S. Z. Imamoglu, "Antecedents and consequences of team potency in software development projects," *Information & Management*, vol. 44, no. 7, pp. 646–656, 2007.

[13] S. M. Gully, K. A. Incalcaterra, A. Joshi, and J. M. Beaubien, "A meta-analysis of team-efficacy, potency, and performance: interdependence and level of analysis as moderators of observed relationships," *Journal of Applied Psychology*, vol. 87, no. 5, p. 819, 2002.

[14] R. A. Guzzo, P. R. Yost, R. J. Campbell, and G. P. Shea, "Potency in groups: Articulating a construct," *British Journal of Social Psychology*, vol. 32, no. 1, pp. 87–106, 1993.

[15] N. Sivasubramaniam, W. D. Murry, B. J. Avolio, and D. I. Jung, "A longitudinal model of the effects of team leadership and group potency on group performance," *Group & Organization Management*, vol. 27, no. 1, pp. 66–96, Mar. 2002. [Online]. Available: http://gom.sagepub.com/content/27/1/66

[16] S. W. Lester, B. M. Meglino, and M. A. Korsgaard, "The antecedents and consequences of group potency: A longitudinal investigation of newly formed work groups," *Academy of Management Journal*, pp. 352–368, 2002. [Online]. Available: http://www.jstor.org/stable/10.2307/3069351

[17] A. D. Stajkovic, D. Lee, and A. J. Nyberg, "Collective efficacy, group potency, and group performance: meta-analyses of their relationships, and test of a mediation model," *Journal of Applied Psychology*, vol. 94, no. 3, p. 814, 2009.

[18] D. M. Paskevich, L. R. Brawley, K. D. Dorsch, and W. Neil, "Relationship between collective efficacy and team cohesion: Conceptual and measurement issues," *Group Dynamics: Theory, Research, and Practice*, vol. 3, no. 3, pp. 210–222, 1999.

[19] T. Ogwang, "A convenient method of computing the gini index and its standard error," *Oxford Bulletin of Economics and Statistics*, vol. 62, no. 1, pp. 123–129, 2000.

[20] S. M. Gully, D. J. Devine, and D. J. Whitney, "A meta-analysis of cohesion and performance effects of level of analysis and task interdependence," *Small Group Research*, vol. 26, no. 4, pp. 497–520, 1995.

ISSN 2395-8618

[21] V. L. Schwanda, K. Barron, J. Lien, G. Schroeder, A. Vernon, and J. T. Hancock, "Temporal patterns of cohesiveness in virtual groups," in *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, ser. CSCW'11. New York, NY, USA: ACM, 2011, pp. 709–712.

[22] Y. R. Tausczik and J. W. Pennebaker, "Improving teamwork using real-time language feedback," in *Proceedings of Human Factors in Computing Systems (CHI)*, 2013, pp. 459–468.

[23] F. C. Serçe, K. Swigger, F. N. Alpaslan, R. Brazile, G. Dafoulas, and V. Lopez, "Online collaboration: Collaborative behavior patterns and factors affecting globally distributed team performance," *Computers in Human Behavior*, vol. 27, no. 1, pp. 490–503, Jan. 2011.

[24] J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth, "The development and psychometric properties of LIWC2007," *Austin, TX, LIWC. Net*, 2007.

[25] R. R. Hirschfeld, M. S. Cole, J. B. Bernerth, and T. E. Rizzuto, "Voluntary survey completion among team members: Implications of noncompliance and missing data for multilevel research," *Journal of Applied Psychology*, vol. 98, no. 3, pp. 454–468, 2013.

[26] D. A. Newman and H.-P. Sin, "How do missing data bias estimates of within-group agreement? Sensitivity of SD WG, CVWG, rWG(J), rWG(J)* , and ICC to systematic nonresponse," *Organizational Research Methods*, vol. 12, no. 1, pp. 113–147, Jan. 2009.

[27] D. A. Newman, C. Lance, and R. Vandenberg, "Missing data techniques and low response rates," *Statistical and methodological myths and urban legends*, pp. 7–36, 2009.

[28] M. K. Lindell, C. J. Brandt, and D. J. Whitney, "A revised index of interrater agreement for multi-item ratings of a single target," *Applied Psychological Measurement*, vol. 23, no. 2, pp. 127–135, Jun. 1999. [Online]. Available: http://apm.sagepub.com/content/23/2/127

[29] A. J. Smola and B. Schoelkopf, "A tutorial on support vector regression," 1998, NeuroCOLT2 Technical Report NC2-TR-1998-030.

[30] H. D. Young, *Statistical treatment of experimental data*. McGraw-Hill, 1962.

[31] I. Brooks and K. Swigger, "Using sentiment analysis to measure the effects of leaders in global software development," in *International Conference on Collaboration Technologies and Systems (CTS)*, 2012, pp. 517–524.

[32] J. Pennebaker, *The Secret Life of Pronouns: What Our Words Say About Us*. Bloomsbury USA, 2013.

[33] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of language and social psychology*, vol. 29, no. 1, pp. 24–54, 2010.