Learning the Creative Potential of Students by Mining a Word Association Task

Cristian Olivares-Rodríguez
Engineering Faculty, Universidad Andres Bello
Avenida República 237
Santiago 8370146, Chile
colivares@unab.cl

Mariluz Guenaga
DeustoTech - Deusto Institute of Technology - University of Deusto
Avenida Universidades 24
Bilbao 48007, Spain
mlguenaga@deusto.es

ABSTRACT

Creativity is a relevant skill for human beings in order to overcome complex problems and reach novel solutions based on unexpected associations of concepts. Thus, the education of creativity becomes relevant, but there are not tools to automatically track the creative potential of learners over time. This work provides a novel set of behavioural features about creativity based on associative skills. These associations are processed to define two models that depict students' creative potential. This way, we have reached an acceptable accuracy rate in the classification of creative potential, hence we have found concrete evidence regarding the ability to automatically predict the creative potential of students based on their association capabitilies.

1. INTRODUCTION

Creativity generally emerges when people face a problematic or new situation, where constraints and concepts are probably unknown. An intensive search of novel solutions is required to solve real problems. This search can be done by exploration, transformation or combination of concepts. Therefore, there is a need of new associations of concepts to reach unknown solutions [7].

Nowadays, students are the centre of their learning when solving authentic problems, and creative skills provide students adaptation abilities to overcome heuristic environment, where nor the path to the solution nor the solution are known and therefore, you need to establish strategies to achieve the goal. In this context, the intensive use of technology by students to produce web searches and social data is a rich source of information to learn how students behave [2], thus a monitoring framework of creativity, based on associative features, becomes feasible.

A set of creative challenges have been applied to 64 students of sixth grade in two primary schools in Spain. First, we have applied an "unusual uses" test as a measure of creativity. After that, we have applied two "word association" tasks in order to depict their associative skills over time, similarly to [4]. Based on data captured from these activities we have extracted a set of relevant features regarding to creativity in order to model the user behaviour.

Our hypothesis bases on the fact that the local frequency of words and the time when they came out provide relevant information about originality. Also, we set that time provides a measure about fluency and that part of speech gives information about flexibility.

We have tested the strength of features associated to creativity with a supervised classification approach. We propose a model to track the creative potential of students based on their associative skills, but it still requires a more powerful set of semantic features and a learning algorithm that works properly with sequential data. However, the developed model provides an acceptable accuracy rate (over 81% in the best case) and outperform a Bag-of-Words approach.

(Sec-2) describes the concept of creativity and provides formal definitions of existing models (Sec-3). (Sec-4) defines the experiment carried out to collect data and evaluates the predictability of proposed models. Then, we present the main results using these models (Sec-5) and provide a discussion regarding the monitoring of creativity (Sec-6). Finally, we summarize our contributions and outline future works (Sec-7).

2. BACKGROUND

Creativity is a mental process based on associations in our mind and it has been characterized by: fluency, flexibility, originality and elaboration. The conceptual model of creativity of Amabile [1] defines a general process to solve problems that is grouped in three phases: a *conceptualization* phase to establish several problem definitions; a *search* phase to reach concepts, make new associations and establish new solutions; and a *development* phase to implement a solution and to update the knowledge. The model also introduces relevant skills related to creativity: associative and executive, which have been studied against the creative performance [4].

Mednick [7] proposed an associative theory of creativity where affirms that a creative person is able to find new solutions to real problems making as many associations as possible (fluency), as diverse as possible (flexibility) and as unexpected as possible (originality). Benedek et al. have found a positive relation between fluency and creativity [4].

There are some procedures to measure the creativity [8] [4]. Generally, it is measured by an *unusual uses* test, where each participant must achieve as many uses as possible for a particular object (e.g. a brick) in a short period of time, and the expert evaluation of creativity is based on a Likert-scale of fluency, flexibility and originality. This methods do not include measures of the creative potential obtained during data from the process to carry out the activity (i.e. search on the Web, social networks, etc.) and, thus there is an opportunity to model the creativity to implicitly depict students' behaviour.

A *word association* task makes visible the association skills by retrieving as many words as possible with respect to a query word, in a short period of time. This task is an heuristic process of word retrieval, where a user defines an association model Q^u in order to provide words w_i, in a certain time t_i and related to a query word q^s (Eq-1).

$$Q^u(q^s) = [(w_1, t_1), (w_2, t_2), \dots, (w_n, t_n)] \forall w_i \in W^u \tag{1}$$

Moreover, each user defines an heuristic measure h_u^* based on a hidden similarity measure S^u (Eq-2).

$$h_u^*(q^s, w) = S^u(q_j^s, w|t) \mid u \in U \tag{2}$$

Even heuristic is hidden, we can derive an empirical model through a set of association features [4].

2.1 Computational Models of Creativity

In art, a search behaviour analysis based on a visual creative task has been developed to figure out the hidden process [5]. A computerized aesthetic composition task was implemented in order to capture the search flow followed by each participant to design a new image. Thus, the user actions are used to depict the heuristic applied in the search process.

In education of creativity, a personalized creativity learning system (PCLS) based on decision trees has been proposed [6]. Nevertheless, they are not focused in track the creativity, but in enhance the process to teach creativity. The purpose of the PCLS is to adapt the student path based on a set of creativity measures and demographic information (gender, college, etc.).

3. MODEL OF WORD ASSOCIATIONS

We have defined two user models: the Bag-of-words (U_{BoG}) and the Features of Creativity (U_{FoC}).

3.1 Bag-of-Words for Association Tasks

The bag-of-words (BoW) is a basic model to describe the content of documents in the information retrieval domain (Eq-3).

$$BoG(d_j) = (f(w_1), f(w_2), \dots f(w_n)) \forall w_i \in W \tag{3}$$

Where d_j is a document, $f(w_i)$ is a function that defines the relation of the word w_i with the document d_j and w_i is in the dictionary W. This model provides a measure of the originality of words. A document d_j can be defined as the set of all associated words that users have provided against a query word q^s (Eq-4).

$$d_j(q^s) = \bigcup_{k=1}^{|U|} Q^k(q^s) \tag{4}$$

And, the word frequency wf (Eq-5) is defined as an originality measure of the word regarding each document and a relative time measure [3].

$$wf(w_i) = \sum_{j=1}^{|D|} \left[\frac{f(w_i, d_j)}{\max\{f(w, d_j) | w \in d_j\}} \times \frac{t_i}{|T|} \right] \tag{5}$$

We define a model of the user based on BoW (Eq-6), where $\overline{W}^u$ is the set of all words provided by the user u.

$$U_{BoW} = \bigcup_{i=1}^{|W|} \begin{cases} wf(w_i) & , w \in \overline{W}^u \\ 0 & , otherwise \end{cases} \tag{6}$$

3.2 Features of Creativity

We measure *Fluency* as the time variance of each query word of the user as you can see in the Eq-7, where T^u is the set of timestamp of each answer of the user u to the query word q^s.

$$t_v^u(q^s) = var[(t_1), (t_2), \dots, (t_n)] \forall t_i \in T^u \tag{7}$$

We measure *Flexibility* as the variance in the Part of Speech (PoS) of associated words (Eq-8), where W^u are the answers of the user u to the query word q^s.

$$PoS_v^u(q^s) = var[PoS(w_1), \dots, PoS(w_n)] \forall w_i \in W^u \tag{8}$$

We define *Originality* through two features based on the word frequency: 1) the variance of the word frequency (Eq-9), where W^u is the set of answers of the user u to the query word q^s.

$$wf_v^u(q^s) = var[f(w_1, d_j), \dots, f(w_n, d_j)] \forall w_i \in W^u \tag{9}$$

And 2) the dot product of frequency and time (Eq-10)

$$t.f^u(q^s) = [\frac{t_1}{|T|}, \ldots, \frac{t_n}{|T|}] \cdot [f(w_1, d_j) \ldots f(w_n, d_j)] \quad (10)$$

Hence, we define a feature vector $fv^u(q^s)$ integrating the Equations 7, 8, 9 and 10, as follows:

$$fv^u(q^s) = [t_v^u(q^s), PoS_v^u(q^s), wf_v^u(q^s), t.f^u(q^s)] \quad (11)$$

Finally, we model the user behaviour U_{FoC} (Eq-12) concatenating the feature vector fv^u of each association task (Eq-11), where $|Q|$ is the set of all association tasks driven by the query words q_i^s.

$$U_{FoC} = \bigcup_{i=1}^{|Q|} fv^u(q_i^s) \quad (12)$$

4. EXPERIMENTAL SETUP

This work aims to model the hidden heuristic in association tasks based on the behaviour of creative people. We have designed a practical experiment based on a Web platform to generate a novel dataset that relates associative skills of users and their creative potential. We applied this experiment on sixth grade students from two different schools in Spain, in a relation of 67% from one school and 33% from the other. The whole sample was composed by 47% of male and 53% of female.

The experiment involved two creative challenges developed during a class: *unusual uses* and *word association* tasks. First, the users were asked to: *write down as many unusual uses as possible for the object 'Shoe' during 60 seconds.* With this task we captured data about the divergent thinking potential of each user and, thus, we can compute a measure of their creative potential. The users were also asked to: *write down as many associated words as possible for a 'Book' ('Door') during 60 seconds.*

In order to form the dataset, the platform has registered the query object, the unusual uses listed by students and the timestamp for each use. It also has saved the query words, the associated words provided by students and the timestamp of each word. These data is represented by equation 1. Demographic data collected from each student includes age, gender and country.

In addition, a label about their creativity was provided based on the unusual uses challenge and the intrinsic characteristics of creativity. Two reviewers labelled each user as creative or non-creative using a Linkert-scale (5) of flexibility, fluency and originality. Accordingly, a labelled dataset was defined to perform a supervised learning of the creative behaviour of users. The dataset structure is depicted in the Table 1.

We have modelled the user behaviour (Sec-3) using the modified Bag-of-Words (BoW) and the Feature of Creativity (FoC). We have designed a two-class supervised learning experiment and trained a set of learning algorithms: Naïve Bayes (NB), Decision Tree ($dTree$), Support Vector Machine (three kernels) and Random Forest ($rTree$). In order to evaluate the accuracy of the learning algorithms we performed a cross-validation method. Thus, we have iteratively divided the dataset in k subsets, where the $k-1$ subsets were used to train the algorithms and the last one was used to validate the prediction quality based on its accuracy. Finally, we have performed an analysis of accuracy results against the percentage of the instances used in the cross-validation method.

Table 1: User information in dataset

Attribute	Description
`Gender`	The user gender
`Age`	The user age
`Country`	The user country
`Creative tag`	Creative (+) or No creative (-)
`Unusual Uses`	A set of tuples *(use, time)* per object
`Associations`	A set of tuples *(word, time)* per query

Table 2: Dataset statistics

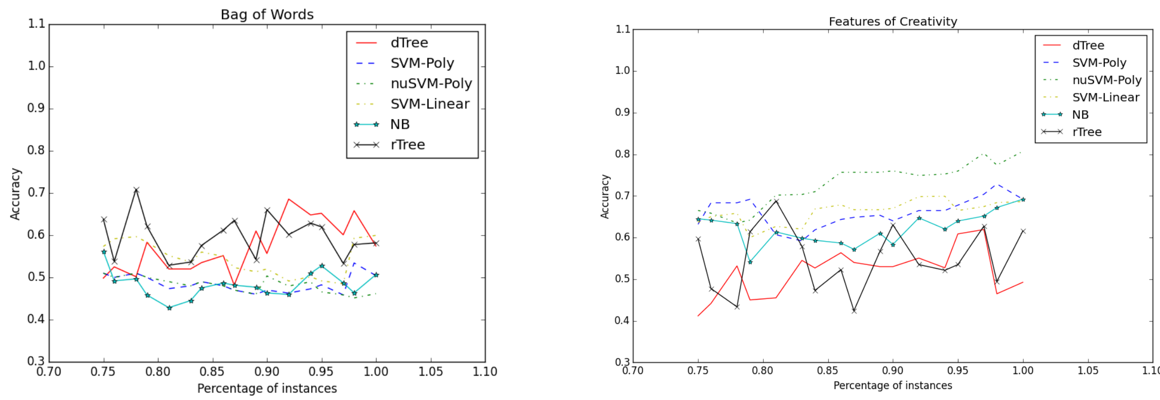
Avg. Attr. per minute	Creative (53%)		No Creative (47%)	
	M (35%)	F (65%)	M (60%)	F (40%)
#Uses 'Shoe'	3.33	4.20	**4.88**	4.10
#Asoc 'Book'	**6.67**	**6.05**	5.30	5.73
#Asoc 'Door'	**5.92**	**5.77**	4.5	4.18
	Age of Students			
	11	12	11	12
#Uses 'Shoe'	3.94	3.33	3.91	4.00
#Asoc 'Book'	**6.42**	4.67	5.52	5.33
#Asoc 'Door'	**5.93**	4.67	4.72	5.33
Diccionary Size (# unique words)				
Global	'Libro'		'Puerta'	
247	127		129	

5. RESULTS

By applying the challenges, a dataset was defined based on Eq-1. We highlight that creative students are more fluent than non-creative ones and younger students provide more associated words per minute. We have also defined a global dictionary with all associated words W provided by users and local dictionaries (W^{q^s}) for each association task. A more detailed information is shown in the Table 2

We have analysed the size of the dataset, because the features are based on statistics. In the figure 1a you can see the accuracy of the U_{BoW} model, which approximately ranges in 10 points at each model. The results of the model are similar for different sizes of the dataset, so this model can be seen independent of the size of the dataset. In the figure 1b we show the accuracy of the U_{FoC} model, which generally increases with respect to the size of the dataset, except in the case of the tree-based method. This model can be seen as dependent of the dataset size and it should improve as the dataset grows.

The most stable algorithms are the kernel-based (SVM) because they fit more precisely with the features of creativity.



(a) (b)

Figure 1: Accuracy performance against different sizes of dataset: The strength of a) U_{BoW} and b) U_{FoC}.

Also, we reach high levels of accuracy in the classification of creative behaviour based on a simple set of features and a moderate number of available samples, which reach up to 81% in the U_{FoC}.

6. DISCUSSION

The work by Jennings et al. is a not context free proposal (art) and, it is too invasive for students [5]. The associative actions of users are mined to figure out the hidden strategy of users through a design task. We define a model based on a ordinary task of word association that is common in problem-solving contexts, web searches and social networks. Therefore, the provided models can be applied in active learning contexts where students make associations. The proposal of Lin et al. [6] is seeking to improve the learning of creativity by recommendation in a personalized tutoring system. We propose a complementary work to identify the creative potential and, thus, it could be possible to provide better learning paths to students based on such prediction.

The bag-of-words approach U_{BoW} has reached an acceptable accuracy level, but it has a high variance. The features of creativity approach U_{FoC} is more stable and it has a growing accuracy along the number of samples. This model is based on a small number of features, which are highly related with the theoretical features of creativity: fluency, frequency and originality.

7. CONCLUSION

We have proposed two user models to identify creative students when they associate words: U_{BoW} and U_{FoC}. These models outline the creativity of students over time by exploiting their word associations (Web search, Social Network, etc). Thus, we depicted that it is possible to learn a classifier based on associative features with an acceptable accuracy. We have developed a dataset that relates the association skills and the creative potential of students.

In the future we will integrate the sequentially of associations, so there is the possibility to use sequential learning algorithms. The flexibility could be described using the sense/meaning of the word as a more informative similarity. Finally, we have depicted that a higher number of instances improves performance, then a more diverse set of samples should be considered.

8. ACKNOWLEDGMENTS

This work is partially founded by the Erasmus Mundus SUD-UE program.

9. REFERENCES

[1] T. Amabile. The social psychology of creativity: A componential conceptualization. In *Journal of Personality and Social Psychology*, volume 45, pages 357–376, 1983.

[2] P. André, M. c. Schraefel, J. Teevan, and S. D. T. Discovery is never by chance: designing for (un) serendipity. *C and C '09*, pages 305–314, 2009.

[3] R. E. Beaty and P. J. Silvia. Why do ideas get more creative across time? an executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts*, 6(4):309–319.

[4] M. Benedek and A. C. Neubauer. Revisiting mednick's model on creativity-related differences in associative hierarchies. evidence for a common path to uncommon thought. *The Journal of Creative Behavior*, 47(4):273–289, 2013.

[5] K. E. Jennings, D. K. Simonton, and S. E. Palmer. Understanding exploratory creativity in a visual domain. In *Proceedings of the 8th ACM conference on Creativity and cognition*, pages 223–232. ACM, 2011.

[6] C. F. Lin, Y.-C. Yeh, Y. H. Hung, and R. I. Chang. Data mining for providing a personalized learning path in creativity: An application of decision trees. *Comput. Educ.*, 68:199–210, Oct. 2013.

[7] S. Mednick. The associative basis of the creative process. In *Psychological review*, volume 69, pages 220–232, 1962.

[8] E. P. Torrance. The nature of creativity as manifest in its testing. In *The Nature of Creativity: Contemporary Psychological Perspectives*. Cambridge University Press, New York, 1988.