Using Feelings in Images to Support Storytelling Games

Greice Pinho Dal Molin
Faculdade de Informatica
Pontificia Universidade Catolica
do Rio Grande do Sul, PUCRS
Porto Alegre, RS, Brazil
greice.molin@acad.pucrs.br

Soraia Raupp Musse
Faculdade de Informatica
Pontificia Universidade Catolica
do Rio Grande do Sul, PUCRS
Porto Alegre, RS, Brazil
soraia.musse@pucrs.br

***Abstract*—Research has emerged in the area of story-based game management with respect to components that direct the player to choices to experience games in a variety of ways, not leaving the game monotonous. The purpose of this paper is to suggest a new approach of interactive game model, which allows the player to add images indicating their feelings regarding the game narrative as a condition for the player to be directed to a new specific scenario in the game. To achieve this, pattern recognition techniques in images are applied to classify the emotion/feeling in images regarding its polarity (positive and negative), including neutrality. By using storytelling the player unleashes his/her ability to create and reinvent each move in the game, by exploring its limits.**

***Keywords*-storytelling; drama manager; interactive narrative; sentiment analysis in images.**

I. Introduction

One of the most important aspects in the interaction that happen among game players is the exchange of thoughts and experiences. Therefore, there are many interfaces handled in games which use computer vision and special hardware that can help to provide such interaction. In particular, texts and images may reflect a feeling about a certain subject, expressing a favorable, unfavorable or neutral opinion. The user interaction, through images, can be useful to point out the user feeling, maybe providing changes in the game narrative.

The purpose of this paper is to suggest a new approach to interactive game models, allowing the player to add images indicating his/her feelings regarding the game narrative as a condition to be directed to a new specific scenario in the game. To achieve this, pattern recognition techniques in images are applied to classify the emotion in images regarding its polarity (positive and negative), including neutrality.

The remainder of this paper is organized as follows: Section II presents the basis of the work so that we can propose a game model using image analysis of feelings in this work, and Section III describes the game model, using the flowchart for its representation. Experiments with image classification are presented in Section IV and some analyzes are discussed in the Section IV-A. Sections V discuss final comments and future work.

II. Related Work

Many studies in the area of classification of feelings using texts have been proposed mainly in the context of social networks [1]. However, the classification of feelings from images and even videos is still under development.

Riedl and Bulitko [2] define telling stories in human culture as a narrative, also known as a cognitive tool, for understanding various subjects and organizing experiences in a narrative format. It is a fascinating area in computer systems because it is the way to make the human interact with stories using computers. There are many activities in the area of narrative intelligence focused on the computational perspective, one of the most interesting applications is the interactive narrative. In such area, users create or influence a plot through actions. It is considered a form of interactive entertainment, but can be used for education and training. Indeed, the most common form of interactive narrative involves the user being the protagonist in a story.

Li [3] proposes the narrative intelligence as an effective way of simulating human intelligence because it has the ability to create, tell, understand, and respond to stories. The major challenge in this area is the knowledge-intensive nature to be able to generate and tell stories in any domain. Using the proposed Open Story Generation system, the players can learn about a particular domain. They can also use crowd sourcing to build a common opinion about everyday activities like going to a restaurant or going to a movie theater.

Regarding related work on feeling analysis in image, Vadicamo et al. [4] propose a visual sentiment classifier starting from a large set of user-generated and unlabeled contents. They compared studies and evaluations for visual sentiment analysis and found out that texts are noisy and weakly correlated with the image content, but it can be further exploited by using a deep Convolutional Neural Network.

Although there are some works in the area of this article, our main contribution is to propose a structure to classify feelings in images used to indicate the next stage of the game. The intention is to try to capture the player's emotion to continue entertaining.

III. The Game Model

In this section, we present our game narrative proposal using feelings detection in images analysis. The model developed in this article aims to assist in the immersion of the individual in each stage of the game. The user actively participates in the indication of one or several images to demonstrate their feelings in order to build or influence the new narrative. In this way, the story continues from the perspective of the player, taking into account their actions to continue the game. The main objective is to propose the interaction of the player in the game through the polarity ratings that are determined by the images reported by the player for the definition of the new scenario. With this, we believe that the player should feel more integrated with the fictional virtual world. Next section presents the used method to classify feeling in images.

A. The VGG-T4SA model

In the case of emotions classification in images, we use the convolutional neural network proposed in [4] and trained with the dataset called Twitter for Sentiment Analysis (T4SA) dataset, also proposed by the authors. The trained visual sentiment classifier, named VGG-T4SA is used to comparative studies and evaluations in the present work. T4SA is composed of about 1 million tweets for which the authors provide textual sentiment classification, and the corresponding 1.4M images. All this data is publicly available [1]. The neural network VGG [5] is a convolutional neural network of image classification. Given an image, the VGG network produces probabilities of the different classes that an image may belong to. For instance, regarding specifically the present paper, the VGG network indicate that a image has a chance of 80% confidence that it presents a positive emotion, 10% of chance of presenting negative emotions as well as 10% having a neutral emotion.

B. A Game Model as Interactive Narrative

The first step is for the player to be introduced to the goal of the game. It can be done using any known technique to familiarize the player with the context, Then the player indicates one or n-images, for example in Figure 2, at any plot in the game, to demonstrate his/her feelings about the narrative presented. The classifier indicates the polarity of each user-defined image. If there is more than one image, the polarity of each image will be defined and quantified to define the highest polarity defining the user's emotion. With the player's emotion set, the next plot of the narrative will be made available.

The image polarity detection method, in this case, serves as an interface to analyze information about the game's last plot, as well as to record the emotional aspects of the players during the game, being contextualized in space and time. In addition, the image polarity is used to activate the next plot in the game narrative. For example, during the game, a set of information about the player's profile and their perceptions about the activities performed during the game and between the narratives can be made available for analysis. From this analysis, some factors such as learning situations and skills developed can be worked out.

In Figure 1 we show the main components involved in the suggested game model, as following described:

- **Read Game History:** This is the story processing part, which is based on a configuration file containing the stories.The file is structured with 4 columns, the first being the plot; the second column indicates the negative class and contains the next plot to be played if the player's emotion is defined by the chosen images; the third column contains the next plot if the class is neutral and the third column contains the next plot if the class is positive. Each of next plot that then player can interact with images is structure in the same way, having the polarity possibilities as trigger to new plots.
- **Choose image:** the player can interact with one or n-images in each interactive plot step. For each selected image, we classify the emotion and trigger next step of the game. In addition, each interaction, classification, time and space where it occurred along the game is recorded for further statistical analysis.

[1] http://www.t4sa.it/

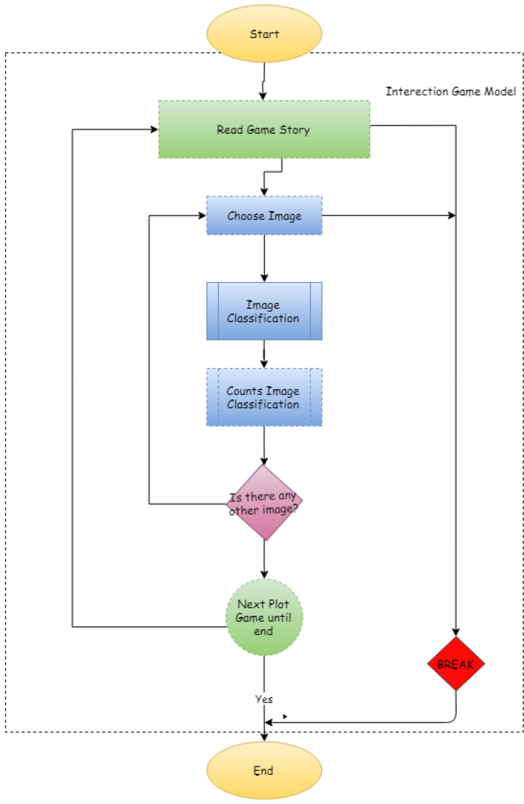

Figure 1. A game model for use in story telling for the player to decide the next steps of the narrative.

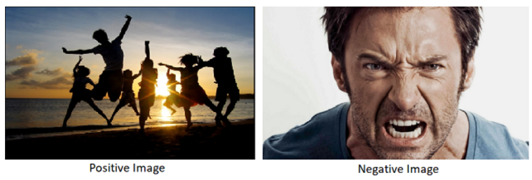

Figure 2. Examples of positive and negative images that can be chosen by the player

- **Is there any other image?** in this step it is verified if there is more than one image to be classified.If yes, the other images are also classified, so the method should decide the emotion based on more than one image by simply averaging the obtained images classification.
- **Next Plot Game until end:** according to the result of the classification the player will be sent to the next plot of the game narrative.
- **Break:** it is possible to stop the game if a break indication is activated. This is modeled in order to have some escape for non adequate images, for instance.

The game implementation is an ongoing work, so in next section we present some results obtained with the model for feelings detection in images.

IV. Experiments

For the evaluation of images classification to be performed in the game, the following steps are performed:

1) We used 571 searched images from Google [2]. The search criteria used was the search for images including emotions that demonstrate a positive or negative feeling. The positive emotions were searched using the following keywords: amusement, contentment, excitement and awesome. For the negative ones we used fear, anger, sad and disgust. The evaluation was done based on a corpus created from a Brazilian blog dataset [6] including texts and images. This corpus is composed of 880 texts and images and manually evaluated using Figure Eight framework [3]. The total amount of images is divided into training and test, being 456 and 115, respectively.
2) Manipulation of uploaded images: Resize the image to the size of 256x256 because it is the resolution used in VGG-T4SA.
3) Classification Methods: We used 4 very known methods as presented in literature: SVM [7], Ada Boost [8], Decision Tree [9] and Random Forest [10].
4) Results are stored for further analysis.

The images collected, according to the criteria already explained, present only positive and negative emotions. Table I presents the classification methods used in the experiments and their parameters. The results do not show many variations between the classification methods used in terms of precision, recall, F1-Scores, considering the 571 samples collected, being 20% of the sample for testing.

Table I
Methods and parameters for classification

METHODS	PARAMETERS	VALUE
SVM	kernel	linear or rbf
SVM	C	1.5
SVM	gamma	1.0
SVM	degree	5
Decision Tree	criterion	entropy
Decision Tree	max_depth	1
Random Forest	criterion	entropy
Random Forest	n_estimators	2000
AdaBoost	default	default
	TOTAL OF PARAMETERS:	**9**

Table II presents a list of images classified as positive and Table III presents a list of the images classified as negative.The order of the names in the table (from top to the bottom) is vertical, while in the figure it is horizontal (from left to the right and also top to the bottom).

A. Preliminary Results

The results do not show many variations between the classification methods used in terms of precision, recall, F1-Scores, considering the 571 samples collected, where 20% is used for testing. However, we find that the emsemble methods have the best percentages as we can see in /Tables IV, V, VI, VII. For this reason, for an improvement of the prediction algorithm, we would use the Random Forest method.

[2]http://www.google.com

[3]https://www.figure-eight.com/

Table II
Results of the selected images for positive classification.

	IMAGE	PREDICTION
1	awe_0849.jpg	Positive
2	excitment_0180.jpg	Positive
3	amusement_0544.jpg	Positive
4	awe_0585.jpg	Positive
5	contentment_0068.jpg	Positive
6	excitement_0266.jpg	Positive
7	excitement_0208.jpg	Positive
8	amusement_0539.jpg	Positive
9	contentment_0102.jpg	Positive
10	excitement_0235.jpg	Positive
11	awe_0577.jpg	Positive

Table III
Results of the selected images for negative classification.

	IMAGE	PREDICTION
1	fear_google2.jpg	Negative
2	sad_0420.jpg	Negative
3	anger_006.7jpg	Negative
4	sad_0421.jpg	Negative
5	fear_0971.jpg	Negative
6	disgust_0765.jpg	Negative
7	disgust_0790.jpg	Negative
8	fear_0684.jpg	Negative
9	sad_0330.jpg	Negative

Table IV
Results of selected images for classification, using as input the correct value of the polarity of the image and the prediction made by the method SVM

	PRECISION	RECALL	F1-SCORE	SAMPLE
Negative	0.56	0.90	0.69	20
Positive	0.87	0.48	0.62	27
avg/total	0.74	0.66	0.65	47

Table V
Results of selected images for classification, using as input the correct value of the polarity of the image and the prediction made by the method Random Forest

	PRECISION	RECALL	F1-SCORE	SAMPLE
Negative	0.59	0.95	0.73	20
Positive	0.93	0.52	0.67	27
avg/total	0.79	0.70	0.69	47

Table VI
Results of selected images for classification, using as input the correct value of the polarity of the image and the prediction made by the method Desicion Tree

	PRECISION	RECALL	F1-SCORE	SAMPLE
Negative	0.55	0.90	0.68	20
Positive	0.86	0.44	0.59	27
avg/total	0.72	0.64	0.63	47

The results of Table V, referring to the Random Forest method, can be interpreted as follows: Once precision is defined as the relation $\frac{tp}{(tp+fp)}$, where tp is the number of true positives and fp the number of false positives, it is intuitively the ability of the classifier not to label as positive a sample that is negative. In the same way, recall is $\frac{tp}{(tp+fn)}$ and this metric is intuitively the ability

Table VII
RESULTS OF SELECTED IMAGES FOR CLASSIFICATION, USING AS INPUT THE CORRECT VALUE OF THE POLARITY OF THE IMAGE AND THE PREDICTION MADE BY THE METHOD ADABOOST

	PRECISION	**RECALL**	**F1-SCORE**	**SAMPLE**
Negative	0.58	0.95	0.72	20
Positive	0.93	0.48	0.63	27
avg/total	0.78	0.68	0.67	47

of the classifier to find all positive samples. In addition, f1-score provides the harmonic mean of precision and recall. The scores for all classes report the accuracy of the classifier in comparison to all other classes. Support indicates the amount of data used in each class and its total. Accuracy is the proportion of correct predictions, regardless of what is positive and what is negative. It is a measure susceptible to data set unbalance and may lead to a wrong conclusion about performance. The calculation of the accuracy occurs as follows: $\frac{(tp+tn)}{(p+n)}$.

By analyzing Table V, it is possible to see that the algorithm has a precision of 79% in the prediction of the image classes, since it is the average between the percentages of the predictions of true positive and true negative. As for the Recall metric, which shows the ability of the classifier to find all positive samples, it indicates that 95% of the image predictions were correctly performed for the negative class and 52% for the positive classes. That goes for Recall's average of 70%. The F1-Score presents the harmonic mean across the classes for the precision and recall metrics, resulting in a negative percentage of 73% and 67% positive class for the negative class. This information is based on the sample informed in support, in which the negative class has 20 images and the positive class 27. The last line of Table V provides a weighted average of the precision, recall and f1-score, in which the weights are the values of the support. Thus, for accuracy, the mean is $\frac{(0.59\times20+0.93\times27)}{47} = 0.79$. The same calculation is executed for recall, indicating that mean$=0.70$ and F1-Score$=0.69$. Finally, we can conclude that the algorithm presented accuracy of 69% of classification, being below that exposed in Section II. We have to consider that we tested it with "wild" images, i.e. coming from Internet without any pre-selection procedure. We are investigating the possibility of replacing the method VGG-T4SA by another one that could provide better accuracy, it is a future work.

V. FINAL CONSIDERATIONS

We present a way of user interaction using images classification that can be used to improve the way players interact in a game narrative. Indeed, emotions in images are used to change characters personalities and other virtual character parameters. Our goal is to provide changes in game narrative when detecting images polarity as informed by the player. In this way, the player will actively participate in the game story, enriching his/her experiences and being more involved with the game. In this paper, we showed the achieved results of 59% of precision in negative images and 93% in positive ones. Images were selected from internet based on a simple selection using keywords: "positive images" or "negative images". Future work include to test another CNN to emotion in images to be used in a game model. As future work, we will develop the proposed model, containing the three categories of games (entertainment, educational, training), using multimodal resources such as audio and video, as well as images.

REFERENCES

[1] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic, "A survey of multimodal sentiment analysis," *Image and Vision Computing*, vol. 65, pp. 3–14, 2017.

[2] M. O. Riedl and V. Bulitko, "Interactive narrative: An intelligent systems approach," *Ai Magazine*, vol. 34, no. 1, p. 67, 2012.

[3] B. Li, M. Thakkar, Y. Wang, and M. O. Riedl, "Storytelling with adjustable narrator styles and sentiments," in *International Conference on Interactive Digital Storytelling*. Springer, 2014, pp. 1–12.

[4] L. Vadicamo, F. Carrara, A. Cimino, S. Cresci, F. DellOrletta, F. Falchi, and M. Tesconi, "Cross-media learning for image sentiment analysis in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 308–317.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[6] H. D. P. dos Santos, V. Woloszyn, and R. Vieira, "Blogset-br: A brazilian portuguese blog corpus," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, N. C. C. chair), K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, and T. Tokunaga, Eds. Paris, France: European Language Resources Association (ELRA), may 2018.

[7] S. Zhao, H. Yao, Y. Gao, G. Ding, and T.-S. Chua, "Predicting personalized image emotion perceptions in social networks," *IEEE Transactions on Affective Computing*, 2016.

[8] J. C.-W. Chan and D. Paelinckx, "Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sensing of Environment*, vol. 112, no. 6, pp. 2999–3011, 2008.

[9] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote sensing of environment*, vol. 86, no. 4, pp. 554–565, 2003.

[10] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006.