Gamification for Machine Learning: The Classification Game

Giorgio Maria Di Nunzio
Dept. of Inf. Eng. (DEI)
University of Padua
dinunzio@dei.unipd.it

Maria Maistro
Dept. of Inf. Eng. (DEI)
University of Padua
maistro@dei.unipd.it

Daniel Zilio
Dept. of Inf. Eng. (DEI)
University of Padua
daniel.zilio.3@studenti.unipd.it

Abstract

The creation of a labelled dataset for machine learning purposes is a costly process. In recent works, it has been shown that a mix of crowd-sourcing and active learning approaches can be used to annotate objects at an affordable cost. In this paper, we study the gamification of machine learning techniques; in particular, the problem of classification of objects. In this first pilot study, we designed a simple game, based on a visual interpretation of probabilistic classifiers, that consists in separating two sets of coloured points on a two-dimensional plane by means of a straight line. We present the current results of this first experiment that we used to collect the requirements for the next version of the game and to analyze i) what is the 'price' to build a reasonably accurate classifier with a small amount of labelled objects, ii) and compare the accuracy of the player to the state-of-the-art classification algorithms.

1 Introduction

Supervised machine learning algorithms require labelled examples to be trained and be evaluated properly. However, the labelling process is a costly, time-consuming and non-trivial task. Manual annotation by experts is the obvious choice [24], but it is slow and expensive. In the last years, mixed approaches that use crowd-sourcing [13] and active learning [19] have shown that it is possible to create annotated datasets at affordable costs. In this paper, we want to apply game mechanics to the problem of classification of objects, a supervised machine learning problem, with a two-fold goal in mind: i) how the gamification of a classification problem can be used to understand what is the 'price' of labelling a small amount of objects for building a reasonably accurate classifier, ii) to analyze the classification performance given the presence of small sample sizes and little training [3, 10].

In this first pilot study, we designed a simple game based on a visual interpretation of probabilistic classifiers [6, 5, 7]. The game consists in separating two sets of coloured points on a two-dimensional plane by means of a straight line. Despite its simplicity, this very abstract scenario can be, and will be in the next version of the game, substituted with more captivating ones (see Section 5.1). At the beginning of the game, players know nothing about the type of objects that they have to separate (in our case the objects are text documents), but they know that the points they see on the plane are a small subset of the total and their position on the plane is not accurate. Players have a limited amount of resources that can be used to improve the position of the points and/or visualize more points (see Section 3 for a detailed explanation of the game).

To summarize, this experiment has been designed to study:

- how the gamification of a classification problem can be used to understand what is the 'price' a user is willing to pay to build a classifier;
- how the performance of a 'human' classifier compares to the state-of-the-art algorithms on a small scale training dataset.

The first goal is related to the problem of minimizing the cost of labelling the dataset and at the same time

In: F. Hopfgartner, G. Kazai, U. Kruschwitz, and M. Meder (eds.): Proceedings of the GamifIR 2016 Workshop, Pisa, Italy, 21-July-2016, published at http://ceur-ws.org

building a reasonably accurate classifier. The second goal is related to the problem of classification performance given the presence of small sample sizes and little training [3, 10].

The paper is organized as follows: in Section 2, we present some background literature on gamification in Information Retrieval (IR). In Section 3, we describe in detail the game, the rules and the interactive application. Section 4 discusses the pilot study and the initial results. Section 5 is dedicated to the requirements and suggestions for future improvements collected from the players during this study. In Section 6, we give our final remarks.

2 Related Work

Gamification is defined as "the use of game design elements in non-game contexts" [4], i.e. tipical game elements are used for purposes different from their normal expected employment. In this context, we can define our web application as a Game with a Purpose (GWAP), that is a game which presents some purposes, usually boring and dull for people, within an entertaining setting, in order to make them enjoyable and to use human computation to solve problems. Nowadays, gamification spreads through a wide range of disciplines and its applications are implemented in different and various aspects of scientific fields of study. For instances, gamification is applied to learning activities [16, 15], business and enterprise [14, 22, 23] and medicine [9, 2].

In recent years, also IR has dealt with gamification, as witnessed by the Workshop on Gamification for Information Retrieval (GamifIR) in 2014, 2015 and 2016. In [12] the authors describe the fundamental elements and mechanics of a game and provide an overview of possible applications of gamification to the IR process. Moreover, [20] investigates possible approaches to properly gamify web search, i.e. making the search of information and the scanning of results a more enjoyable activity. Actually, many proposals of game applied to different aspects of IR have been presented. For example, [18] describes a game that turns document tagging into the activity of taking care of a garden, with the aim of managing private archives. In [17], the authors propose a method to obtain ranking of images by utilizing human computation through a gamified web application. Finally, [11] introduces a strategy to gamifying the annotation of a French corpora. Although a lot of gamified applications were presented in the IR field in the last few years, to the best of our knowledge there are no applications to machine learning and in particular to machine classifiers.

3 The Classification Game

The game is based on the two-dimensional representation of probabilities [6, 21] which is a very intuitive way of presenting the problem of classification on a two-dimensional space. Given two classes c_1 and c_2, an object o is assigned to category c_1 if the following inequality holds:

$$\underbrace{P(o|c_2)}_{y} < m \underbrace{P(o|c_1)}_{x} + q \tag{1}$$

where $P(o|c_1)$ and $P(o|c_2)$ are the likelihoods of the object o given the two categories, while m and q are two parameters that depend on the misclassification costs that can be assigned by the user to compensate for either the unbalanced classes situation or different class costs.

If we interpret the two likelihoods as two coordinates x and y of a two dimensional space, the problem of classification can be studied on a two-dimensional plot. The decision of the classification is represented by the 'line' $y = mx + q$ that splits the plane into two parts and all the points that fall 'below' this line are classified as objects that belong to class c_1 (see Figure 1 for an example). Without entering into the mathematical details of this approach [6], the basic idea of the game is that players can adapt the two parameters m and q in order to optimize the separation of points and, at the same time, can use their resources to improve the estimate of the two likelihoods by buying training data, and/or add more points to the plot by buying validation data.

3.1 Rules of the game

The game is organized in 10 levels, which are presented from the easiest to the most difficult and which correspond to the different classification tasks of the top 10 classes of the Reuters 21578 dataset[1]. A level is difficult when it is hard to linearly separate the positive and the negative class (class c_1 and c_2 in Equation 1), i.e. when the positive and negative class partially overlap, and/or when there are few examples of objects in the positive class. We used the standard Reuters ModApte split to obtain the training and test documents for the ten classes. An objects in the training set can be used during the game either as a training example or a validation sample, but not both. In this way we are setting the machine learning problem in terms of a hold-out method that, despite being not very accurate compared to a cross-validation, is much easier to use in a game.

[1] `http://www.daviddlewis.com/resources/testcollections/reuters21578/`

The goal of each level (an in general of the game) is to find the best classifier, i.e. the one which maximizes the F1 score, with the least amount or resources. Resources (we intentionally did not use the word credits or money) can be spent to increase the training and the validation set. At the beginning of the game, the players has already a free 10% of the collection annotated: 5% used for training and 5% used for validation. At any point in the game, the player can use some resources to buy additional training or validation objects. When he/she selects the training option, an additional 5% of the collection is added to the training, this action will set the points more precisely (because probabilities are estimated more accurately), while when he/she chooses the validation option, an additional 5% is included in the validation set and more points will appear in the plot. Notice that, when the player starts the game, he/she is provided with a sufficient amount of resources to 'buy' the whole dataset for each level. In this way, the player does not have to think about saving resources for difficult levels. There is no limit to the time the user can spend in a level since the important part of the game is to understand the amount of resources that the player consider sufficient to solve the problem, not how fast he/she can do it. Once the player has found what he/she considers the best classifier, he/she can proceed with the test, thus the classifier is tested on the test set and the F1 score is computed. At this point, the level is completed and the player is forced to go to the next level or conclude the game.

3.2 Interactive Application

We have implemented the game of classification with the Shiny package in R [1], and the source code of the application is freely available for download[2]. This interactive Web application can be used as a show case online, but we had to use a local version in order to avoid lagging and server disconnection that would have made the game very annoying for our players. The objects plotted on the two-dimensional space are news agencies from the Reuters newswires dataset.

In Figure 1, you can see the layout of the web application for this pilot study. The interface can be divided in three main areas: the left panel, which contains buttons and controls to adjust the parameters of the classifier, the top right panel, which displays the plot of the two classes, and the bottom right area, which summarizes the employed resources and the F1 scores.

In the left side panel, there is a text field that allows the player to choose a username, and two buttons to start the game and to move to the next category. Below these two buttons, there is a field which indicates the remaining resources, and the available 'clicks', to training/validate the model for the current level. The player can utilize the positive and negative check boxes to visualize only the positive and negative points respectively. The player can center the line by clicking directly on the plot image and then shift and rotate it by using the sliders, which corresponds to the q and m parameter of Equation 1. Finally, we added a button that allows the player to go back to the best parameters setting that he/she has found during the current level. The color of the button indicates whether it is active (blue) or disabled (red).

In the top right panel, there is the main window that shows the two setsw of points: light blue for the positive class and red for the negative one. The dark blue line is the decision line used to split the two classes. In the left upper corner of the plot, there are three numbers: the first one in yellow is the current F1 score, the second one in blue is the best F1 score obtained by the user in the current class with the amount of training and validation set, and the red one is the value that the player should 'beat' [3].

Finally, below the plot area, on the right bottom box, there is a summary of the user actions and results. The histogram describes the amount of resources spent in training actions, with the green color, and validation actions, with the pink color. For each class, the F1 score obtained on the test set is reported below the corresponding column. At the end of the game, i.e. when the user has completed all the levels, he/she will see his/her mean F1 score, averaged on each category, and the mean goal score, which is the average of the F1 scores obtained by the automatic algorithm on each category.

4 Pilot Study

A pilot study was carried out to test this preliminary version of the game and to collect opinions and suggestions regarding possible improvements of the game. The experiment was conducted on a sample of 20 student and researchers of the University of Padua. As future work, we aim at spreading this game through the use of social media in order to collect a bigger and a more diversified sample of users.

The majority of the users that participate in this test had just a naïve idea of what machine learning is and how an automatic classifier operates. Since the majority of our users were not machine learning expert, we provided them a brief explanation of the prob-

[2] https://github.com/gmdn/Classification

[3] This value corresponds to the best score of the Bernoulli Naïve Bayes classifier trained on whole dataset. We used the results reported in [6] and we rounded down some values in order to motivate the players.

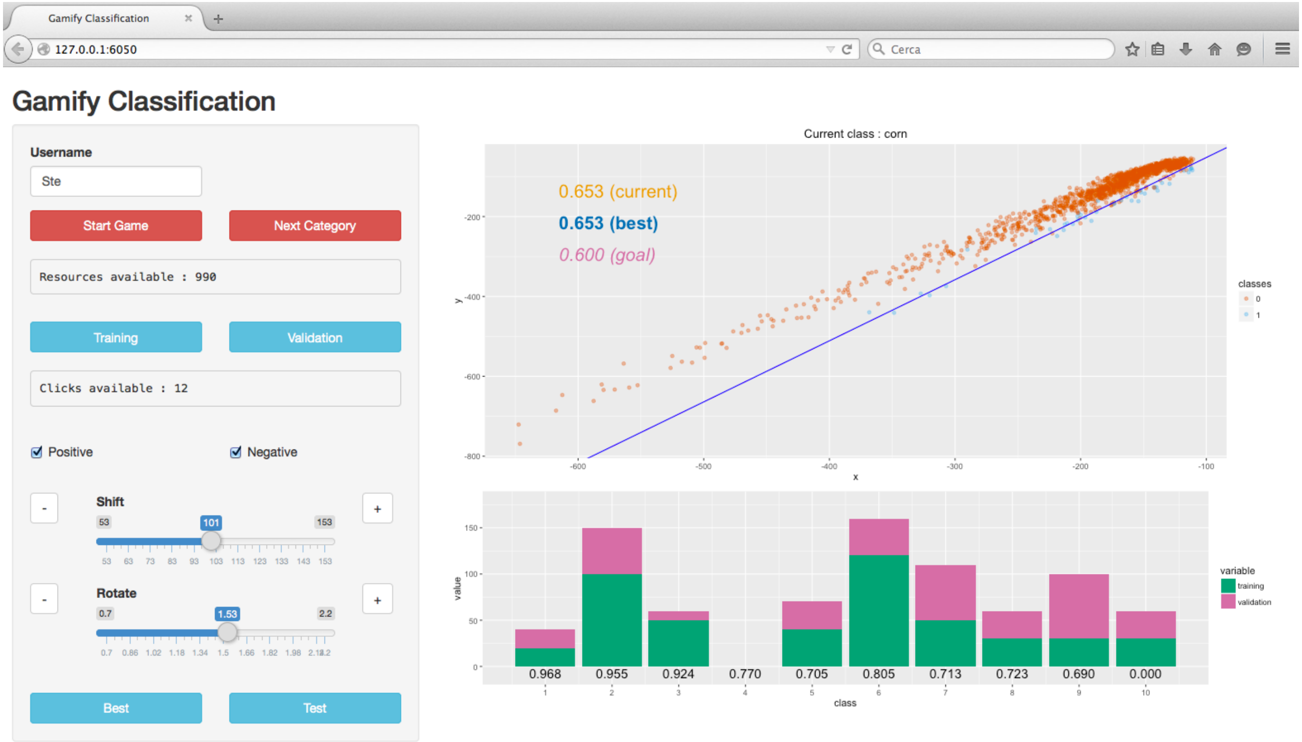

Figure 1: Layout of the web application

lem and the fundamental concepts (especially about training and validation) before starting the game. Then, we introduced them to the interface as described in section 3.2.

4.1 Results

In our experiments, for each player we collected the F1 score for each class and the amount of resources used. For a comparison with the state-of-the-art, we trained on the exact same training and validation set used by each player for each class a SVM with linear kernel using the 'kernlab' package in R[4] together with the 'caret' package[5]. For a fair comparison, since the players acted as 'optimizers' of the Naïve Bayes classifier decision, we also optimized the SVM by adjusting the cost parameter C within the range $[0.01, 0.05]$ (smaller or larger values of C did not produce any significant change).

In Table 1, we report the F1 measure for each player on each level. The name of the column are the original names of the top 10 Reuters-21578 classes. The last three column shows the average of the F1 measure across the classes for each player, the average F1 measure of the SVM across the data of each player, the percentage of resources used by each player in a game. The last two rows represents: the average F1 score for each class for the players and for the SVM.

[4] https://cran.r-project.org/web/packages/kernlab/
[5] https://cran.r-project.org/web/packages/caret/

In this initial analysis, we were impressed by two results: on average, the players could beat the 'goal' score more easily than expected, which means that the probabilistic classifiers can be trained/validated with just 25% of the original dataset and obtain in many cases even better results than a cross-validation on the whole dataset. We will investigate this problem in future works. The second interesting aspect is that SVM performed as well as the whole dataset with only 25% of the annotated dataset and without cross-validation. This results is very promising since, potentially, the gamification problem may give a strong indication about how to stop the labelling process and use the annotated dataset to train with a very high accuracy state-of-the-art-algorithm. This second part will require a deep analysis and further experiments to confirm the statistical significance of this process.

5 Further Developments

During the game and at the end of each game session, we discussed together with each player about improvements and issues of the interface and the game in general. We report in this section a summary of the discussions.

5.1 Game Scenarios

Together with the users who participated to this pilot study, we started to sketch some possible scenarios of

Table 1: F1 results for each player and level from easiest to hardest. Average performance of the SVM on the same training validation is shown for each player and level. Last column shows the resources used by each player.

username	earn	acq	grain	crude	money.fx	ship	wheat	interest	trade	corn	average	SVM	resources
airamoigroig	0.97	0.95	0.88	0.83	0.76	0.78	0.74	0.74	0.78	0.56	0.80	0.86	29%
Alan	0.96	0.91	0.79	0.67	0.75	0.68	0.68	0.63	0.76	0.59	0.74	0.85	16%
Ale	0.94	0.95	0.83	0.86	0.76	0.79	0.77	0.63	0.71	0.49	0.77	0.85	18%
CalebTheGame	0.96	0.90	0.88	0.87	0.76	0.84	0.77	0.75	0.76	0.61	0.81	0.87	56%
ClaudioBarba	0.96	0.96	0.85	0.83	0.72	0.74	0.75	0.69	0.63	0.63	0.78	0.80	12%
dz	0.96	0.95	0.85	0.79	0.74	0.76	0.72	0.74	0.74	0.62	0.79	0.83	17%
edoardo_verona	0.97	0.96	0.86	0.89	0.75	0.80	0.76	0.74	0.75	0.57	0.80	0.87	42%
Erica	0.96	0.94	0.82	0.78	0.75	0.72	0.59	0.67	0.59	0.53	0.74	0.84	23%
gadaleta	0.96	0.96	0.91	0.87	0.74	0.81	0.70	0.73	0.72	0.53	0.79	0.87	26%
Giada	0.96	0.95	0.83	0.82	0.75	0.80	0.71	0.74	0.67	0.62	0.78	0.87	23%
Hector	0.95	0.95	0.90	0.80	0.74	0.86	0.77	0.56	0.60	0.56	0.77	0.86	31%
jeppy	0.96	0.95	0.81	0.67	0.69	0.74	0.60	0.65	0.71	0.51	0.73	0.79	1%
ottoX8	0.97	0.94	0.89	0.71	0.75	0.78	0.75	0.75	0.67	0.56	0.78	0.82	26%
pil	0.95	0.93	0.84	0.83	0.72	0.82	0.71	0.73	0.60	0.61	0.77	0.84	12%
poiopoio	0.95	0.95	0.85	0.83	0.77	0.77	0.73	0.73	0.65	0.52	0.78	0.85	17%
power23	0.96	0.95	0.86	0.90	0.73	0.78	0.74	0.77	0.68	0.52	0.79	0.86	28%
renberche	0.96	0.95	0.88	0.77	0.71	0.75	0.76	0.72	0.72	0.55	0.78	0.87	39%
signoraMaria	0.97	0.96	0.85	0.88	0.78	0.78	0.74	0.74	0.72	0.60	0.80	0.84	23%
Ste	0.97	0.96	0.92	0.77	0.71	0.80	0.71	0.72	0.69	0.59	0.78	0.86	45%
veronica	0.97	0.96	0.89	0.87	0.73	0.82	0.69	0.74	0.47	0.54	0.77	0.83	12%
average	0.96	0.95	0.86	0.81	0.74	0.78	0.72	0.71	0.68	0.57			
SVM	0.97	0.95	0.89	0.85	0.73	0.82	0.89	0.74	0.82	0.80			

this game that will improve the game experience. We have come up with four possible alternatives:

- Plants and gardening: we have a field sown with different types of seeds, but we do not know exactly where these seeds are. Some of the seeds will grow into edible plants, others will grow into weeds. The goal is to build a fence that separated the field in a way that "our" part of the field will contain the most of the edible plants and the least of the weeds. We can ask some help to our animal granivore friends (like pigeons and doves) to fly over and check some part of the fields to see whether the seeds are good or not.

- Gold mines: there is an area with a lot of gold nuggets as well as useless common stones and you are the first explorer to mine this area. Your resources are limited, and you can only choose one part of the area while the rest remains untouched. The goal is to choose the area with loads of gold and the smallest number of stones. You have a friend who is an expert in gold mining and can probe the area to understand whether there is gold or stone.

- Aerial warfare: in this war scenario, we have an army that is involved in a military zone, and we are forced to perform a raid to seize an area in order to secure the zone. The goal is to send our air forces to clear the area that contains the most of the enemies and the least of our ground troops and civilians. Before the raid, we can send our helicopters to explore the area and check the current situation.

- Plastic and Glass Recycling: after a music festival in a park, you have to collect all the bottles and cans that have been left on the field. You have limited resources and you can only split the area in two sides: one side should contain more plastic bottles than glass bottles and the other side more glass bottles than plastic bottles. You have some kids that can help you to spot where are the parts with more glass or plastic.

5.2 Game Controls

We received some very good feedback about game controls and interaction and how to improve them to obtain a better feeling of the game.

5.2.1 Main Window

Most of the players would prefer a full screen window to see the points more clearly without any distraction. The amount of credits used per class is not very relevant for their game when they play as well as the score obtained in previous classes. It would be better to have a window with the ranking of the scores that players can open when they need to see their status.

5.3 Line Control

Players began to understand the use of the sliders after a few attempts. In particular, the rotation of the line was not immediately clear since the plot is centred around the minimum and maximum values of the coordinates, while the slope is computed given the intercept of the line with the y axis. The thing that puzzled the players was the non-intuitive rotation around a point far from the plot limits. For this reason they suggested two alternatives:

- to select a fixed point within the plot (i.e. the center of the line) and to rotate the line along this point;

- to maintain the slope fixed and rotate the plane instead of the line.

Moreover, the control of the slope of the line would be easier with a "knob" rather than a slider[6]. There are also new ideas about the interactions with the game in terms of touch screen technologies. In fact, it would be much easier for the players to interact with the game with the gestures that are now "natural" on mobile devices: rotation, swipe, zooming in and out, may enhance the user experience and bring the game to a different level. Finally, line controls should be overlaid on top of the main window instead of being on one side, in this way the eye of the player does not have to move from one side to the other of the screen every time the line has to be adjusted.

5.4 Game Incentives

From the live interaction with the players during the game sessions, it was clear that one of the strongest motivations to replay the game was to have ranking of the players with the scores obtained, to know whether the friends/colleagues performed better or worse. At the time of the pilot experiments, we could give them hints about their performance compared to the other players and just that information was enough for them to sparkle their sense of competition. Some of them were willing to play a second time to just beat their competitors. This is in line with the literature on gamification [12, 8]. In addition to the ranking of the scores of 'human' players, we want to introduce the comparison of scores between each player and a set of state-of-the-art classification algorithm trained on the exact same game. We want to see how strong this incentive would be for a human to know that his performance is better or worse compared to a computer. We are also planning a set of virtual goods and badges to motivate the player during the game and after each session.

Since the competitiveness is one of the main motivations which encourages the users to play the game and reach high performances, it is very important to find the definition of a formal criterion to rate and rank players by taking into account the F1 score and the resources spent. As described in section 3.1, the goal of the game is multiobjective, indeed the main task consists in defining, at the same time, a classifier which is effective, i.e. it reaches high values of accuracy and precision, and efficient, i.e. it uses a few amount of resources in terms of training and validation.

Let $C = \{c_1, \ldots, c_i, \ldots, c_N\}$ be the set of categories, we denote with s_i the F1 score obtained on the test set of the i-th category by the player, and with g_i the F1 score obtained by the automatic algorithm that we called 'goal' score on the same category. We indicate with t_i and v_i the amount of resources spent by the user in training and validation documents respectively, and R is the total amount of resources provided at the beginning of the game, 1800 in our game. We define the user rating J as

$$J = a \sum_{i=1}^{N} \frac{s_i}{g_i} + b \left(1 - \frac{1}{R} \sum_{i=1}^{N} (t_i + v_i) \right) \qquad (2)$$

where a and b are two parameters, which range in the interval $[0, 1]$. Notice that these parameters represent the significance assigned to the two different tasks which define the game, indeed, a influences the importance of the effectiveness objective, while b determine the importance of the efficiency purpose. For this preliminary version of the game, we chose a and b with equal value, $a = b = 1$, since we consider both the efficiency and effectiveness task of equivalent significance. Future work and further applications of this game can justify the preference for one of the two tasks, motivating the choice of different weights for a and b.

6 Final Remarks

In this first pilot study of the gamification in machine learning, we set up a simple game, based on a visual interpretation of probabilistic classifiers, that consists in separating two sets of coloured points on a two-dimensional plane by means of a straight line. The 20 players that participated in this study already gave us important suggestions in order to improve both the game mechanics and the game controls. We believe that, with the right game scenario (plants or gold mine for example), this game could be easily played by users that do not need any information about training/validation.

Moreover, the classification results of the game were also very high compared to the small amount of labelled objects and we also found a very promising relation with the results of the SVM trained on the same labelled dataset. We are currently study a new version of the game with some options that implements cross validation, which would bring the machine learning aspect to a new different level.

References

[1] Winston Chang. *Shiny: Web Application Framework for R*, 2015. R package version 0.11.

[2] Yu Chen and Pearl Pu. Healthytogether: Exploring social incentives for mobile fitness applications. In *Proceedings of the Second International Symposium of Chinese CHI*, Chinese CHI '14, pages 25–34, New York, NY, USA, 2014. ACM.

[6] See for example the gallery of this type of control realized with d3js `https://radmie.github.io/ng-knob/`

[3] Giorgio Corani and Marco Zaffalon. Learning reliable classifiers from small or incomplete data sets: The naive credal classifier 2. *J. Mach. Learn. Res.*, 9:581–621, June 2008.

[4] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From Game Design Elements to Gamefulness: Defining "Gamification". In *Proc. of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek '11, pages 9–15, New York, NY, USA, 2011. ACM.

[5] Giorgio Maria Di Nunzio. Using Scatterplots to Understand and Improve Probabilistic Models for Text Categorization and Retrieval. *Int. J. Approx. Reasoning*, 50(7):945–956, 2009.

[6] Giorgio Maria Di Nunzio. A New Decision to Take for Cost-Sensitive Naïve Bayes Classifiers. *Information Processing & Management*, 50(5):653 – 674, 2014.

[7] Giorgio Maria Di Nunzio and Alessandro Sordoni. A Visual Tool for Bayesian Data Analysis: The Impact of Smoothing on Naive Bayes Text Classifiers. In *Proc. of the ACM SIGIR'12 conference on research and development in Information Retrieval, Portland, OR, USA, August 12-16, 2012*, page 1002, 2012.

[8] David Easley and Arpita Ghosh. Incentives, gamification, and game theory: An economic approach to badge design. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, pages 359–376, New York, NY, USA, 2013. ACM.

[9] Carsten Eickhoff. Crowd-powered experts: Helping surgeons interpret breast cancer images. In *Proceedings of the First International Workshop on Gamification for Information Retrieval*, GamifIR '14, pages 53–56, New York, NY, USA, 2014. ACM.

[10] George Forman and Ira Cohen. Learning from Little: Comparison of Classifiers Given Little Training. In *Knowledge Discovery in Databases: PKDD 2004, Pisa, Italy, September 20-24, 2004, Proceedings*, pages 161–172, 2004.

[11] Karën Fort, Bruno Guillaume, and Hadrien Chastant. Creating zombilingo, a game with a purpose for dependency syntax annotation. In *Proceedings of the First International Workshop on Gamification for Information Retrieval*, GamifIR '14, pages 2–6, New York, NY, USA, 2014. ACM.

[12] Luca Galli, Piero Fraternali, and Alessandro Bozzon. On the Application of Game Mechanics in Information Retrieval. In *Proc. of the 1st Int. Workshop on Gamification for Information Retrieval*, GamifIR'14, pages 7–11, New York, NY, USA, 2014. ACM.

[13] Chien-Ju Ho, Shahin Jabbari, and Jennifer Wortman Vaughan. Adaptive task assignment for crowdsourced classification. In *ICML (1)*, volume 28 of *JMLR Proceedings*, pages 534–542. JMLR.org, 2013.

[14] Jose Luis Jurado, Alejandro Fernandez, and Cesar A. Collazos. Applying gamification in the context of knowledge management. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*, i-KNOW '15, pages 43:1–43:4, New York, NY, USA, 2015. ACM.

[15] Karl M Kapp. *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons, 2012.

[16] Isabella Kotini and Sofia Tzelepi. *Gamification in Education and Business*, chapter A Gamification-Based Framework for Developing Learning Activities of Computational Thinking, pages 219–252. Springer Int. Publ., Cham, 2015.

[17] Mathias Lux, Mario Guggenberger, and Michael Riegler. Picturesort: Gamification of image ranking. In *Proceedings of the First International Workshop on Gamification for Information Retrieval*, GamifIR '14, pages 57–60, New York, NY, USA, 2014. ACM.

[18] Carlos Maltzahn, Arnav Jhala, Michael Mateas, and Jim Whitehead. Gamification of private digital data archive management. In *Proceedings of the First International Workshop on Gamification for Information Retrieval*, GamifIR '14, pages 33–37, New York, NY, USA, 2014. ACM.

[19] Burr Settles. Closing the loop: Fast, interactive semi-supervised annotation with queries on features and instances. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, 27-31 July 2011, John McIntyre Conference Centre, Edinburgh, UK, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1467–1478, 2011.

[20] Mark Shovman. The Game of Search: What is the Fun in That? In *Proc. of the 1st Int. Workshop on Gamification for Information Retrieval*,

GamifIR'14, pages 46–48, New York, NY, USA, 2014. ACM.

[21] Rita Singh and Bhiksha Raj. Classification in likelihood spaces. *Technometrics*, 46(3):318–329, 2004.

[22] Laurentiu Catalin Stanculescu, Alessandro Bozzon, Robert-Jan Sips, and Geert-Jan Houben. Work and play: An experiment in enterprise gamification. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 346–358, New York, NY, USA, 2016. ACM.

[23] Jennifer Thom, David Millen, and Joan DiMicco. Removing gamification from an enterprise sns. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 1067–1070, New York, NY, USA, 2012. ACM.

[24] Salud M. Jiménez Zafra, Giacomo Berardi, Andrea Esuli, Diego Marcheggiani, Maria Teresa Martín-Valdivia, and Alejandro Moreo Fernández. A multi-lingual annotated dataset for aspect-oriented opinion mining. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 2533–2538, 2015.