A Game of Lines:
Developing Game Mechanics for
Text Classification

Giorgio Maria Di Nunzio[1], Maria Maistro[1], and Daniel Zilio[2]

[1] Dept. of Information Engineering – University of Padua, Italy
`giorgiomaria.dinunzio@unipd.it, maria.maistro@dei.unipd.it`
[2] Dept. of Cultural Heritage – University of Padua, Italy `daniel.zilio@unipd.it`

Abstract. In this paper, we describe a set of experiments that turn the machine learning classification task into a game, through gamification techniques, and let non expert users to perform text classification without even knowing the problem. The application is implemented in R using the Shiny package for interactive graphics. We present the outcome of three different experiments: a pilot experiment with PhD and post-doc students, and two experiments carried out with primary and secondary school students. The results show that the human aided classifier performs similarly and sometimes even better than state of the art classifiers.

1 Introduction

The creation of a labelled dataset for supervised learning is slow and expensive. In the last years, mixed approaches that use crowd-sourcing and interactive machine learning [1] have shown that it is possible to create annotated datasets at affordable costs [12]. One major challenge in motivating people to participate in these labelling tasks is to design a system that promotes and enables the formation of positive motivations towards work as well as fits the type of the activity. In this context, an approach named 'gamification' has become popular. Gamification is defined as "the use of game design elements in non-game contexts" [4], i.e. tipical game elements, like rankings, leaderboards, points, badges, etc, are used for purposes different from their normal expected employment.

Nowadays, gamification spreads through a wide range of disciplines and its applications are implemented in different areas. For instance, an increasingly common feature of online communities and social media sites is a mechanism for rewarding user achievements based on a system of badges and points. They have been employed in many domains, including educational sites like Khan Academy[3], and tourist review sites like Tripadvisor[4]. At the most basic level, these game elements serve as a summary of a users key accomplishments; however, experience with these sites also shows that users will put in non-trivial

---

[3] `https://www.khanacademy.org/`
[4] `https://www.tripadvisor.com/`

amounts of work to achieve particular badges, and as such, badges can act as powerful incentives [2].

The use of gamification in academic research areas has been introduced very recently and its potential is still to be explored and validated. Information Retrieval (IR) has recently dealt with gamification, as witnessed by the GamifIR in 2014, 2015 and 2016[5]. In [10], the authors describe the fundamental elements and mechanics of a game and provide an overview of possible applications of gamification to the IR process. In [13], approaches to properly gamify Web search are presented, i.e. making the search of information and the scanning of results a more enjoyable activity. Other approaches of game applied to different aspects of IR have been proposed. For example in [11], the authors describe a game that turns document tagging into the activity of taking care of a garden, with the aim of managing private archives.

In this paper, we present the recent studies of gamification in text classification and the development of a Web application written in R with the package Shiny [3]. This application, initially designed to understand probabilistic models, has been redesigned as a game to train a text classifier with the aid of non experts, especially kids from primary and secondary schools, during the European Researchers' Night in September 2016 at the University of Padua[6]. We tested this application with a two-fold goal in mind: i) how the gamification of a classification problem can be used to understand 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.

2   The Classification Game

In this section, we present the refinements of a visualization approach of probabilistic text classifiers that was transformed into a game. The application was implemented with the Shiny package in R that allows to build interactive graphics [3]. This two-dimensional representation allows non experts to visually interact with the algorithm and, at the same time, to gather new training labels. In this section, we first describe the mathematical idea that supports the game, then we describe the rule of the game and how players can interact with the algorithm.

2.1   Math Background

The game is based on the two-dimensional representation of probabilities, also known as Likelihood Spaces [14], which is a very intuitive way of presenting the problem of classification on a two-dimensional space (full mathematical details

---

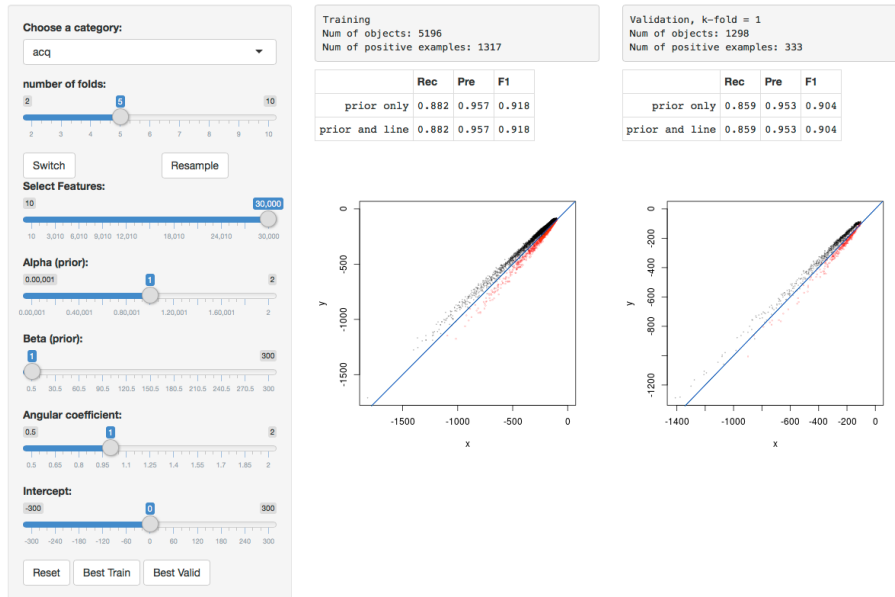[5] http://gamifir.com
[6] http://www.venetonight.it/

Fig. 1. Layout of the web application designed for experts

can be found in [8, 7, 6, 5]). 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 can be either set automatically, for example by optimizing a measure of classification accuracy, or semi-automatically by asking to a user to suggest the initial conditions based on a visual inspection of the problem. In fact, 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 where: i) the decision of the classification is represented by the line $y = mx + q$ that splits the plane into two parts, ii) the points that fall 'below' this line belongs to class c_1.

2.2 Game Mechanics

The initial version of the interface[7], shown in Figure 1, was designed to be used by experts to understand how to optimize the search of the optimal parameters.

───────────

[7] Available at `https://gmdn.shinyapps.io/shinyK/`

In the "gamified" version of this problem, players have to find the best combination of m and q having a fixed amount of resources available to train and validate the algorithm. The game is organized in N levels (that corresponds to the binary classification problems), which are presented from the easiest to the most difficult and which correspond to the different classification tasks of the top N classes of the Reuters 21578 dataset[8]. A level is difficult when it is hard to linearly separate the positive class c_1 and the negative class c_2. An object can be used during the game either as a training example or a validation sample, but not both. The goal of each level (and in general of the game) is to find the best classifier, i.e. the line which best separates the two categories, c_1 and c_2 and therefore which maximizes the F1 score, with the least amount or resources.

Resources can be used to increase the number of objects of the training and/or the validation set. At any point in the game, the player can use some resources to buy additional training or validation objects. By doing so, an additional 5% of the collection is added to either the training set (more precise) or the validation set (more objects on the screen). 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   Experiments

In the previous section, we presented how the players can interact with the classification game by "investing" a limited amount of resources to buy training and validation data and, consequently, to find a better combination of the two parameters m and q.

In this section, we present the results of three different experiments of the gamification of text classification that involved different users and different interfaces.

3.1   Pilot Experiment: PhD and Post-doc students

A second version of the interface was designed for PhD and post-doc students[9] and 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. In this first experiment, we were positively surprised by two results (a complete description of the results can be found in [9]). First, on average, the players could reach the 'goal' (i.e., the score that a state-of-the-art classification algorithm would reach with the whole labelled dataset) more easily than expected, by using only 25% of the available data. The second interesting aspect is that a support vector machine trained on the same reduced dataset (around

---

[8] http://www.daviddlewis.com/resources/testcollections/reuters21578/
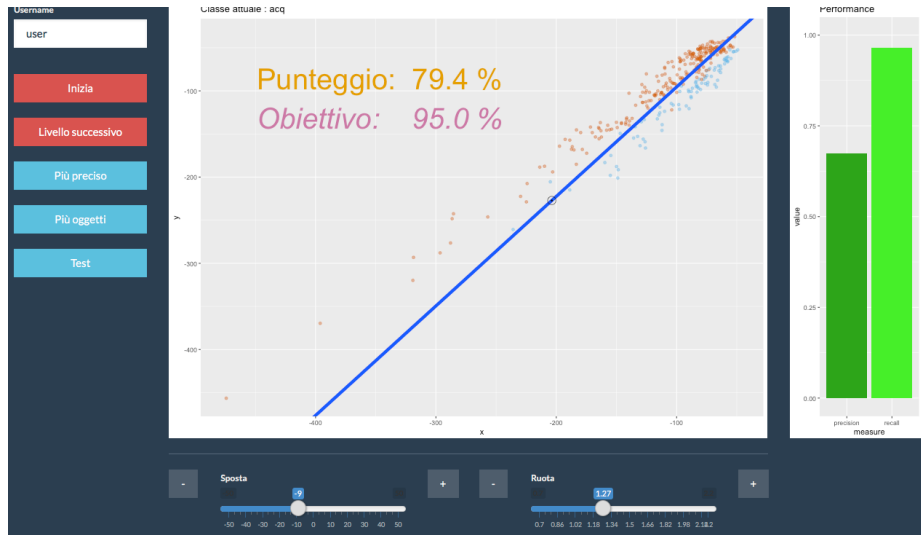[9] Available at https://gmdn.shinyapps.io/Classification/

Fig. 2. Layout of the web application designed for students

25% of the annotated dataset) performed as well as the same SVM trained on the whole dataset. This results are very promising since, the gamification of text classification may give a reliable indication about when to stop the labelling process and use the annotated dataset to train with good classification performances a state-of-the-art-algorithm. This second part will require a deep analysis and further experiments to confirm the statistical significance of this process.

3.2 Second Experiment: primary and secondary school students

During the European Researcher's Night at the University of Padua in September 2016, we designed a new interface to make the game easier for kids of primary and secondary schools who played the application. The interface, shown in Figure 2, lets users play only three levels (each level corresponds to a different category) and give feedback about the current performance whenever the line is adjusted. In this experiment, we also added some incentives like a public leaderboard that was displayed and regularly updated and chocolate candies for the top scorer. A total of 28 players used the interface.

Considering that these users did not know anything about machine learning or text classification, the results in terms of classification performance were even more surprising compared to the first experiment. In Table 1, we compare the average results of the classification performance of the players (column *manual*) with the classification performance of a Naïve Bayes classifier (NB) and a Support Vector Machine (SVM) as well as the 'goal' performance. You can notice that the results obtained by participants are very close to the one obtained with the NB and in the case of the second class, the users achieves better performances

Table 1. Manual vs NB and SVM classifiers. Classification performance during the European Researcher's Night. The averaged F1 measure of 28 participants is reported for each class.

Classes	Goal	Manual	NB	SVM
1	0.950	0.931	0.943	0.940
2	0.850	0.784	0.768	0.840
3	0.750	0.715	0.715	0.730
average	0.850	0.810	0.809	0.837

Table 2. Manual vs NB and SVM classifiers. Classification performance during the week at the Banca d'Italia. The averaged F1 measure of 27 participants is reported for each class.

Classes	Goal	Manual	NB	SVM
1	0.950	0.940	0.942	0.939
2	0.850	0.807	0.786	0.841
3	0.750	0.714	0.710	0.723
average	0.850	0.830	0.813	0.834

than NB. On average, the classifier with the human contribution is performing better than NB and worse than SVM.

3.3 Third experiment: General Public

The first week of April 2017, during an event at one of branches of Banca d'Italia in Padua for the brand new 50 euro note, we presented a third version of the game that was available for the public a whole week. For this study, we decided to make the layout cleaner, see Figure 3, and add keyboard controls to change the decision line instead of using sliders. We kept the same game incentives, chocolate candies and leaderboard, and we added an instructional presentation of the problem to help the player to understand what 'machine learning' and 'training set' are.

A total of 27 participants played with the game and their results are reported in Table 2. Even in this case the human aided classifier achieves good results and the interaction of users with the algorithm through the gamified approach reached performances close to SVM and often better than NB. In this case the results are much closer to SVM than NB even if the amount of resources used was comparable to the second experiments: players tend to consider the performance of the classifier satisfactory when 30% of the resources are used.

Finally, notice that the algorithms were trained on a different amount of data during the game, the scores in Table 1 and Table 2 are not directly comparable. This explains the different results reported for NB and SVM in Table 1 and Table 2.

Fig. 3. Layout of the web application designed for general public

4 Final Remarks and Future Work

In this paper, we presented the ongoing work on gamification for text classification that involves non expert users in the task of labelling data and produce an estimate of the monetary cost of creating the training dataset. Considering the very abstract game (a line and some dots), the first three preliminary studies were successful in terms of participation and initial results. The goals of these studies is to have feedback and collect enough data to study how to design the game in order to make it open to the general public; in addition, we want to understand whether a 'serious' game can be implemented in order to gather labelled data for machine learning.

Future work aims at extending the proposed game and transform it in an application for different mobile devices. Therefore, further effort is needed to design the interface of the mobile application with integrated environments like Unity[10]. Moreover, considering that the players are not expert in classification the rules of the game should be presented clearly and some concepts, as for example the validation phase, need to be explained in an easier way.

Finally, we aim at investigating a different game mode with two players collaborating together to reach a common goal. For instance, the users can share the controls so they need to cooperate to find the best solution, or an alternative is to assign different tasks to each user, one user will control the classification line while the other user will assess documents to help him or her to get more training examples.

---

[10] https://unity3d.com

References

1. Saleema Amershi, Maya Cakmak, W. Bradley Knox, and Todd Kulesza. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine*, 35(4):105–120, 2014.
2. Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Steering user behavior with badges. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 95–106, New York, NY, USA, 2013. ACM.
3. Winston Chang. *Shiny: Web Application Framework for R*, 2015. R package version 0.11.
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. Interactive machine learning with r. In Francesco Mola and Claudio Conversano, editors, *CLADAG 2015 10th Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society. Book of Abstracts.*, pages 333–338. 2015.
8. Giorgio Maria Di Nunzio. *Interactive Text Categorisation: The Geometry of Likelihood Spaces*, pages 13–34. Springer International Publishing, Cham, 2017.
9. Giorgio Maria Di Nunzio, Maria Maistro, and Daniel Zilio. Gamification for machine learning: The classification game. In *Proceedings of the Third International Workshop on Gamification for Information Retrieval co-located with 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016), Pisa, Italy, July 21, 2016.*, pages 45–52, 2016.
10. 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.
11. 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.
12. B. Morschheuser, J. Hamari, and J. Koivisto. Gamification in crowdsourcing: A review. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 4375–4384, Jan 2016.
13. 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.
14. Rita Singh and Bhiksha Raj. Classification in Likelihood Spaces. *Technometrics*, 46(3):318–329, 2004.