Cheap, Fast and Good!
Voting Games with a Purpose

Karën Fort, Mathieu Lafourcade, Nathalie Le Brun
STIH/Sorbonne Université, Paris, France - LIRMM, Montpellier, France - Imaginat, Lunel, France
karen.fort@sorbonne-universite.fr, mathieu.lafourcade@lirmm.fr, imaginat@imaginat.name

Abstract

We present in this paper the voting games with a purpose that were developed around JeuxDeMots, a central game aiming at creating a lexical network for French. We show that such lightweight applications can help collect quality language resources very efficiently and we advocate for a common platform for such voting games for language resources.

Keywords: crowdsourcing, GWAP, voting games

1. Introduction

JeuxDeMots[1] is a game with a purpose (GWAP) aiming at creating a lexical semantic network for French (Lafourcade, 2007). The game was created in 2007, in the wake of the ESP Game (von Ahn and Dabbish, 2004), and is therefore one of the first GWAPs for natural language processing with Phrase Detectives (Chamberlain et al., 2008), long before wordrobe (Bos and Nissim, 2015) and ZombiLingo (Guillaume et al., 2016).

Since September 2007, more than 4,000 players have registered on JeuxDeMots and 1,523,321 games have been played. As of today (January 2018), the network contains more than 2 million terms linked by more than 180 million relations.

The idea to develop complementary games came naturally, as the main game interface and features did not seem adequate to gather some specific information. More specifically, very simple click-only games, which can be played casually without registering and on smartphones, looked promising. Also, multiplying game designs would compensate for the multiple biases of the JeuxDeMots original design, hence producing wider coverage and more accurate lexical and semantic data.

The first voting game which was added aside of JeuxDeMots is PtiClic (Lafourcade and Zampa, 2009). Its aim is to distribute terms according to their relation to a given target term. The player has to click and drag terms toward the appropriate box associated to a specific semantic relation (see Figure 1). Once finished, the proposals are compared to those of others players.

Now 12 complementary games are available through a portal[2] (see Figure 2), 10 of which are simple voting games.

To our knowledge, very few voting GWAPs exist for language resources creation. Apart from the ones we just mentioned, the only GWAP that relates to this type of game is wordrobe (Bos and Nissim, 2015)[3]. On this platform, players are invited to participate to a variety of tasks, related to semantic disambiguation (noun *vs* verb, co-reference identification, named entity annotation, etc) and to choose an answer from a limited list of solutions. Although the concept resembles that of the JeuxDeMots portal, wordrobe tasks are much more complex and require more concentration and some more advanced (at least school-level) knowledge.

Figure 1: PtiClic: the term *repos* (*rest*) is the target term. Each term of the cloud should be dragged and dropped in one of the three boxes on the right-hand side.

Figure 2: The JeuxDeMots portal. Note that Totaki and top10 are not voting games.

[1] See: http://www.jeuxdemots.org/.

[2] See: http://imaginat.name/JDM/Page_Liens_JDMv2.html.

[3] See: http://wordrobe.housing.rug.nl/Wordrobe/public/HomePage.aspx.

2. A Galaxy of Voting Games

2.1. Common Features

We define voting games as very simple games in which the players have to choose between a predefined, limited number of answers, without any training. The selection (or vote) of the player is compared to those of the other players, and more specifically to the state of the resource, in order to perform two tasks: a) including the answer in the resource and b) computing some reward points, which are part of the game functionalities.

Contrary to a quiz game, in which the correct answers are known, we obviously cannot compare the votes which are cast to a reference. Therefore, the majority of answers is used to generate rewards corresponding to what is considered as the right one. It has to be noted that two games (distant in time) with the same instructions might not yield the same results, as the underlying resource might have been modified in the meantime. The created resource (in our case, the `RezoJDM` lexical-semantic network) is dynamic and evolves over time.

We decided to exclude from the definition more complex games, like the ones allowing for free-text answers, like `Totaki` (a guessing game where clues are given to the system which tries to infer the target word) or `top10` (another guessing game, where the players can identify words selected by the system from a simple definition). We also exclude games requiring training, like `Argotario` (Habernal et al., 2017).

The interface of the simplest voting games is quite easy to develop, as it generally consists of a question, a term, and a couple of buttons to choose from. Beside being simple to master, such games are also well-adapted to mobile devices (smartphones and tablets) and they can be played quickly, anytime, anywhere.

An example of such an interface is presented in Figure 3, for `LikeIt` (Lafourcade et al., 2015): the balloons represent the possible answers (in this case, "Yes, I like the idea" / "I don't mind" / "No, I don't like the idea"), the term to decide on is centered and highlighted (here, *obscurcir*, i.e. *to darken*) and the votes on the previous term (*pendre*, i.e. *to hang*) are shown in a horizontal colored bar at the top of the page.

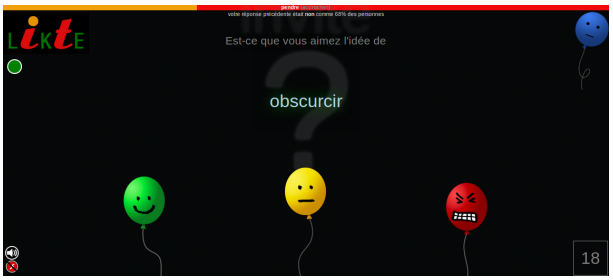

Figure 3: Interface of `LikeIt`, a polarity game with the term *obscurcir* (*to darken*).

Obviously, given the simplicity of the games, a nice design helps attracting players, so funny images are used (balloons in `LikeIt`) instead of simple buttons.

The main challenge is for the system to select adequate terms to be proposed to the player. The approach, based on the idea of potential information propagation consists of the following steps:

- identify a set of symbols/values that we want to tag the terms with. Adding a neutral value if needed. For example, in the case of `LikeIt` the values are: $\{pos\_positive, pos\_negative, pos\_neutral\}$;
- select of a term to tag. Randomly choose a target T (in a set of terms), which is already tagged (not with the neutral value). Then, there is p chance that you propose this term and $p - 1$ that you propose one of its neighbors in the network. We set p to 0.5 in our experiments.
- bootstrap by tagging manually with a non neutral value at least one word. In the case of `LikeIt`, we tagged *bon* (*good*) with one positive vote and *mal* (*evil*) with one negative vote.

This simple selection algorithm allows to crawl the network, tagging terms through a propagation approach by maximizing the chance of proposing a target term that is relevant to the task. Increasing the value of p tends to slow down the propagation but increases the number of votes for each term.

2.2. Obtained Results

As reported in Lafourcade et al. (2015), during the first 3 months of `LikeIt` more than $25,000$ terms have been polarized (i.e. tagged with a combination of positive, negative and neutral votes), with a total of over $150,000$ votes. After 7 years, more than $360,000$ terms have been polarized for 75 million votes representing 70% of the terms contained in the network at that time. The Polarimot project (Gala and Brun, 2012) aimed at building a similar resource of polarized terms, but with classical means, i.e. manually. For this project, in the course of 3 months, 3 experts tagged (with 3 votes) a set of $2,400$ terms. The comparison between the resources (Polarimot and `LikeIt`) showed that for the common terms (corresponding to the $2,400$ terms of Polarimot) the obtained polarities are almost identical. The only difference (for less than 20 terms) is a more subtle polarization for terms that are polysemous with some contrastive polarity (like *affection* that refers both to love or to disease).

`SexIt` (Lafourcade and Fort, 2014) is based on the same principle as `LikeIt` (and the same internal engine). The purpose of the game is to assess if a given term is related to sex (in its broadest meaning). As reported in Lafourcade and Fort (2014), the propagation algorithm is especially efficient in crawling the underlying network to propose relevant terms.

`Selemo` (see Figure 4) is a voting game in which the number of choices depends on the target term and relation. The point of the game is to select the most (or least) relevant associations amongst those displayed. For example, what is the most relevant: "a bird can fly" or "a bird can sing"? In 4 months, more than $300,000$ relations have been tagged

Figure 4: Interface of `Selemo`. In this example, are the listed characteristics (*edible, hot, delicious, ...*) relevant or not to *casado* (a Costa Rican dish)?

(as relevant or not relevant) with this game. The accuracy of the results when 3 votes or more were cast is 100%, for 2 votes it is of 95% and 70% for just 1 vote.

Figure 5: `Askit` aims at assessing uncertain semantic relations, especially concerning polysemous words. In this example, can *archives* have the characteristic *pleine* (*full*)?

`AskIt` (see Figure 5) allows to validate/invalidate proposed relations inferred automatically from the `JeuxDeMots` lexical network. The `AskIt` engine selects a relation concerning a word meaning and ask if it holds for another meaning. For example: Does a *bank (river)* contain money? This strategy allows to build contrastive knowledge, which is instrumental in word sense disambiguation, especially when taking advantage of negative (i.e. inhibitory) relations. Since its launch in 2010, this game has allowed to validate/invalidate 1.5 million relations (corresponding to around 23 million votes) with an accuracy of 99.83%.
Similarly, `Emot` (see Figure 5) proposes a target term and a set of emotion/sentiment from which the player has to select the most appropriate (Lafourcade et al., 2016). Since its launch, more than 660,000 emotion/sentiment relations

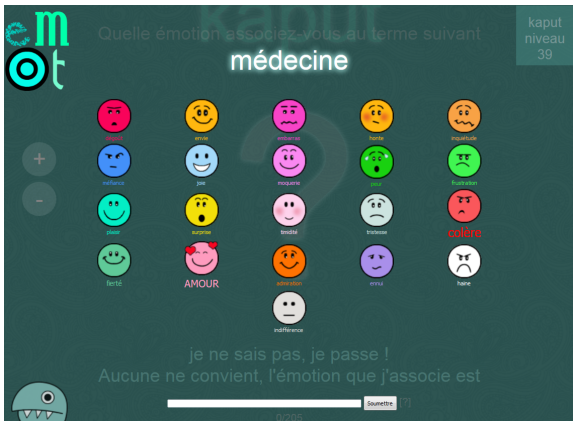

Figure 6: `Emot` aims at collecting sentiment associations with words. In this example, what are the sentiments that best correspond to *médecine* (*medicine*)?

have been created for 120,000 terms by 24 million votes.
`ColorIt` (Lafourcade et al., 2014) is based on the principle of `Emot` but adapted to color/appearance information. Since its start, more than 20,000 terms have been colorized with more than 3.7 million votes.
`PolitIt` (Tisserant and Lafourcade, 2015) is based on the same principle, but is adapted to political associations (for example, *market economy* with *liberalism*). Since its start, more than 8,000 terms have been *politized* with more than 500.000 million votes.
`Yakadirou` (see Figure 7) allows to associated a place preposition to a place relation. For example, in the relation *cat r_place sofa* what is the most relevant preposition: on, over, under? More than 380,000 bets have been placed in 2 years.

Figure 7: `Yakadirou` aims at associating prepositions to relations of place. In this example, what is the preposition of place to associate to *marchandise* and *postal parcel*?

`Tierxical` (see Figure 8) is a bit different from the previous games as it allows to bet on the first mostly associated terms for a given target term. The choice of the player slightly impacts the distribution of the relation weights. More than 750,000 bets have been placed in 5 years.

3. Limitations of the Approach

3.1. The Perils of Majority Voting

Although the influence of the other players' vote is limited, as the previous answers are only shown **after** the vote is

Figure 8: `Tierxical` aims at reordering word associations from the strongest to the weakest. In this example, what are the 3 best synonyms for *débiteur* (*debtor*)?

cast, majority voting still presents some important drawbacks.
First, the players are all considered equally, so a person who just plays around clicking randomly is considered the same as a highly skilled player.
Second, players can easily cheat if they agree on casting the same vote ("always click on Yes", for example).
These two limitations should be compensated by the number of players, provided enough of them play honestly.
Therefore, in such games, attracting a lot of players is especially important.

3.2. The Perils of Simplification

Another danger of voting GWAPs is that they can lead to over-simplification. One example of such a drift in a (microworking) crowdsourcing task is presented in (Bowman et al., 2015), in which in order to identify entailment relations, workers were asked if most people would say that if the first sentence is true, then the second must be too.
In our case, the voting tasks are complementary to a central game, `JeuxDeMots`, which allows to compensate, at least partly, for this effect.

4. Conclusion

Voting games are easy to develop and they provide a very efficient way of collecting large amounts of speakers' decisions in a very limited time. A common platform for such games would allow to easily gather language data, with very little development work.
In our case, the created resources are copyleft and can be downloaded directly from the games' Web sites, with a click on the upper left hand-side image.[4]

[4]For example, for `LikeIt`: `http://www.jeuxdemots.org/JDM-POLA-FR/?C=M;O=A`.

References

Bos, J. and Nissim, M. (2015). Uncovering noun-noun compound relations by gamification. In *Proc. of the Nordic Conference of Computational Linguistics (NODALIDA)*, pages 251–255, Vilnius, Lithuania, May.

Bowman, S. R., Angeli, G., Potts, C., and Manning, C. D. (2015). A large annotated corpus for learning natural language inference. *arXiv preprint arXiv:1508.05326*.

Chamberlain, J., Poesio, M., and Kruschwitz, U. (2008). Phrase Detectives: a web-based collaborative annotation game. In *Proc. of the International Conference on Semantic Systems (I-Semantics)*, Graz, Austria.

Gala, N. and Brun, C. (2012). Propagation de polarités dans des familles de mots : impact de la morphologie dans la construction dun lexique pour lanalyse dopinions. In *Proc. of Traitement Automatique des Langues Naturelles*, Grenoble, France, June.

Guillaume, B., Fort, K., and Lefebvre, N. (2016). Crowdsourcing complex language resources: Playing to annotate dependency syntax. In *Proc. of the International Conference on Computational Linguistics (COLING)*, Osaka, Japan, December.

Habernal, I., Hannemann, R., Pollak, C., Klamm, C., Pauli, P., and Gurevych, I. (2017). Argotario: Computational argumentation meets serious games. In *Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*.

Lafourcade, M. and Fort, K. (2014). Propa-l: a semantic filtering service from a lexical network created using games with a purpose. In *Proc. of the International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland, May.

Lafourcade, M. and Zampa, V. (2009). Pticlic : a game for vocabulary assessment combining jeuxdemots and lsa. In *Proc. of Conference on Intelligent text processing and Comptational Linguistics (CICLing)*, Mexico City, Mexico, March.

Lafourcade, M., Le Brun, N., and Zampa, V. (2014). Crowdsourcing word-color associations. In *Proc. of the International Conference on Application of Natural Language to Information Systems (NLDB)*, Montpellier, France, June.

Lafourcade, M., Le Brun, N., and Joubert, A. (2015). Collecting and evaluating lexical polarity with a game with a purpose. In *Proc. of the International Conference on Recent Advances in Natural Language Processing (RANLP)*, Hissar, Bulgaria, September.

Lafourcade, M., Le Brun, N., and Joubert, A. (2016). Construire un lexique de sentiments par crowdsourcing et propagation. In *Proc. of Traitement Automatique des Langues Naturelles*, Paris, France, July.

Lafourcade, M. (2007). Making people play for lexical acquisition. In *Proc. of the 7th Symposium on Natural Language Processing (SNLP 2007)*, Pattaya, Thailand, December.

Tisserant, G. and Lafourcade, M. (2015). Politit, du crowd-sourcing pour politiser le lexique. In *Proc. of Etudier le Web politique : Regards croiss, Institut des Sciences de l'Homme*, Lyon, France, June.

Name of the game	Created complementary resource	Information
LikeIt	polarized lexicon	150,000,000 votes - 740,000 terms - 1,700,000 polarities
AskIt	negative relations	25,000,000 votes - 860,000 negative relations
SexIt	sex/no sex relations	410,000 votes - 19,000 terms
PolitIt	political relations	540,000 votes - 8,900 politically tagged terms
ColorIt	color relations	3,700,000 votes - 20,000 colorized terms - 37,000 color relations
Selemot	annotations	23,000,000 votes - 1,500,000 annotations
Yakadirou	prepositions of place	380,000 votes - 27,000 place preposition annotated relations

Figure 9: Obtained results for the voting games.

von Ahn, L. and Dabbish, L. (2004). Labeling images with a computer game. In *Proc. of SIGCHI conference on Human factors in computing systems*, CHI '04, pages 319–326, New York, NY, USA. ACM.