



HAL
open science

Une plateforme de recommandation automatique d'emojis

Gaël Guibon, Magalie Ochs, Patrice Bellot

► **To cite this version:**

Gaël Guibon, Magalie Ochs, Patrice Bellot. Une plateforme de recommandation automatique d'emojis. Traitement Automatique du Langage Naturel, Jun 2017, Orléans, France. hal-01795576

HAL Id: hal-01795576

<https://amu.hal.science/hal-01795576>

Submitted on 18 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Une plateforme de recommandation automatique d’emojis

Gaël Guibon^{1, 2} Magalie Ochs¹ Patrice Bellot¹

(1) Aix Marseille Université, Université de Toulon, CNRS, ENSAM, LSIS, Marseille, France

(2) Caléa Solutions, 1 place Francis Chirat, 13002 Marseille, France
prenom.nom@lsis.org

RÉSUMÉ

Nous présentons une interface de recommandation d’emojis porteurs de sentiments qui utilise un modèle de prédiction appris sur des messages informels privés. Chacun étant associé à deux scores de polarité prédits. Cette interface permet également d’enregistrer les choix de l’utilisateur pour confirmer ou infirmer la recommandation.

ABSTRACT

An emoji recommendation platform

We show an emoji recommendation web interface dedicated to sentiment-related emojis. This application uses a model learnt on private informal short text messages associated with two predicted polarity scores. The application also saves the user’s choices to validate or invalidate the recommendation.

MOTS-CLÉS : emoji, recommandation, apprentissage automatique, medias sociaux, messagerie.

KEYWORDS: emoji, recommandation, machine learning, social media, messaging application.

1 Introduction

Les emojis sont l’un des principaux vecteurs d’émotions et de sentiments. Depuis leur création dans les années 90 et leur instauration dans le clavier de l’iPhone en 2011, les emojis sont de plus en plus présents dans le paysage actuel de la communication écrite sous presque toutes ses formes : SMS, message instantané, chat, forum, email, etc. Avec 2389 emojis “standards” fin 2016, et 2683 à ce jour¹, leur nombre ne cesse d’augmenter. Il convient donc de pouvoir recommander les emojis non plus uniquement à l’aide d’un lexique, comme c’est le cas dans Mood Messenger² et iMessage sous iOS 10 d’Apple, mais également de manière plus intelligente en prenant en compte la phrase entière. L’application proposée montre le résultat de travaux en cours sur la recommandation d’emojis porteurs de sentiments dans un contexte phrastique.

2 Architecture de l’application

Le système de recommandation de l’application propose une prédiction d’emojis porteurs de sentiments sélectionnés selon l’Emoji Sentiment Ranking (Kralj Novak *et al.*, 2015), un lexique d’emojis

1. Selon les emojis Unicode de mai 2017 : <http://unicode.org/emoji/charts/full-emoji-list.html>

2. <http://moodmessenger.com/>

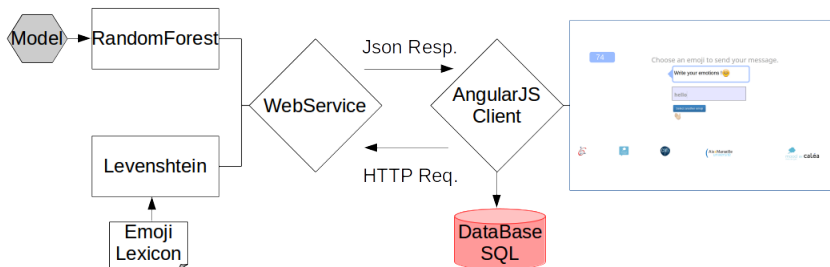


FIGURE 1 – Architecture globale de l'application

avec polarités associées issu de 83 annotateurs humains dans 13 langues. Cette recommandation se limite à la phrase en cours et non au message (Barbieri *et al.*, 2017) ni à l'ensemble de la conversation (Xie *et al.*, 2016). La séparation des phrases, le découpage en unités lexicales élémentaires et la lemmatisation sont effectués à l'aide de modèles de nltk³.

Le système de recommandation proposé combine deux approches (Figure 1) : une prédiction automatique d'emojis par apprentissage automatique et une distance d'édition.

Prédiction automatique. La prédiction automatique d'emojis est effectuée à l'aide de *Random Forest* (Breiman, 2001) pour classification multi-étiquettes appris sur 9700 phrases contenant des emojis. L'étendue de l'apprentissage est volontairement limitée à 169 emojis porteurs de sentiments. Lors de l'apprentissage et lors de la prédiction chaque phrase est traduite en une représentation vectorielle tf-idf qui est ensuite enrichie avec des n-grammes de 1 à 5 mots, et 2 scores de polarité prédits à l'aide d'Echo(Hamdan *et al.*, 2015) et de SentiStrength(Theilwall *et al.*, 2012). Le tout est ensuite utilisé comme caractéristiques pour le classifieur (Guibon *et al.*, 2017).

Distance d'édition. La distance d'édition utilisée est celle de Levenshtein (Levenshtein, 1966). Elle est associée à un lexique propriétaire de correspondance mots-clés - emojis. Plus précisément, l'insertion et la suppression ont une valeur de 1, tandis que la substitution a une valeur de 2 correspondant aux deux opérations de suppression puis d'insertion nécessaires pour les lettres déjà tapées.

Interface. L'interface web consiste en une SPA (*Single Page Application*) effectuée en AngularJS⁴ et dont le code source est disponible sur github⁵. Il s'agit d'un client qui consomme une API par requêtes HTTP pour récupérer du json. L'interface et le moteur de prédiction sont indépendants et communiquent via 2 ports. Enfin, à des fins de validation du système par l'utilisateur, l'interface est également liée à une base de données MySQL qui permet d'enregistrer les phrases tapées, les emojis sélectionnés et s'il s'agit d'un emoji recommandé ou d'un autre disponible parmi ceux représentant les 7 émotions basiques d'Ekman (Ekman, 1993).

L'application est hébergée sur le site du laboratoire à l'adresse suivante : <http://lsis-mood-emoji.lsis.org/>. Quant au système de recommandation, en voici un exemple de requête : <http://lsis-mood-emoji.lsis.org:8080/?query=hello>

3. <http://www.nltk.org/>

4. <https://angularjs.org/>

5. <https://github.com/gguibon/lsis-mood-emoji>

Références

- BARBIERI F., BALLESTEROS M. & SAGGION H. (2017). Are emojis predictable? *arXiv preprint arXiv :1702.07285*.
- BREIMAN L. (2001). Random forests. *Machine learning*, **45**(1), 5–32.
- EKMANN P. (1993). Facial expression and emotion. *American psychologist*, **48**(4), 384.
- GUIBON G., OCHS M. & BELLOT P. (2017). Prédiction automatique d'emojis sentimentaux. In *Conférence en Recherche d'Information et Applications (CORIA) 2017*.
- HAMDAN H., BELLOT P. & BECHET F. (2015). Sentiment lexicon-based features for sentiment analysis in short text. In *In Proceeding of the 16th International Conference on Intelligent Text Processing and Computational Linguistics*.
- KRALJ NOVAK P., SMAILOVIC J., SLUBAN B. & MOZETIC I. (2015). Sentiment of emojis. *PLOS ONE*, **10**(12).
- LEVENSHTAIN V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, p. 707–710.
- THELWALL M., BUCKLEY K. & PALTOGLOU G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, **63**(1), 163–173.
- XIE R., LIU Z., YAN R. & SUN M. (2016). Neural emoji recommendation in dialogue systems. *arXiv preprint arXiv :1612.04609*.