

Analyse de sentiments

Frédéric WANTIEZ – Pierre VIGIER

14 mai 2016

1 Description du problème

L'analyse de sentiments est un problème d'apprentissage supervisé. Notons W l'ensemble de tous les mots possibles et $W^* = \cup_{n \geq 0} W^n$ l'ensemble des textes sur ce vocabulaire. Soit $x_1, \dots, x_N \in W^*$ des textes et $y_1, \dots, y_N \in S$ le sentiment associé à chacun des textes. Ces sentiments peuvent être des valeurs dans $S = [0, 1]$ où 0 signifie que le texte est "très négatif" et 1, "très positif". Une variante plus simple est d'avoir les $(y_i)_{i \in 1, \dots, N}$ dans $S = 0, 1$ où 0 signifie "négatif" et 1, "positif". Notre objectif est de déterminer une fonction f telle que $\forall i \in 1, \dots, N, y_i \approx f(x_i)$ et qui devra, de plus, bien généraliser sur des textes jamais vus auparavant. Dans le cas où S est fini, il s'agit d'un problème de classification. Dans le cas contraire, il s'agit d'un problème de régression.

Nous allons essentiellement nous concentrer sur le problème de classification. Plusieurs types d'entrée et plusieurs types de classifieurs seront essayés sur le problème. La mesure de performance choisie est la précision $A(y_1, \dots, y_N, \hat{y}_1, \dots, \hat{y}_N) = \frac{\sum_{i=1}^N 1_{y_i = \hat{y}_i}}{N}$. Elle nous permettra de comparer les performances des différents algorithmes.

{Technologie qu'on utilise ?

2 Ensemble de données

3 Méthodologie

4 Premiers essais

5 Prise en compte de l'ordre des mots

6 Conclusion