

Cours de Traitement Automatique du Langage

Chapitre 06 : Analyse et classification de textes
DI5 – 2021-22

Analyse de sentiments

- Critique de films positive ou négative ?
 - + *...many characters and richly applied satire, and some great plot twists*
 - *It was pathetic. The worst part about it was the boxing scenes...*
 - + *...awesome caramel sauce and sweet toasty almonds. I love this place!*
 - *...awful pizza and ridiculously overpriced...*

Analyse de sentiments

- Critique de films positive ou négative ?
 - + ...zany characters and **richly** applied satire, and some **great** plot twists
 - It was **pathetic**. The **worst** part about it was the boxing scenes...
 - + ...**awesome** caramel sauce and sweet toasty almonds. I **love** this place!
 - ...**awful** pizza and **ridiculously** overpriced...

Text Classification

Définition

- *Input*: a document d + a fixed set of classes $C = \{c_1, c_2, \dots, c_J\}$
- *Output*: a predicted class $c \in C$

A partir de règles

- Règles basées sur des combinaisons de mots ou d'autres caractéristiques
- La précision peut être élevée si les règles sont soigneusement affinées par un expert mais l'élaboration et la maintenance de ces règles sont coûteuses
- Spam : black-list-address OR ("dollars" AND "you have been selected")

Supervised Machine Learning

- *Input*: A training set of m hand-labeled documents $(d_1, c_1), \dots, (d_m, c_m)$
- *Output*: a learned classifier $\gamma: d \rightarrow c$
- Different classifieurs possible : Bayes, kppv, NN, ...

Classifieur naïf de Bayes

- Méthode ("naïve") basée sur la règle de Bayes

- S'appuie sur une représentation très simple du document → Bag of Words
- Première tentative : estimations par maximum de vraisemblance
 - utiliser simplement les fréquences d'apparition des mots dans les données → creation d'un mega-document par concaténation
 - Apprentissage

$$\hat{P}(c_j) = \frac{N_{c_j}}{N_{total}} \quad \hat{P}(w_i | c_j) = \frac{\text{count}(w_i, c_j)}{\sum_{w \in V} \text{count}(w, c_j)}$$

- Classification

$$c_{MAP} = \operatorname{argmax}_c \hat{P}(c) \prod_i \hat{P}(x_i | c)$$

- Probleme des mots absents du corpus $p = 0$



Prenons un exemple

1. Prior from training:

$$\hat{P}(c_j) = \frac{N_{c_j}}{N_{total}}$$

$$P(-) = 3/5$$

$$P(+) = 2/5$$

2. Drop "with"

3. Likelihoods from training (avec lissage de Laplace):

$$p(w_i|c) = \frac{\text{count}(w_i, c) + 1}{(\sum_{w \in V} \text{count}(w, c)) + |V|}$$

$$P(\text{"predictable"}|-) = \frac{1+1}{14+20} \quad P(\text{"predictable"}|+) = \frac{0+1}{9+20}$$

$$P(\text{"no"}|-) = \frac{1+1}{14+20} \quad P(\text{"no"}|+) = \frac{0+1}{9+20}$$

$$P(\text{"fun"}|-) = \frac{0+1}{14+20} \quad P(\text{"fun"}|+) = \frac{1+1}{9+20}$$

	Cat	Documents
Training	-	just plain boring
	-	entirely predictable and lacks energy
	-	no surprises and very few laughs
	+	very powerful
	+	the most fun film of the summer
Test	?	predictable with no fun

4. Scoring the test set:

$$P(-)P(S|-) = \frac{3}{5} \times \frac{2 \times 2 \times 1}{34^3} = 6.1 \times 10^{-5}$$

$$P(+)P(S|+) = \frac{2}{5} \times \frac{1 \times 1 \times 2}{29^3} = 3.2 \times 10^{-5}$$

Améliorations

- Binary Multinomial Naïve Bayes
 - utiliser l'occurrence des mots plutôt que leur fréquence → suppression des mots dupliquer avant l'énumération
- Utilisation de lexiques (des mots positifs ou négatifs)
 - MPQA Subjectivity Cues Lexicon
- Prise en compte des négations
 - Simple baseline method: Ajouter NOT_ devant les mots dans la négation
didn't like this movie , but I → didn't NOT_like NOT_this NOT_movie but I
- Machine Learning → Apprentissage
- Etc...

Impossible de conclure ce cours sans....

Quelques mots sur les
Chatbots et les systèmes
de dialogue

Agents de dialogue basés sur des Frames

- Parfois appelés « Chatbots basés Tâches ».
- Systèmes dont l'objectif est d'aider un utilisateur à résoudre une tâche
 - Exemple : la réservation d'un voyage ou l'achat d'un produit.
- **Idée principale**
 - Une structure de connaissances (Frame → cadre, canevas) représentant les intentions de l'utilisateur
 - Un ou plusieurs Frames (chacun composé d'emplacements avec des valeurs)
- **2 architectures principales**
 - **GUS (Genial Understander System) architecture (1977)**
 - Très vieux mais toujours très utilisé !
 - Bobrow, Daniel G., Ronald M. Kaplan, Martin Kay, Donald A. Norman, Henry Thompson, and Terry Winograd. 1977. "GUS, a frame-driven dialog system." *Artificial Intelligence* 8, 2:155-173.
 - **Dialogue-State architecture**
 - Extension de GUS

The Frames...

Frame

- Un ensemble de **slots**, à remplir avec des informations d'un type donné.
- Chacun est associé à une question posée à l'utilisateur
- Le système pose des questions à l'utilisateur jusqu'à remplir tous les slots
- L'utilisateur peut remplir plusieurs slots à la fois :
- Lorsque le Frame est complètement rempli, la base de données est interrogée pour **produire une réponse**

Slot	Type	Question
ORIGIN city		"What city are you leaving from?"
DEST	city	"Where are you going?"
DEP DATE	date	"What day would you like to leave?"
DEP TIME t	ime	"What time would you like to leave?"
AIRLINE	line	"What is your preferred airline?"

Dialogue → Je veux un vol de San Francisco à Denver, aller simple, partant après 17 heures le mardi.

Mutli-Frames

- Système additionnel de "**Frame detection**"
 - Les Chatbots sont souvent constituer de plusieurs Frames

Remplissage des slots

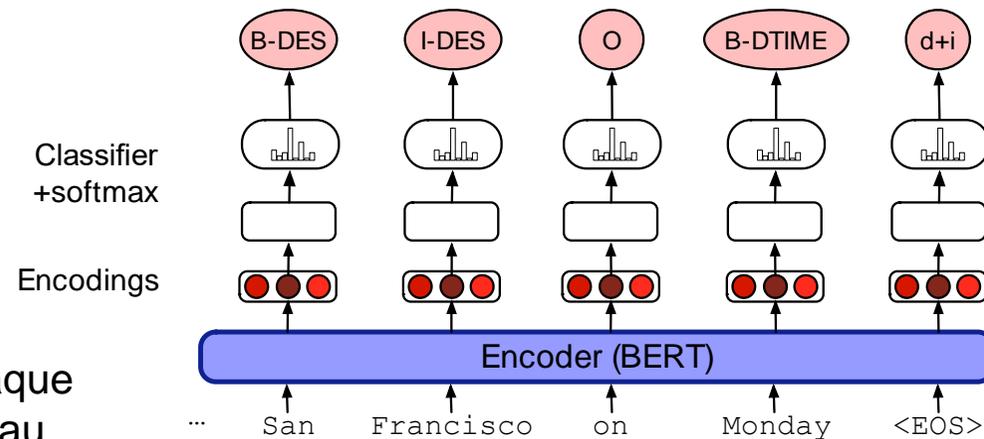
Comment sont remplis les slots?

1/ Analyse à base de règles

- CF Chapitre Traitement de base → Tokenization + Expressions régulières

2/ ML → classifieurs permettant d'associer des mots aux slots sémantique

- Étant donné un ensemble de phrases étiquetées
- Entrée: «I want a flight for San-Francisco on Monday afternoon please».
- Sortie: Destination : SF - Date de départ : Lundi - ...
- Nécessité de beaucoup de données étiquetées
- Marquage BIO (Begin, intérieur, Outside)
- L'idée : Entraîner un classificateur à étiqueter chaque mot en entrée avec une étiquette qui correspond au slot correspondant (s'il y en a un)
- Apprentissage : Ajout d'étiquettes B et I pour chaque type de slot dans les données d'entraînement



0 0 0 0 0 B-DES I-DES 0 B-DEPTIME I-DEPTIME 0
I want to fly to San Francisco on Monday afternoon please

Et ensuite....

0 0 0 0 0 B-DES I-DES 0 B-DEPTIME I-DEPTIME 0
I want to fly to San Francisco on Monday afternoon please

Post traitements

- Extraction de la chaîne de remplissage pour chaque slot.
- Normalisation à la forme correcte avec une ontologie.
 - Utilisation de dictionnaires d'homonymes (SF=SFO=San Francisco)
 - A template is a pre-built response string

Génération des réponses → template-based generation

- Templates fixes → "Hello, how can I help you?"
- Templates avec variables :
 - "What time do you want to leave CITY-ORIG?"
 - "Will you return to CITY-ORIG from CITY-DEST?"

Un dernier mot...

Attention aux biais en TAL et ML

De nombreuses fois constatés en TAL

- Kiritchenko et Mohammad (2018) ont constaté que la plupart des classificateurs de sentiments attribuent une émotion plus négative aux phrases contenant des noms afro-américains
- Certains classificateurs de toxicité (discours de haine, abus, harcèlement ou autres) signalent à tort comme toxiques des phrases qui ne le sont pas mais qui mentionnent simplement des termes spécifiques comme « aveugle » ou « homosexuel »
- ...
- **Peut être causé par :**
- Des problèmes dans les données d'entraînement ; les systèmes d'apprentissage automatique sont connus pour amplifier les biais.
- Des problèmes dans les étiquetages humains
- Des problèmes dans les ressources utilisées (comme les lexiques)
- Problèmes dans l'architecture du modèle
- **Un domaine de recherche encore ouvert**

Model Cards (Mitchell et al., 2019)

Pour chaque algorithme que vous diffusez, documentez à minima :

- l'utilisation et les utilisateurs prévus
- les algorithmes et paramètres d'entraînement utilisés
- les sources de données utilisées
- les prétraitements effectués sur les données et leur motivation
- Les protocoles d'évaluation suivis
- Les performances du modèle testé sur dans différentes configurations (groupes démographiques ou autres)