UNIVERSIDAD DIEGO PORTALES

Tarea 6: Minería de Datos y Procesamiento de Lenguaje Natural

Rodrigo Fuenzalida

Profesor: Alejandro Figueroa
Ayudante: Nicolás Olivares

Facultad de Ingeniería
Escuela de Informática y telecomunicaciones

28 de octubre de 2013

Índice

Capítulo 1

Introducción

El objetivo de esta tarea es comparar dos clasificadores multi-clases: SVM y MaxEnt. La idea es realizar lo mismo de la tarea dos, pero esta vez utilizando modelos de maxima entropia. Muestre una tabla comparativa.
Notese que debe mantener la consistencia de los folders de datos, es decir, utilizar los mismos folders de entrenamiento y prediccion para ambos clasificadores.

Capítulo 2

Desarrollo

2.1. SVM

Las máquinas de soporte vectorial o máquinas de vectores de soporte (Support Vector Machines, SVMs) son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik y su equipo en los laboratorios ATT.

Estos métodos están propiamente relacionados con problemas de clasificación y regresión. Dado un conjunto de ejemplos de entrenamiento (de muestras) podemos etiquetar las clases y entrenar una SVM para construir un modelo que prediga la clase de una nueva muestra. Intuitivamente, una SVM es un modelo que representa a los puntos de muestra en el espacio, separando las clases por un espacio lo más amplio posible. Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de su proximidad pueden ser clasificadas a una u otra clase.

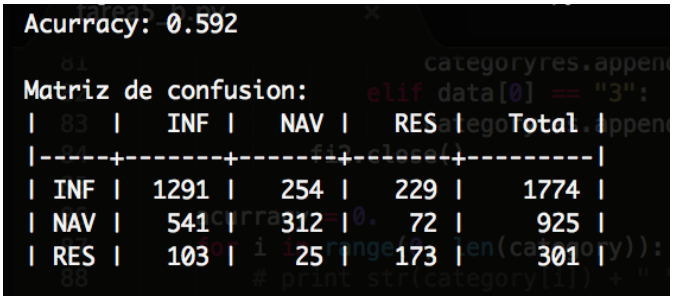

Figura 2.1: SVM

En la figura 2.1 se presentan los valores obtenido por SVM, en el cual los datos de prueba fueron 300 querys y los datos a testear fueron las 2700 etiquetas restantes, lo cual arrojo los valores que se presentan en dicha figura.

2.2. MaxEnt

En la figura 2.2 se aprecian los resultados obtenido por una implementación de MaxEnt realizada en python y considerando las mismas propiedades propuestan en el caso de SVM, como se puede ver, los resultados son muy distintos unos de otros, por lo que podemos determinar la calidad que entrega una técnica de la otra.

```
Acurracy: 0.100333333333

Matriz de confusion:
|     |   INF |   NAV |   RES |   Total |
|-----+-------+-------+-------+---------|
| INF |     0 |     0 |  1774 |    1774 |
| NAV |     0 |     0 |   925 |     925 |
| RES |     0 |     0 |   301 |     301 |
```

Figura 2.2: MaxEnt