

Capítulo 4

Resultados

4.1. Resultados generales

A continuación se presentan y analizan los resultados obtenidos por el clasificador creado.

En la tabla 4.1, se presentan los resultados promediados de la validación cruzada con 10 pliegues para cada uno de los cinco experimentos efectuados en este trabajo. En esa tabla se observa que todos los experimentos superaron el baseline de la simple predicción aleatoria de clases. Esta predicción implica una exactitud de 50 %. Como se ve en el primer renglón de la tabla, el mejor resultado de exactitud se obtuvo usando los enunciados con adjetivos o adverbios y el resultado más pobre se obtuvo usando los enunciados con disparadores de presuposición. Como era de esperarse, el error más alto se registró usando este último método, mientras que el error más bajo fue para el método usando los enunciados con adjetivos o adverbios. También se puede notar que los resultados de razón-tp y razón-fp son coherentes con los resultados de precisión y error obtenidos, es decir, los mejores resultados se aprecian en el primer renglón y los peores en el quinto (enunciados con adjetivos o adverbios y enunciados con disparadores de presuposición, respectivamente). Con respecto a la última columna, número de rasgos, se podría inferir que existe un umbral en el número de rasgos que afecta directamente el desempeño del clasificador, ya que el experimento con dispara-

4.1 Resultados generales

dores de presuposición fue el que tuvo el menor número de rasgos y es posible que esta haya sido la causa del bajo desempeño del clasificador. Sin embargo, el resto de los experimentos (renglones 1 a 4) tuvieron un número de rasgos similar (alrededor de 20,000) y aún así sus desempeños cambiaron considerablemente, es decir, a pesar de tener un número similar de rasgos, los experimentos de los renglones 2, 3 y 4 no alcanzaron la exactitud del experimento del primer renglón.

	Método	Exactitud	Error	Razón-tp	Razón-fp	Número de rasgos
(1)	enunciados con adjetivos o adverbios	72.60	27.40	72.65	27.45	20,016.70
(2)	enunciados agrupados con FNM	69.30	30.70	69.08	30.48	19,332.20
(3)	enunciados con disparadores o adjetivos o adverbios	67.56	32.44	66.45	31.33	20,434.40
(4)	todos los enunciados	66.63	33.37	65.87	32.60	19,101.80
(5)	enunciados con disparadores de presuposición	63.84	36.16	64.50	36.83	9,277.80

Tabla 4.1: Resultados promediados de la validación con 10 pliegues del clasificador creado.

En la figura 4.1 se observan cinco puntos graficados en el espacio ROC, cada uno de ellos correspondiente a los cinco métodos usados para entrenar el clasificador. El punto correspondiente al uso de oraciones con adjetivos o adverbios es el que se encuentra más cercano a la clasificación perfecta (0.0,1.0), lo que indica que obtuvo el menor error falso positivo y clasificó mejor que los demás los casos positivos. En la figura 4.2 se presentan las curvas ROC, una para cada método utilizado. En ellas se confirman los resultados presentados anteriormente: el mejor resultado se obtiene entrenando el clasificador con enunciados que contenga adjetivos o adverbios. En ellas también se puede apreciar el trade-off que se obtendría si se deseara clasificar con mayor efectividad los casos realmente positivos. Por ejemplo, en el caso de los enunciados con adjetivos o adverbios, si se deseara una razón-tp de 90 %, se tendría que aceptar una razón-fp de poco más del 50 %.

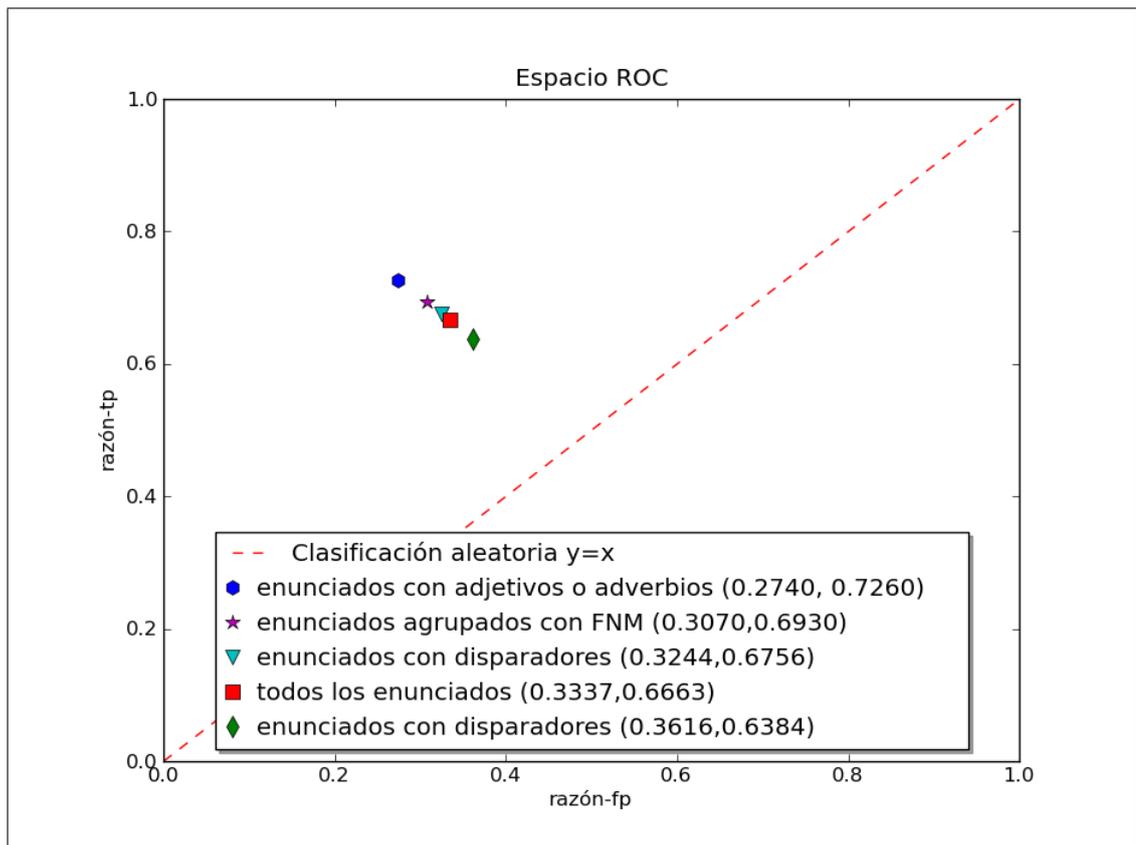


Figura 4.1: Ubicación de los clasificadores en el espacio ROC - La figura muestra los puntos que representan a cada una de las técnicas usadas para entrenar el clasificador. Cuando se usan enunciados con adjetivos o adverbios se logra la mejor posición.

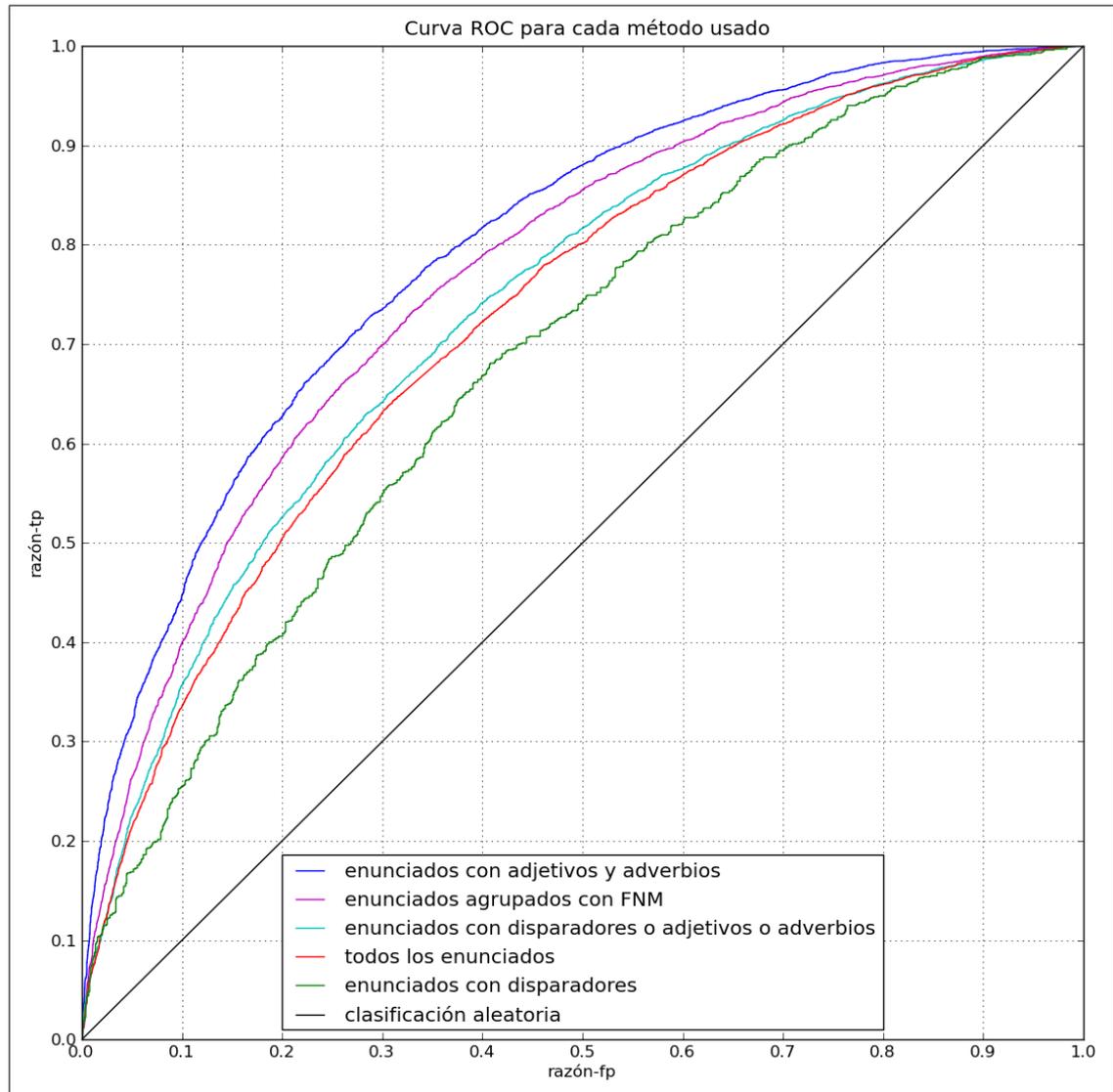


Figura 4.2: Curva ROC para cada método utilizado - La figura muestra la curva ROC obtenida por cada método usado para entrenar cada clasificador. Las curvas son congruentes con los otros resultados presentados: cuando se usan enunciados con adjetivos o adverbios se logra el mejor desempeño.

4.1 Resultados generales

		Método	Exactitud	Error	Número de rasgos
Resultados con los datos de [47]	(1)	unigramas	79.63	20.36	23,486
	(2)	adjetivos	76.37	23.62	4,434
Resultados de [47]	(3)	unigramas	78.70	21.30	16,165
	(4)	adjetivos	77.00	23.00	2,633

Tabla 4.2: Resultados promediados de la validación cruzada con 3 pliegues del clasificador creado para cada experimento. Se observa que los valores de exactitud para cada experimento (unigramas y adjetivos) son muy similares para el clasificador creado en esta tesis y el creado en [47].

A pesar de que no son comparables los resultados obtenidos en este trabajo con los obtenidos en [47], debido a la diferente metodología seguida, se decidió realizar dos experimentos con el corpus de comentarios usado en [47] para conocer el desempeño del clasificador bayesiano programado en este trabajo. Los resultados obtenidos en estos experimentos se encuentran en la tabla 4.2.

La prueba consistió en reproducir dos de los experimentos reportados en ese trabajo. El primer experimento en cuestión es el que hace uso exclusivamente de los unigramas más frecuentes, así como el uso de palabras negadas¹. Durante la reproducción del experimento no se hizo uso de las palabras negadas, porque se consideró innecesario; esto es, se utilizaron únicamente los unigramas más frecuentes.

Como se ve en el tercer renglón de la tabla 4.2, en ese trabajo se reportó una exactitud de 78.70 % para el primer experimento. Como se observa en el primer renglón, el resultado obtenido con el clasificador de este trabajo fue de 79.63 %.

En el segundo y cuarto renglón están los resultados para la segunda prueba (adjetivos más frecuentes). En esta prueba se obtuvo una exactitud de 76.37 % mientras que en el trabajo de comparación se obtuvo una exactitud de 77 %. Como se observa, la exactitud disminuye tal como en el trabajo citado.

Es importante notar que, como en [47], en ambos casos (con unigramas y con adjetivos) se usaron solo aquellos cuya frecuencia absoluta era mayor o igual a cuatro.

¹El uso de palabras negadas consiste en concatenar la etiqueta “NOT_” a todas las palabras que siguieron, en una oración, a las negaciones *not*, *isn't*, *didn't*, etc.)

Con esta prueba se pudo concluir que los experimentos de este trabajo y los de [47] tienen una exactitud similar, aun sin considerar pasos especiales como el uso de palabras negadas. También se observa que a pesar de que la exactitud es menor usando solo adjetivos, la cantidad de rasgos utilizados también es mucho menor a la cantidad usada con los unigramas. Esto podría significar la importancia de los adjetivos y adverbios en la emisión de un juicio.

4.2. Enunciados con adjetivos o adverbios

Respecto a los experimentos llevados a cabo en este trabajo, los resultados pueden apreciarse en la tabla 4.1. A continuación se analizará cada uno de los experimentos realizados.

Las oraciones subjetivas, detectadas por medio de aparición de adjetivos o adverbios, resultó el método más exitoso, alcanzando 72.6 % de exactitud. Sus valores de razón-tp y razón-fp son considerablemente mejores que los de los métodos restantes. Esto resulta contradictorio con los resultados de [47], donde los mejores resultados son obtenidos con los unigramas únicamente (sin distinguir si son o no adjetivos u otra parte del discurso) y confirma que los adjetivos y adverbios tienen información que determina la opinión sobre algo, como se muestra también en [2].

Aunque cabe mencionar que además de los adjetivos y adverbios, se incluyeron todas las palabras de esas oraciones donde ocurrieron, intentando así incluir palabras que formaran parte del contexto de la opinión.

4.3. Enunciados agrupados con FNM

Este método resultó el segundo mejor con una exactitud de 69.3 %. Cabe recordar que este método usó los adjetivos y adverbios, los unigramas y los trigramas (sin sustantivo en la segunda palabra del trigramas) más frecuentes.

Si no se seleccionan los rasgos de las matrices de datos, los enunciados se agrupan generalmente, según lo experimentado, no por sentimiento sino por tema,

4.4 Enunciados con disparadores de presuposición o adjetivos o adverbios

en este caso por película, lo cual no ayuda para determinar la orientación de la opinión.

4.4. Enunciados con disparadores de presuposición o adjetivos o adverbios

Se obtuvo una exactitud de 67.56 % con estos enunciados, elegidos solo si contenían algún disparador considerado o algún adjetivo o adverbio. Su desempeño fue mejor que si se tomaran solo los disparadores pero inferior a tomar solo los adjetivos o adverbios. Esto podría indicar que los disparadores no ofrecen información acerca de la opinión de una declaración.

4.5. Todos los enunciados

Este método sirvió como control, ya que se tomaron todas las palabras de todas las oraciones seleccionadas (aleatoriamente) de los datos de entrenamiento. Todos los métodos usados, excepto el que usa solo enunciados con disparadores de presuposición, superaron la exactitud de 66.63 % obtenida con este enfoque.

4.6. Enunciados con disparadores de presuposición

Con la menor exactitud reportada, de 63.84 %, este método entregó los resultados más pobres de toda la experimentación. Esto puede implicar que los disparadores no son más útiles para la detección de opiniones que simplemente usar las palabras de todos los enunciados, sin embargo usando una lista más amplia de disparadores o buscando otro tipo de características relacionadas a ellos, se podrían obtener mejores resultados.

4.6 Enunciados con disparadores de presuposición

Este bajo desempeño se podría deber a que el número de enunciados con disparadores es muy reducido, y los clasificadores requieren de una cantidad considerable de ejemplos (conocimiento previo) para poder predecir exitosamente nuevos casos.