
Índice general

Agradecimientos	II
Resumen	III
Resum	V
Abstract	VII
1. Introducción	1
1.1. Opiniones falsas	3
1.2. Detección de las opiniones falsas	5
1.3. Objetivos	7
1.4. Antecedentes	8
1.4.1. Clasificadores de una sola clase	9
1.4.2. El método de PU-Learning	9
1.5. El método de PU-Learning*	11
1.6. Metodología	13
1.6.1. Método PU-Learning	14
1.6.2. Variante del método PU-Learning	15
1.6.3. Método PU-Learning*	17
1.6.4. Detección de opiniones falsas en dominios cruzados	18
1.7. Organización de la tesis	20
2. Using PU-Learning to detect deceptive opinion spam	22
2.1. Introduction	23

2.2. Related Work	25
2.3. PU-Learning for opinion spam detection	26
2.4. Evaluation	28
2.4.1. Datasets	28
2.4.2. Evaluation Measure	29
2.4.3. Results	30
2.5. Conclusions and future work	33
3. Detecting positive and negative deceptive opinions using PU-Learning	35
3.1. Introduction	36
3.2. Related Work	38
3.3. PU-Learning for Opinion Spam Detection	41
3.4. Datasets	43
3.5. Experimental Evaluation	46
3.5.1. Experimental settings	46
3.5.2. Experiment 1: Lower and upper bounds for the PU-learning approach	47
3.5.3. Experiment 2: Original vs modified PU-learning	49
3.5.4. Experiment 3: Polarity and deception under PU-learning	52
3.5.5. Experiment 4: On the choice of features and classifier	53
3.6. Conclusions and Future Work	53
4. Detection of opinion spam with character n-grams	56
4.1. Introduction	57
4.2. Related Work	59
4.3. Experimental setup	60
4.4. Experiments	61
4.4.1. Experiment 1: Character vs. word n-grams	61
4.4.2. Experiment 2: Character n-grams robustness	63
4.5. Conclusions and future work	66
5. Discusión de los resultados	68
5.1. Evaluación experimental	69

5.2. Configuración experimental	69
5.2.1. Colecciones	71
5.2.2. Pre-procesamiento	75
5.2.3. Representaciones	76
5.2.4. Medidas de evaluación	77
5.2.5. Análisis estadístico	78
5.3. PU-Learning* con n-gramas de palabras	79
5.3.1. Opiniones favorables	80
5.3.2. Opiniones desfavorables	81
5.3.3. Opiniones falsas de ambas polaridades: favorables y desfavorables	82
5.3.4. Análisis de significancia estadística	84
5.4. PU-Learning* con n-gramas de caracteres	85
5.5. Comparación con otros métodos	87
5.6. PU-Learning* en dominios cruzados	90
6. Conclusiones	95
6.1. Conclusiones	96
6.2. Trabajo futuro	98
6.3. Publicaciones	99
Bibliografía	100

Índice de figuras

1.1. Construcción del clasificador con el método PU-Learning*	14
2.1. Classifier construction with PU-Learning approach.	27
2.2. Summary of best Results; f-measure.	33
3.1. Baseline and Upperbound results for the different subsets of positive and negative opinions.	47
3.2. Results of the baseline, original PU-learning, and modified PU-learning in the classification of deceptive and truthful opinions from both po- larities.	49
4.1. The macro f-measure variation with the training set size.	65
5.1. f-measure para diferentes tamaños del conjunto de entrenamiento. . .	85
5.2. f-measure para diferentes tipos de atributos.	86

Lista de algoritmos

1.	El método de PU-Learning	11
2.	El método de PU-Learning*	13
3.	PU-Learning for opinion spam detection	28
4.	Original PU-learning algorithm. P and U are the sets of positive and unlabeled examples respectively; C_i is the binary classifier at iteration i ; Q_i represents the set of unlabeled examples from U_i classified as negative by C_i , and RN_i is the set of reliable negative examples gathered from iteration 1 to iteration i	42
5.	Modified PU-learning algorithm. P and U are the sets of positive and unlabeled examples respectively; Q_i and RN_i represent the sets of identified and retained reliable negative examples at iteration i , and C_i is the binary classifier at iteration i	43

Índice de tablas

1.1. Resultados de la aplicación del método de PU-Learning.	15
1.2. Resultados de la aplicación del método de PU-Learning*.	16
1.3. Resultados de la aplicación del método de PU-Learning* con n-gramas de caracteres.	18
1.4. Resultados de la aplicación del método de PU-Learning* en dominios cruzados con hoteles y restaurantes.	19
1.5. Resultados de la aplicación del método de PU-Learning* en dominios cruzados con hoteles y médicos.	19
2.1. Comparison of the performance of different classifiers when using 20, 40 and 60 examples of deceptive opinions for training; in this table D refers to deceptive opinions and U to unlabeled opinions.	31
2.2. Comparison of the performance of different classifiers when using 80, 100 and 120 examples of deceptive opinions for training; in this table D refers to deceptive opinions and U to unlabeled opinions.	32
3.1. Detailed results on the classification of <i>positive</i> opinions using 60, 80, 100 and 120 labeled deceptive opinions (DP) and 520 of unlabeled examples (UN) for training. In this table, P, R and F state for preci- sion, recall and f-measure respectively.	50
3.2. Detailed results on the classification of <i>negative</i> opinions using 60, 80, 100 and 120 labeled deceptive opinions (DP) and 520 of unlabe- led examples (UN) for training. In this table, P, R and F state for precision, recall and f-measure respectively.	51

3.3. Detecting Deceptive opinions when using 120, 160, 200 and 240 samples of Deceptive opinions and 1040 opinions of mixed polarities in the Unlabeled set (520 Deceptive and 520 Truthful).	52
3.4. Results of the classification of positive and negative opinion spam by Naïve Bayes (NB) and SVM using unigrams and bigramas as features. The values correspond to the F_1 measure for both classes, deceptive and truthful opinions.	54
4.1. Results obtained with word ngrams and character n-grams for positive opinion spam.	61
4.2. Results obtained with word ngrams and character n-grams for negative opinion spam.	62
4.3. Results obtained with word ngrams and character n-grams for the full set of opinion spam.	62
4.4. The 20 character n-grams with highest Information Gain values for postivie and negative opinions.	64
5.1. Sub-corpus de opiniones favorables.	72
5.2. Sub-corpus de opiniones desfavorables.	73
5.3. Sub-corpus de opiniones falsas de ambas polaridades.	74
5.4. Corpus de opiniones utilizados en los experimentos de dominios cruzados.	74
5.5. Bolsas de n-gramas de palabras.	76
5.6. Bolsas de n-gramas de caracteres.	77
5.7. Resultados del experimento con opiniones favorables.	80
5.8. Resultados del experimento con opiniones desfavorables.	82
5.9. Resultados del experimento de opiniones falsas de ambas polaridades.	83
5.10. Comparación de los resultados obtenidos por el método de PU-Learning* contra otros métodos que emplean el mismo corpus de opiniones favorables.	89
5.11. Comparación de los resultados obtenidos por el método de PU-Learning* con los métodos que emplean el mismo corpus de opiniones desfavorables.	90

5.12. Sub-corpus utilizados para el conjunto de entrenamiento de los experimentos de dominios cruzados.	91
5.13. Resultados del experimento de dominios cruzados utilizando como conjunto de entrenamiento hoteles y como conjunto de prueba restaurantes.	92
5.14. Resultados del experimento de dominios cruzados utilizando como conjunto de entrenamiento hoteles y como conjunto de prueba médicos.	92
5.15. Índice Jaccard para los dominios de hoteles-restaurantes.	93
5.16. Índice Jaccard para los dominios de hoteles-médicos.	93