



An Introduction to Text Classification

Exercises

Jörg Steffen, DFKI

Substitute Günter Neumann, DFKI

steffen@dfki.de

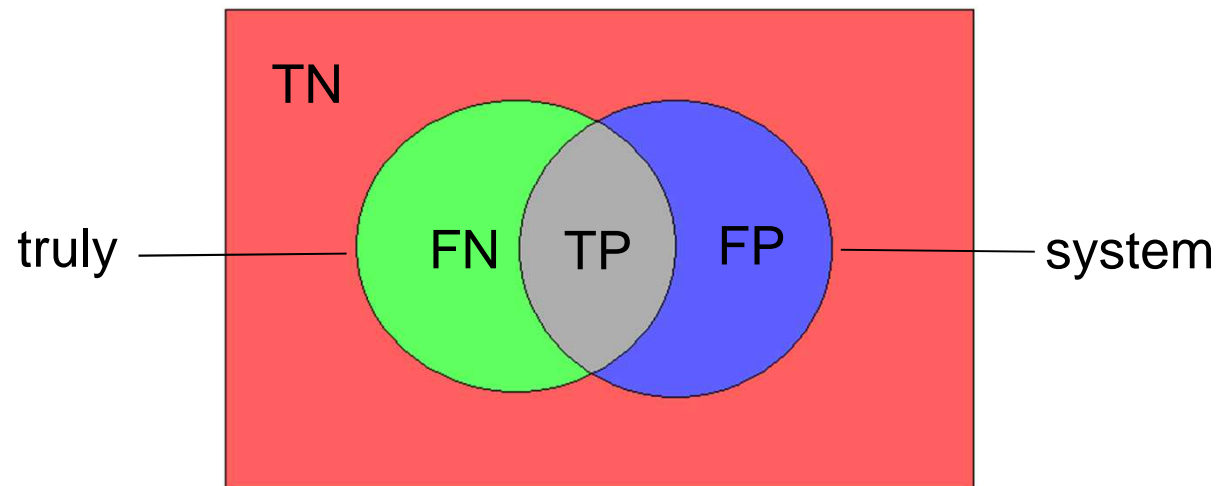
10.11.2014

Classification Evaluation



- Possible results of a binary classification in a **confusion matrix**

	truly YES	truly NO
system YES	true positives	false positives
system NO	false negatives	true negatives





- Precision

- percentage of documents correctly identified as belonging to the category

$$\textit{precision} = \frac{\textit{true positives}}{\textit{true positives} + \textit{false positives}}$$

- Recall

- percentage of documents found belonging to the category

$$\textit{recall} = \frac{\textit{true positives}}{\textit{true positives} + \textit{false negatives}}$$



- F-Measure combines both precision and recall in one value

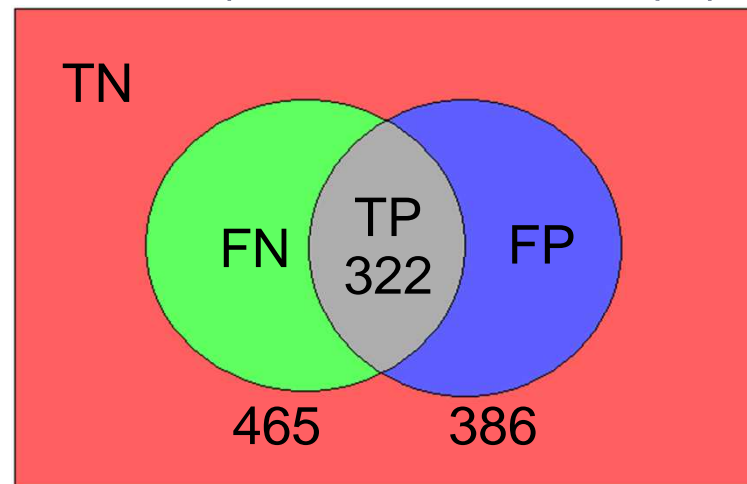
$$F_{\beta} = \frac{(\beta^2 + 1) \times \textit{precision} \times \textit{recall}}{\beta^2 \times \textit{precision} + \textit{recall}}$$

- β allows different weighting of precision and recall
- for equal weighting: $\beta = 1$
- Precision twice as important as Recall: $\beta = 0.5$
- Recall twice as important as Precision: $\beta = 2$

Exercise 1



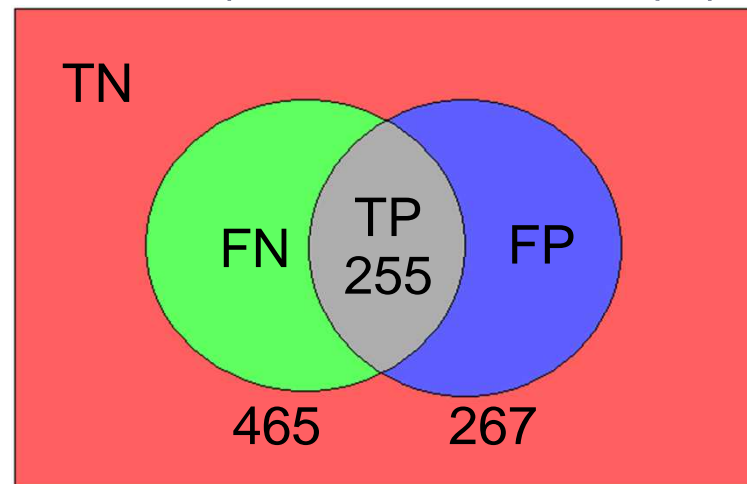
- True Positives: 322
- False Positives: $386 - 322 = 64$
- False Negatives: $465 - 322 = 143$
- Precision: $322 / (322 + 64) = 0.83$
- Recall: $322 / (322 + 143) = 0.69$
- F1-Measure: $(2 \times 0.83 \times 0.69) / (0.83 + 0.69) = 0.75$



Exercise 2



- True Positives: 255
- False Positives: $267 - 255 = 12$
- False Negatives: $465 - 255 = 210$
- Precision: $255 / (255 + 12) = 0.96$
- Recall: $255 / (255 + 210) = 0.55$
- F1-Measure: $(2 \times 0.96 \times 0.55) / (0.96 + 0.55) = 0.70$



Exercise 3



- F1-Measure Ex.1:
$$((1 + 1) \times 0.83 \times 0.69) / (1 \times 0.83 + 0.69) = 0.75$$
- F0.5-Measure Ex.1:
$$((0.25 + 1) \times 0.83 \times 0.69) / (0.25 \times 0.83 + 0.69) = 0.80$$
- F1-Measure Ex.2:
$$((1 + 1) \times 0.96 \times 0.55) / (1 \times 0.96 + 0.55) = 0.70$$
- F0.5-Measure Ex.2:
$$((0.25 + 1) \times 0.96 \times 0.55) / (0.25 \times 0.96 + 0.55) = 0.83$$