Non-Patent Document 2: "Improving text classification by shrinkage in a hierarchy of classes", A. McCallum et al., 1998.

## DISCLOSURE OF INVENTION

### Problems to be Solved by the Invention

All previous methods require the use of held-out data 2 to estimate the degree of interpolation between $p(w|R)$ and $p(w|S)$, as shown in FIG. 1. However, selecting a subset of training data instances of $R$ (held-out data) reduces the data that can be used for training even further. This can out-weight the benefits that can be gained from setting the interpolation parameters with the held-out data. This problem is only partly mitigated by cross-validation, which, furthermore, can be computationally expensive. In FIG. 1, X <= Y means document set Y contains document set X. Due to the analogy of geographic regions, we use the term "region", instead of the term "category" or "class".

It might appear that another obvious solution, would be to use the same training data twice, once for estimating the probability $p(w|R)$ and once for estimating the optimal weight $\lambda$. However, using the approaches like described Non-Patent Document 1 or Non-Patent Document 2, would simply set the weight of $\lambda$ to 1 for $p(w|R)$, and zero for $p(w|S)$. This is because their method requires point-estimates of $p(w|R)$, which is a maximum-likelihood or maximum-a posterior estimate, that cannot measure the uncertainty of the estimate of $p(w|R)$.

### Means for Solving the Problem

According to an aspect of the present invention, there is provided a document classification method comprising:

a first step for calculating smoothing weights for each word $w$ and a fixed class $R$, the first step including, given a set of classes $\{R, S_1, S_2, \ldots\}$ where class $R$ is subsumed by class $S_1$, class $S_1$ is subsumed by class $S_2$, ..., calculating for each class $S$ probability over probability $p(w|S)$ representing probability that word $w$ occurs in a document belonging to class $S$, and, for each of these probabilities over the probabilities $p(w|S)$, calculating the likelihood of the training data observed in class $R$;

a second step for calculating smoothed second-order word probability, the second step including, among all the probabilities over the probability $p(w|S)$ (there is one for each $S \in \{R, S1, S2, \ldots\}$), selecting the one which results in the highest likelihood of the data as calculated in the second step before, the selected probability being used as the smoothed second-order word probability for $p(w|R)$; and

a third step for classifying document including calculating the probability that the document belongs to the class $R$ by using the smoothed second-order word probability to integrate over all possible choices of $p(w|R)$, or by using the maximum a-posteriori estimate of the smoothed estimated of $p(w|R)$.

Our approach compares the distributions of $p(w|R)$ and $p(w|S)$ and use the difference to decide if and how, the distribution $p(w|R)$ should be smoothed using only

In the final step, for a new document $d$ we calculate the probability that document $d$ belongs to class $R$, by using the probability over probability $p(w|R)$. For example, we use the naive Bayes assumption, and calculate $p(d|R)$ by probability over probability $p(w|R)$ (Bayesian Naive Bayes).

Effect of the Invention

~~The~~ Embodiments of the present invention ~~has~~ have the effect of smoothing the probability that a word $w$ occurs in a text that belongs to class $R$ by using the word probabilities of outer-classes of $R$. It is possible to achieve~~s~~ this without the need to resort to additional held-out training data.

## BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram showing the functional structure of the system proposed by previous work.

FIG. 2 is a block diagram showing a functional structure of a document clarification system according to a first exemplary embodiment of the present invention.

FIG. 3 is a block diagram showing a functional structure of a document clarification system according to a second exemplary embodiment of the present invention.

FIG. 4 shows an example related to the first embodiment.

FIG. 5 shows an example related to the second embodiment.

## EXEMPLARY EMBODIMENTS FOR CARRYING OUT THE INVENTION

<First Exemplary Embodiment>

The main architecture usually performed by a computer system is described in FIG. 2. We assume we are interested in whether the text is about region $R$ or not, which