

Maximum Entropy Models in Chief Complaint Classification

Kevin H. O'Connor, MSc., Kieran M. Moore, M.D.,

Bronwen Edgar, MHSc., Don McGuinness, MA

Queen's University Emergency Syndromic Surveillance Team/Cissec Corporation

OBJECTIVE

This paper describes a novel approach to the statistical classification of free-text chief complaints for the purpose of syndromic surveillance.

BACKGROUND

The Queen's University Emergency Syndromic Surveillance Team (QUESST) has been conducting Emergency Department (ED) syndromic surveillance since 2004. The system collects real time data from seven ED's all of which capture a free-text chief complaint. As these chief complaints contain high levels of variability and noise, syndromic classification represents one of the most important steps in the analysis of the data.

The project uses the open-source RODS program [1], which is packaged with CoCo, a Naïve Bayes n-gram classifier [2]. Naïve Bayes n-gram classification has two large drawbacks: it analyzes only at the word level, and it treats all words/word pairs as statistically independent entities.

The authors proposed the use of a classifier built using Maximum Entropy models in an attempt to address these two drawbacks. Maximum Entropy is a well known method of feature based classification [3], but its use has not previously been reported in syndromic surveillance. It was not known whether the characteristics of chief complaints would allow for accurate classification using Maximum Entropy.

METHODS

A classifier was implemented using the open-source Stanford NLP toolkit. The presence or absence of all frequently occurring character sequences, up to a maximum length, were treated as binary features. A randomly selected corpus of 10,000 records hand-classified into one of eight categories was used to train both CoCo and the new Maximum Entropy classifier. A separate corpus of 3,000 records was used to assess the performance of the two classifiers, with the hand-classification used as the gold standard. An F-Score was used as the performance metric.

At the same time, a development server was installed using the Maximum Entropy classifications. This server was otherwise identical to the production server. The development server was used over the course of six months for a comparison of the two classification schemes in the greater context of a full featured syndromic surveillance tool.

RESULTS

The Maximum Entropy classifier had a superior F-score for all three of the most frequent syndromes: 0.940 vs 0.841 for Fever/ILI, 0.971 vs 0.911 for Respiratory, 0.974 vs 0.893 for GI. Performance in the rare syndromes remains relatively low and highly variable due to the low availability of training data.

Epidemiological and clinical assessment from the comparison of the two servers indicated that the Maximum Entropy classifier was preferable. Based on these observations, the new classifier was subsequently deployed on the project's production server.

A selective analysis of errors was performed, citing the computational/linguistic reason for the error. An illustrative example was "Synope". This was classified correctly as Neurological by the Maximum Entropy model due to the presence of n-graphs "syn" and "ope" from "syncope". The complaint was misclassified by CoCo because the word had not been observed in the training set.

CONCLUSIONS

Maximum Entropy classification represents a strong alternative to Naïve Bayes classification in syndromic surveillance. The classifier's ability to analyze text at the sub-word level allows the classifier to handle the high level of noise present in the data. The classifier's ability to analyze strings containing multiple words allows the classifier to form a more sophisticated model with a modicum of semantic information. The ability to recognize the dependence of different 'features' allows for a more precise model.

The model is also easily extended by non-standard features. Meta-text features, such as the string length, could easily be integrated into such a classifier. Demographic, geographic and temporal features of the patient and/or visit could also potentially augment the model. Other feature-based statistical classification methods, such as Support Vector Machines, should be investigated to determine their relative performance.

The only drawback of Maximum Entropy relative to CoCo is the lengthy training process, since the training of a Maximum Entropy model involves an iterative scaling process. However, in a production environment, this training only needs to occur once.

REFERENCES

- [1] J. U. Espino, M. M. Wagner, et al (2003). "The RODS Open Source Project: Removing a Barrier to Syndromic Surveillance." (Submitted to MEDINFO 2004)
- [2] Olszewski RT. "Bayesian classification of triage diagnoses for the early detection of epidemics". In: Proc FLAIRS Conference: AAAI Press; 2003. p. 412-416.
- [3] Nigam, K, J. Lafferty, and A. McCallum. "Using maximum entropy for text classification". In IJCAI-99 Workshop on Machine Learning for Information Filtering.

Further Information:
Kevin O'Connor, kocconnor@cissec.com