Joffe E, et al. J Am Med Inform Assoc 2015;22:962–966. doi:10.1093/jamia/ocv010, Research and Applications

Expert guided natural language processing using one-class classification

RECEIVED 21 July 2014
REVISED 19 January 2015
ACCEPTED 26 January 2015
PUBLISHED ONLINE FIRST 10 June 2015

Erel Joffe[1,2], Emily J Pettigrew[3], Jorge R Herskovic[1,4], Charles F Bearden[1],
Elmer V Bernstam[1,5]

ABSTRACT

Introduction Automatically identifying specific phenotypes in free-text clinical notes is critically important for the reuse of clinical data. In this study, the authors combine expert-guided feature (text) selection with one-class classification for text processing.

Objectives To compare the performance of one-class classification to traditional binary classification; to evaluate the utility of feature selection based on expert-selected salient text (snippets); and to determine the robustness of these models with respects to irrelevant surrounding text.

Methods The authors trained one-class support vector machines (1C-SVMs) and two-class SVMs (2C-SVMs) to identify notes discussing breast cancer. Manually annotated visit summary notes (88 positive and 88 negative for breast cancer) were used to compare the performance of models trained on whole notes labeled as *positive* or *negative* to models trained on expert-selected text sections (snippets) relevant to breast cancer status. Model performance was evaluated using a 70:30 split for 20 iterations and on a realistic dataset of 10 000 records with a breast cancer prevalence of 1.4%.

Results When tested on a balanced experimental dataset, 1C-SVMs trained on snippets had comparable results to 2C-SVMs trained on whole notes ($F = 0.92$ for both approaches). When evaluated on a realistic imbalanced dataset, 1C-SVMs had a considerably superior performance ($F = 0.61$ vs. $F = 0.17$ for the best performing model) attributable mainly to improved precision ($p = .88$ vs. $p = .09$ for the best performing model).

Conclusions 1C-SVMs trained on expert-selected relevant text sections perform better than 2C-SVMs classifiers trained on either snippets or whole notes when applied to realistically imbalanced data with low prevalence of the positive class.

Keywords: one class classification, novelty detection, natural language processing, feature selection

INTRODUCTION

Automatically identifying specific phenotypes in free-text clinical notes is difficult, yet critically important for the reuse of clinical data. For example, a researcher may want to identify all patients with an unusually good response to treatment, or a physician may want to retrieve notes relevant to breast cancer from a large patient record. Natural Language Processing (NLP) methods usually train classifiers on sets of whole notes labeled as *positive* or *negative* (for a phenotype) by domain experts (supervised binary classification). However, often much of the text in the notes is irrelevant to the studied phenotype (e.g., discussions of unrelated conditions or normal findings). This is particularly true for *negative* notes which may negate the studied condition (e.g., *"breast cancer was ruled-out"*), but more commonly are entirely irrelevant (e.g., discuss diabetes). Thus, note classification is actually a multiclass classification task which tries to distinguish between the *positive, negative,* and *irrelevant* classes.

Building classifiers on noisy data and using *irrelevant* notes as examples of the *negative* class results in poor performance and over-fitting.[1,2] However, previous attempts to reduce noise by text selection (feature selection) has had limited success.[3] Further, it is difficult to generate a representative training set of *negative* and *irrelevant* classes. Truly n*egative* notes are rare (e.g., there are few notes that explicitly discuss the absence of breast cancer), and representing the *irrelevant* class uniformly is practically impossible (e.g., it is difficult to represent all the text that "is not" related to breast cancer). A possible solution is to limit the text analysis to the well-characterized positive class, and to train a one-class classification (1-CC) model on examples of that class.[1]

One-class classifiers analyze only positive examples of data to learn a classification boundary that excludes negative outlier values.[1] When dealing with heavily imbalanced datasets characteristic of text classification, this approach has been demonstrated to outperform conventional two-class classification.[4,5] However, since models are trained only on positive examples, they are very sensitive to irrelevant text. In other words, the presence of "noise" makes it difficult to identify text that is unique to the *positive* class.[6,7]

One way to mitigate this problem is to perform some type of dimensionality reduction (feature selection) to remove as much *irrelevant* text as possible. For 1-CC this is difficult because most feature selection methods rely on evaluating performance on counter examples from a *negative* class (i.e., assume that *positive* and *negative* examples can be identified). There has been little prior work on automatic feature selection for 1-CC. The few studies that have been done demonstrated only a modest improvement in performance.[7–9] An alternative to automatic approaches could be to have experts select features.

An expert annotating a clinical note could, with only a little additional effort, highlight the sections (or snippets) of relevant text on which s/he based their decision.[10] Subsequently these snippets could be used for expert-guided feature selection (text selection) in the construction of NLP classifiers, particularly for 1-CC. Following this rationale, Zaidan et al. assigned higher weights to text selected in snippets to augment training of a support vector machine (SVM) classifier, demonstrating improved accuracy in the classification of movie reviews.[10] Yu et al.,[11] used a similar approach for classifying passages of clinical text. However, both Zaidan and Yu evaluated their methods using an

Correspondence to Elmer V. Bernstam, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, 7000 Fannin St., Suite 600, Houston, TX 77030, USA; Elmer.v.bernstam@uth.tmc.edu

For numbered affiliations see end of article.

experimental dataset that was limited to concise positive and negative passages (i.e., short sections of text without an *irrelevant* class) using 2-CC. To extend expert-guided feature selection to 1-CC we constructed models solely on snippets (as opposed to assigning higher weights to expert-selected text). Intuitively, such an approach complements 1-CC as noise is manually excluded from the positive training examples.

This study aimed to evaluate a novel methodology for facilitating machine learning applied to text classification. Our specific objectives were to: 1) evaluate the utility of expert-selected features for NLP, by comparing the performance of classifiers trained on relevant phrases (snippets) to that of classifiers trained on whole notes for note classification; 2) evaluate the combination of 1C-SVM and expert-selected features; and 3) determine how much of the irrelevant surrounding text can a snippet contain without compromising performance (i.e., the required accuracy of snippets).

METHODS

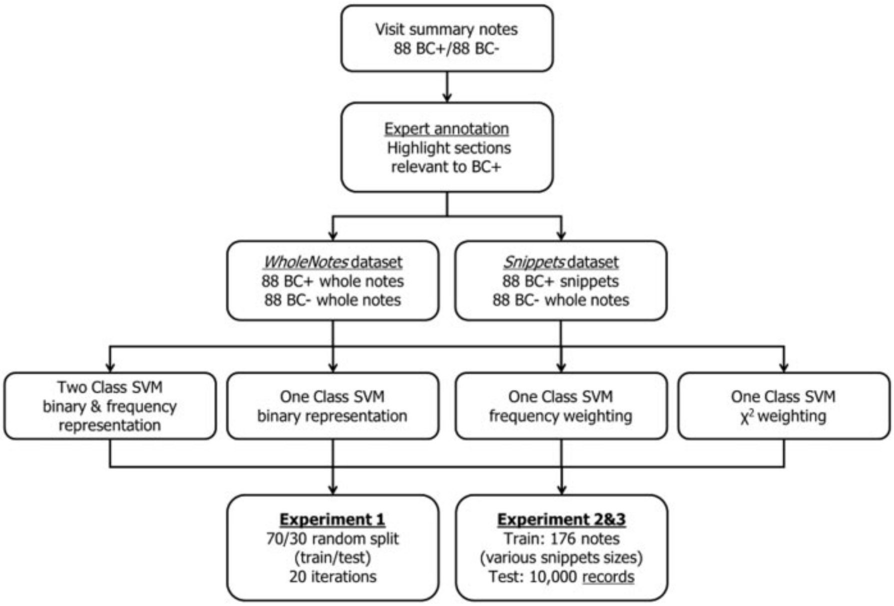
Sample notes and annotation

We built classifiers for identifying notes pertaining to breast cancer using one- and two-class classification systems (see Figure 1 for an overview of the methodology). We used 176 notes of the visit-summary type. Of these, 88 notes discussed breast cancer and 88 notes were from age- and sex-matched patients without breast cancer. Positive notes were derived from a manually annotated corpus of 95 medical records of patients positive for breast cancer (identified by ICD-9-CM code and validated during manual annotation). Annotation was performed by two physicians who were asked to determine whether each of the notes in the record indicated that the patient had breast cancer, and to highlight all the sections of text (from few words to whole sentences) containing information relevant for making that assertion (snippets). Annotation was performed with a simple Graphical User Interface (GUI) using the computer mouse to highlight relevant text. To estimate inter-annotator agreement physicians were asked to annotate a shared set of five additional records. Inter-annotator agreement was calculated based on the overlap in the words annotated by both physicians. Inter-annotator agreement was calculated based on the overlap in the words annotated by both physicians and averaged 56% across the five documents. Notably, one of the reviewers highlighted almost twice as many words in the five shared documents compared to the other reviewer (3303 vs. 1869). Of the 95 annotated records, we extracted 88 visit summaries that discussed breast cancer. We then extracted 88 visit summaries from randomly selected age- and sex-matched patients without breast cancer (negative notes). Each note was preprocessed to remove system generated text such as electronic signatures and headers. We removed stop-words and stemmed the remaining words with a Porter stemmer.[12] We used the 176 notes to generate two datasets. The first, *Whole Note* dataset, included notes labeled as positive or negative. In the second, the *Snippets* dataset, all highlighted snippets from the record were concatenated to create a single positive note. We chose to concatenate all highlighted snippets into a single positive note because records contained a variable number of snippets and snippet length was also highly variable. The set of negative notes remained the same.

Feature representation for one and two class SVMs

For each dataset, we trained a one-class support vector machine (1C-SVM) and a two-class SVM (2C-SVM), both with a linear kernel. We chose the linear kernel for simplicity and applicability to a binary representation of text features.[5] Both SVMs were implemented using LibSVM and the SciKit-learn Python module. In the case of the 1-CC we used Scholkopf's implementation of a 1C-SVM classifier which has been demonstrated to outperform other methods of 1-CC.[1,6,13] For the 2C-SVM, we represented each note as a bag-of-words, annotating each word with either its presence/absence (binary) or its frequency in the data set. For the 1C-SVM we first used a binary bag-of-words representation because cropping the snippets distorts the word frequencies in the notes (i.e., the training set which was constructed of snippets had different word frequencies compared to the test-set

Figure 1: Overview of study methodology.



which contained whole notes). However, when employing a binary bag-of-words representation, common and rare words receive the same value (i.e., 0 or 1). To give a higher ranking to words with a stronger association with snippets in the 1C-SVM we further evaluated two weighting schemes.[5] In the first, we weighted words by their frequency in all members of the positive class (whole notes or snippets depending on the training dataset). For example, there were 39 735 words in the 88 positive notes (with stop words removed), and the word “final” appears 5 times, so the feature “final” had the weight 5/39 735=0.0001258 when the 1C-SVM using this representation was trained. In the second weighting scheme, we weighted each word by its correlation-coefficient, a variant of the χ^2 test, which looks only at words from the *positive* class. We defined the coefficient χ^2 as:

$$\chi 2 = \frac{(N_{r+} + N_{n-} - N_{r-} - N_{n+})\sqrt{\mathrm{N}}}{\sqrt{(N_{r+} + N_{r-})(N_{n+} + N_{n-})(N_{r+} + N_{n+})(N_{r-} + N_{n-})}}$$

where $Nr+/Nr-$ is the number of snippets in which the word occurs/does not occur, and Nn^+/Nn^- is the number of nonsnippet section of *positive* notes in which the word occurs/does not occur (for a detailed explanation see[14]).

Finally, we evaluated the use of 2C-SVM with snippet-based feature selection, χ^2/frequency weighting of whole notes, and χ^2/frequency weighting of snippets. The rationale of this analysis was to establish the independent effect on performance of 1-CC, snippets, and weighting. The parameters of the classifiers were set using a 3-fold cross validation, once to optimize *F* score and once to optimize precision. We trained classifiers on a random 70% split of the dataset, and tested on the remaining 30%. We repeated the process 20 times and report the average precision, recall, and *F* score (Experiment 1).

Evaluation of generalizability on a 10 000 record set

In the second part of the experiment we evaluated the generalizability of classifiers. Best performing classifiers were trained on the entire dataset and tested on notes from 10 000 medical records with a 1.4% prevalence of breast cancer. Records for the test set were randomly sampled from all records belonging to females 40–80 years old that contained at least one visit note produced by a primary care physician. Positive and negative records were defined based on the presence or absence of an ICD-9-CM code corresponding to breast cancer. The classifier was run against all notes in the record. A record was defined as *positive* if it contained one or more notes flagged by the classifier as *positive* for breast cancer.

Evaluation of performance with various lengths of snippets

Selecting text for snippets is not precise and experts varied in the amount of text that they highlighted for inclusion in snippets. For example, one could highlight entire paragraphs, entire sentences, or parts of sentences. Therefore, we determined the effect of varying the amount of highlighted text on performance. To accomplish this, we created two more training datasets: the *Paragraphs* dataset, in which we replaced snippets with the paragraphs that contained them and the *Surrounding Sentences* dataset, in which we replaced snippets with the sentences that contained them. We repeated the performance evaluation of classifiers trained on these datasets against the 10 000 record test set (Experiments 2 and 3). Sentence tokenization was performed with the general English tokenizer implemented by the Punkt module in NLTK. For simplicity a paragraph was defined as the snippet plus the two sentences preceding and the two sentences following it. As an additional baseline, we present the performance of exact string matching for the term “breast cancer.”

Table 1: Classifier performance on the experimental test set maximizing *F*-score

	Precision (SEM)	Recall (SEM)	*F*-score (SEM)
2C-SVM *Whole Notes*	**0.93** (0.05)	**0.92** (0.05)	**0.92** (0.03)
2C-SVM *Whole Notes* (*f*)	**0.94** (0.05)	**0.91** (0.06)	**0.92** (0.05)
2C-SVM *Whole Notes* (χ^2)	**0.93** (0.06)	**0.9** (0.05)	**0.91** (0.05)
2C-SVM Snippets FS*	**0.94** (0.03)	**0.93** (0.04)	**0.93** (0.04)
2C-SVM Snippets*	0.47 (0.17)	0.59 (0.40)	0.45 (0.21)
1C-SVM Whole Notes (f)	0.79 (0.08)	0.88 (0.05)	0.83 (0.04)
1C-SVM Whole Notes (χ^2)	0.87 (0.06)	0.91 (0.07)	0.88 (0.04)
1C-SVM Snippets (f)	**0.91** (0.07)	**0.93** (0.05)	**0.92** (0.04)
1C-SVM Snippets (χ^2)	**0.92** (0.03)	**0.91** (0.06)	**0.91** (0.02)
String match “breast cancer”	0.93†	0.75†	0.83†

In bold, results of best performing classifiers that were also tested on the 10 000 record realistic test-set.

SEM - standard error of mean.

*2C-SVM *Snippets* FS – maintained the structure of positive notes but limited the bag-of-words vector to words appearing in snippets. 2C-SVM *Snippets* – used as positive notes the snippets sections only thus distorting frequency measures.

†No SEM as string matching was evaluated once on the entire dataset ($n=176$).

RESULTS

The 2C-SVM performed best when trained on whole notes (whole notes: $F=0.92$; snippets $F=0.45$). The 1C-SVMs (frequency and χ^2 weights) performed best when trained on snippets (snippets and frequency weights $F=0.92$ vs. whole notes and frequency weights $F=0.83$; χ^2 weights $F=0.91$ vs. 0.88) and had comparable results to those of the 2C-SVM. 2C-SVM trained on the snippets performed poorly (Table 1).

Performance on the 10 000 record set

On the set of 10 000 records, with parameters set to maximize precision (Table 2), 1C-SVMs trained on snippets out-performed the 2C-SVM trained on whole notes (1C-SVMs $F=0.61$ for χ^2 weights; $F=0.40$ for frequency weights; $F=0.11$–0.17 for 2C-SVM). This was due to a large difference in precision (1C-SVMs trained on snippets $p=.53$ for χ^2 weights; $\mathrm{P}=0.88$ for frequency weights; $p=.06$–.09 for 2C-SVM). Use of 2C-SVMs with weighting, snippets-based feature selection, or the combination of the two did not improve performance (2C-SVM + snippets + weighting data not presented).

Performance on varying lengths of snippets

On this set, 1C-SVMs trained on the *Paragraphs* database and on the *Surrounding Sentences* database also outperformed the 2C-SVM trained on whole notes ($F=0.36$ and 0.35 for frequency representations, and $F=0.59$ and 0.62 for χ^2 weights, for Paragraphs and *Sentences*, respectively).

DISCUSSION

When tested on a balanced experimental datasets, 1C-SVMs trained on snippets performed comparably to traditional 2C-SVMs trained on whole notes. However, when evaluated on a realistic imbalanced

Table 2: Classifier performance on a realistic (imbalanced) test set maximizing precision

	Precision	Recall	*F* score
2C-SVM *Whole Notes*	0.06	0.84	0.11
2C-SVM *Whole Notes* (*f*)	0.09	0.9	0.17
2C-SVM *Whole Notes* (χ^2)	0.09	0.9	0.16
2C-SVM *Snippets FS*	0.07	0.89	0.13
1C-SVM *Snippets* (*f*)	**0.88**	**0.26**	**0.40**
1C-SVM *Snippets* (χ^2)	**0.53**	**0.73**	**0.61**
1C-SVM *sentences* (*f*)	0.30	0.40	0.35
1C-SVM *sentences* (χ^2)	**0.63**	**0.60**	**0.62**
1C-SVM *paragraphs* (*f*)	0.84	0.23	0.36
1C-SVM *paragraphs* (χ^2)	**0.88**	**0.45**	**0.59**
String match "breast cancer"	0.24	0.90	0.32

In bold, results of best performing classifiers.

dataset, 2C-SVMs performed poorly. On these data, 1C-SVMs performed considerably better (i.e., these models are generalizable), due mainly to greatly improved precision. Superior performance was noted even with overly broad identification of snippets (i.e., including surrounding sentences or paragraph as well as the relevant text), suggesting that the approach is robust to variation in expert annotation.

Identifying specific phenotypes in large clinical datasets is an increasingly important task. However, even common diseases have relatively low population prevalence. Therefore, large clinical datasets are very imbalanced with respect to most phenotypes (e.g., diseases). The novelty of this work is in adapting supervised methods of feature selection on imbalanced datasets to 1-CC.

This study has several limitations. First, data were derived from a single institution. Therefore, we can only conclude that 1C-SVM had superior generalizability from the experimental dataset to the dataset at our institution. However, the training dataset was very small (88 positive notes) compared to the 10 000 records in the (realistic, imbalanced) test set. Therefore, this approach is likely to work well on other datasets. Second, performance on the realistic dataset was imperfect with maximal *F* measure of 0.6. This is because the study was designed to compare the performance of the 1C-SVM to that of the 2C-SVM rather than to identify an optimal solution to the problem of note classification. Performance could be improved with the combination of other NLP techniques such as collocation (e.g., "breast cancer" as a single feature as opposed to "breast" and "cancer"), negation identification (e.g., "no evidence of breast cancer"), section tagging (e.g., "breast cancer in mother" under the family-history section), and mapping to the Unified Medical Language System (e.g., identification of "infiltrative ductal carcinoma" as indicating that the patient has "breast cancer"). Third, it might be argued that this study addresses the problem of concept-level identification that could have been more simply handled by asking experts to specify all the relevant concepts or using an existing ontology. We picked the task of identifying "breast-cancer" for its simplicity, but the methodology could be used to identify conditions for which an ontology does not exist (e.g., identifying "unusual responders" to chemotherapy). Alternatively, if a tumor registry is available for the population, it could be used to identify patients with breast cancer with much less effort than our approach. Finally, it is possible that a multi-class approach, whereby experts would annotate all possible classes (i.e., positive, negative, neutral, and irrelevant) could outperform 1-CC. However, identifying negative cases and uniformly representing the irrelevant class is a challenging task which was beyond the scope of this study.

The reduction in performance of 2C-SVM between the experimental dataset and the realistic dataset suggests that for rare-class classification binary models tend to over-fit.[5,15] Previous research has demonstrated the advantage of 1-CC in such settings.[4,5] However, all these studies focused on classifying very short passages (as opposed to our task of using whole notes) thus avoiding 1-CC susceptibility to "noise."[6,7]

Few studies evaluated feature selection specifically devised for 1-CC; mostly using unsupervised methods. These demonstrated only a limited improvement in performance.[7,8] Other studies have evaluated supervised feature selection methods for 2C-SVMs which we adapted here to 1-CC. Liu et al. used weights based on a probability metric calculated only for text from the *positive* class to augment 2C-SVMs; an approach similar to our use of weighting based on correlation coefficients. They demonstrated improved performance compared to other methods of feature selection on imbalanced datasets.[2] We found that weighting did not improve 2C-SVM performance; possibly due to the relatively high baseline performance. However, weighting was essential for 1C-SVM. The intuition is that in the context of 1-CC, weighting provides an indication of how unique a feature is to the positive class. Ng et al. suggested limiting the feature vector only to words appearing in the *positive* class while training on both *positive* and *negative* notes. This approach is similar to 2C-SVM with a feature vector limited to words from snippets and was found to outperform other methods of feature selection.[14] Again, in our experiment this approach did not improve the performance of 2C-SVM but had a considerable effect on the performance of 1C-SVMs ($F = 0.83$ vs. $F = 0.92$ for 1C-SVM trained on *Whole notes* vs. *Snippets*, respectively). In this study, we expand on this method through a "semi-explicit" feature selection by having expert(s) select relevant sections of text. A similar approach was used by Zaidan et al. and by Yu et al. who demonstrated a modest improvement over traditional 2-CC (i.e., without feature selection). The modest results of these studies are in-line with our observation that expert guided feature selection does not significantly improve performance on a balanced experimental dataset. However, neither study evaluated this form of feature selection for imbalanced datasets or in combination with 1-CC. We found that in text classification, model evaluation on balanced experimental datasets gives an overly optimistic estimation of performance. Subsequently, when evaluating models on realistic imbalanced dataset the advantage of the combination of expert-guided feature selection and 1-CC over traditional 2-CC becomes apparent.

Manually identifying key concepts using expert annotation is labor-intensive, requires extensive training and is hard to maintain over time.[16–18] Snippet annotation could, at least in part, reduce the required human effort. To this end, it is particularly relevant to note that NLP performance was improved even when using an imprecise identification of snippets within the surrounding sentences or paragraphs. Theoretically in future systems relevant sections of text could be selected implicitly as part of the routine interaction of experts with the electronic health record; for example, by eye tracking or mouse tracking to identify which sections of the text experts read more attentively.[19]

CONCLUSIONS

Using expert-selected text of interest (snippets) combined with one class classification results in a considerably improved performance on realistically imbalanced datasets. Performance improves even with an overly broad identification of snippets (i.e., within surrounding sentences or paragraph).

FUNDING

This work was supported by the Brown Foundation, National Institutes of Health NCATS grants UL1 TR000371 and UL1 TR001105, National Science Foundation grant III 0964613, and NIH NCI grant U01 CA180964 and by a training fellowship from the Keck Center Computational Cancer Biology Training Program of the Gulf Coast Consortia (CPRIT grant no. RP101489).

COMPETING INTERESTS

None.

CONTRIBUTORS

E.J. and E.V.B. conceived the methods. E.J., E.P., J.R.H. and C.F.B. implemented and tested the software used to collect data and perform the analyses. E.J. and E.V.B. drafted the manuscript. All authors read and agreed with the analysis and the manuscript.

REFERENCES

1. Khan SS, Madden MG. One-class classification: taxonomy of study and review of techniques. *Knowl Eng Rev.* 2014;29:345–374.
2. Liu Y, Tong H, Sun A. Imbalanced text classification: a term weighting approach. *Expert Syst Appl.* 2009;36:690–701.
3. Zheng Z, Wu X, Srihari R. Feature selection for text categorization on imbalanced data. *ACM SIGKDD Explor Newsl.* 2004;6:80.
4. Zhuang L, Dai H. Parameter optimization of kernel-based one-class classifier on imbalance learning. *J Comput.* 2006;1:32–40.
5. Raskutti B, Kowalczyk A. Extreme re-balancing for SVMs. *ACM SIGKDD Explor Newsl.* 2004;6:60.
6. Manevitz L, Yousef M. One-class document classification via Neural Networks. *Neurocomputing.* 2007;70:1466–1481.
7. Villalba SD. An evaluation of dimension reduction techniques for one-class classification. *Artif Intell Rev.* 2008;27:273–294.
8. Lian H. On feature selection with principal component analysis for one-class SVM. *Pattern Recognit Lett.* 2012;33:1027–1031.
9. Tax D. “One-class Classification Concept-learning in the Absence of Counter-examples.” Thesis. Delft University of Technology, 2001.
10. Zaidan OF, Eisner J, Piatko CD. Using “Annotator Rationales” to improve machine learning for text categorization. *Comput Linguist.* 2007;260: 260–267.
11. Yu S, Farooq F, Krishnapuram B, *et al.* Leveraging rich annotations to improve learning of medical concepts from clinical free text. *AMIA Annu Symp Proc.* 2011;2011:1603–1611.
12. Porter M. An algorithm for suffix stripping. *Program.* 1980;14:130–137.
13. Schölkopf B, Platt JC, Shawe-Taylor J, *et al.* Estimating the support of a high-dimensional distribution. *Neural Comput.* 2001;13:1443–1471.
14. Ng HT, Goh WB, Low KL. Feature selection, perceptron learning, and a usability case study for text categorization. *SIGIR Forum (ACM Spec Interes Gr Inf Retrieval).* 1997;31:67–73.
15. Hempstalk K, Frank E. Discriminating against new classes: one-class versus multi-class classification. In: Wobcke W, Zhang M, eds. *AI 2008: Advances in Artificial Intelligence*. Berlin: Springer; 2008.
16. Roberts A, Gaizauskas R, Hepple M, *et al.* Building a semantically annotated corpus of clinical texts. *J Biomed Inform.* 2009;42:950–966.
17. Wilbur WJ, Rzhetsky A, Shatkay H. New directions in biomedical text annotation: definitions, guidelines and corpus construction. *BMC Bioinformatics.* 2006;7:356.
18. Neves M, Leser U. A survey on annotation tools for the biomedical literature. *Brief Bioinform.* 2014;15:327–340.
19. Joffe E, Havakuk O, Herskovic JR, *et al.* Collaborative knowledge acquisition for the design of context aware alert systems. *JAMIA.* 2012;19: 988–994.

AUTHOR AFFILIATIONS

[1]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Texas

[2]Department of Hematology and Bone Marrow Transplantation, Tel Aviv Medical Center, Tel Aviv Israel

[3]Department of Computer Science, Rice University, Houston, Texas

[4]The University of Texas, M.D. Anderson Cancer Center, Houston, Texas

[5]Department of Internal Medicine, Medical School, The University of Texas Health Science Center at Houston, Houston, Texas