Classification Of Nonprofit Organizations: A Supervised Machine-learning Approach

Ji Ma and Isha Kanani *

University of Texas at Austin

November 21, 2018

Abstract

This research note reports the use of supervised machine-learning algorithms in classifying the nonprofit organizations in the United States. Mission statements and project descriptions are collected from the 990 forms as text data, and classifications using National Taxonomy of Exempt Entities are collected from the National Center for Charitable Statistics at the Urban Institute. Three text classification algorithms are experimented: Naïve Bayes, Random Forest, and Neural Network. The Neural Network classification achieves the best results with an average accuracy of 9*.9% (standard deviation **), recall *** (standard deviation **), and precision *** (SD **). An open-source Python package *npocat* is developed and shared using the trained algorithms. Future projects are discussed.

*J.M.: maji@austin.utexas.edu, LBJ School of Public Affairs and RGK Center for Philanthropy and Community Service; I.K.: ishakanani@utexas.edu, School of Information.

Todo list

1 Introduction

Although the voluntary and philanthropic organizations have long been existent for numerous centuries, the so-called "nonprofit sector" was only coined in the 1970s by scholars, policy makers, and nonprofit practitioners (Hall, 2006). A major reason for assembling the diverse organizations as a conceptual whole is to legitimize the existence of these organizations and the benefits these organizations receive (Barman, 2013; Hall, 2006). From Durkheim's 2012 perspective, the order and structure of a society can be reflected by a classification system. The National Taxonomy of Exempt Entities (NTEE) developed by the National Center for Charitable Statistics (NCCS), the most widely used classification system, is one of the efforts legitimizing the existence of nonprofit sector (Hodgkinson, 1990; Hodgkinson & Toppe, 1991). As Barman (2013, p. 105) cite Clarke and Casper (1996, p. 601): "The ways in which different entities (people, animals, plants, diseases, etc.) are organized into classificatory groups reveal something of the social, cultural, symbolic, and political contexts within which classifications occur."

NTEE practical use (Hodgkinson, 1990; Hodgkinson & Toppe, 1991).

The development of NTEE classifications can date back to the 1980s.

Brief history and introduction of NTEE.

(Hodgkinson, 1990; National Center for Charitable Statistics, 2006, p. 16)

The NTEE classifications has been used for numerous practical and academic purposes. For example, it provides a framework on which the social and economic activities of nonprofits can be mapped and compared with other types of organizations in a society (e.g., Roeger, Blackwood, & Pettijohn, 2015). Scholars also use NTEE codes for sampling purposes (e.g., Carman & Fredericks, 2010; Okten & Weisbrod, 2000) or as independent variables (Sloan, 2009). The invention of an international classification system, although challenging, can be the cornerstone for studying "global civil society" (Lester M. Salamon & Anheier, 1992; Lester M Salamon, Anheir, & coaut, 1996; Vakil, 1997).

The NTEE classification system, although one of the best we have, still has several major drawbacks. First, because it only assign one major category code to an organization, it cannot accurately describe a nonprofit organization's programs which are usually diverse and across several domains (Grønbjerg, 1994, p. 303). Although another classification system assigning purpose codes to programs was developed (Lampkin, Romeo, & Finnin, 2001), it is not widely used (why?). Second, the assignment of NTEE codes is not complete because it is "based on an assessment of program descriptions contained in Parts 3 and 8 of the Form 990" and "program descriptions were only available for some organizations" (National Center for Charitable Statistics, 2006, p. 16). A recent study found the number of organizations in Washington State with a specific NTEE code could be significantly increased if the mission statements were used for coding (Fyall, Moore, & Gugerty, 2018). Third, NTEE codes are static but nonprofit organizations' activities may change over time. Recoding existent NTEE assignments is extremely onerous for human.

NTEE limitations: Lester M. Salamon and Anheier (1992).

Contribution of this study.

2 Method

2.1 Working with Texts and Research Workflow

The classification of texts is a typical task of automatic content analysis, and three types of methods are common to this task: dictionary, supervised, and unsupervised methods (Grimmer & Stewart, 2013, pp. 268–269). The dictionary methods use a predefined dictionary of words to classifying the texts. Although accurate, this approach is not capable to deal with the variations and contexts of language. An improved solution is to use supervised methods which are computer algorithms that can "learn" the linguistic patters in a dataset classified by human coders. Unlike the dictionary and supervised methods which require predefined categories of interest, unsupervised methods can discover linguistic patters in texts without inputting any knowledge of classification. However, the validity of unsupervised methods is a serious challenge because the classifications returned may not be theoretically meaningful. This study employs supervised methods to make the use of existing classifications and human-coded records and deal with linguistic variations and contexts.

Research workflow.

2.2 Datasets and Sampling

There are two types of datasets for supervised text classification: training dataset and testing dataset. Both datasets are collections of text records that have been classified by human coders. The machine-learning algorithms can "learn" the linguistic patters from the training dataset and then classify the records in testing dataset using the patters learned. The results generated by algorithms can be compared with those coded by human coders, and ultimate goal is to use trained models to replace human. The quality of training dataset is decisive because it must be a representative sample of the whole corpus. The training dataset can be generated by randomly sampling the whole corpus, but a better strategy is proportional sampling according to the distribution of classification scheme (Grimmer & Stewart, 2013, p. 278).

2.2.1 Text Data

How NTEE codes are assigned: by whom according to what? Where to obtain text data?

(Hodgkinson, 1990; National Center for Charitable Statistics, 2006, p. 16)

2.2.2 Classification Data

BMF file.

2.2.3 Sampling

Why and how to do the bootstrap sampling (Erceg-Hurn & Mirosevich, 2008, p. 596).

NTEE classification confidence rating (A/B/C).

Figure **??** shows the proportions of NTEE major categories from 1989 to 2015 ...

Figures or tables describing the patterns of datasets.

2.3 Machine-Learning Algorithms

Introduction to the algorithms.

This study uses three supervised machine-learning classification algorithms: Naïve Bayes, Random Forest, and Neural Network. Other than these three individual methods, an ensemble approach that combines all the three models is also experimented.

2.3.1 Naïve Bayes Classification

Naïve Bayes Classification is a machine learning algorithm that works on probabilistic approach. Given a set of features, this classifier predicts the class with the highest probability of the feature set. The algorithm is primarily built on Bayeś theorem.

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Bayeś theorem gives probability of the occurrence of class A, while B is provided. B here is subset of all parameters. Bayeś theorem assumes all parameters to be independent of each other, it simply follows principles of conditional probability.

In the context of text classification: The classifier depends on bags of word representation, which consists of all important words for classification and their frequency. In training data set, each "text" is converted to bag of word, and given as an input along with it's "label" need to define text and label to use it throughout the paper. For testing purpose, the classifier is given a bag of words and asked to predict the label with the highest probability for given set.

2.3.2 Random Forest Classification

Random forest classifier consists of a group of decision trees. Each decision tree is trained with unique subset of the training set. During the prediction, each forest predicts the class of the input on it's own and the final class predicted is derived from predictions of all decision trees.

2.3.3 Neural Network Classification

Intro and mechanism.

2.4 Measuring Algorithm Performance

The performance of a classification algorithm can be measured by accuracy, precision, and recall. The *accuracy* measures the percentage of correctly classified organizations as showed in Eq. 1, where i is one of the three classification algorithms (i.e., NB, RF, and NN), $Org^{correct}$ is the number of organizations correctly classified by the algorithm i, and Org^{total} is the total organizations to be classified. For example, $Accuracy^{RF} = 0.6$ indicates that, when RF classifies an organization, the chance of getting right is 60%.

$$Accuracy^{i} = \frac{Org^{correct}}{Org^{total}} \tag{1}$$

The *precision* and *recall* measures the performance of a classifier on a specific category. In Eq. 2, k is one of the NTEE codes, $Org_{k}^{correct}$ is the number of organizations correctly classified as k by algorithm i, and Org_{k}^{i} is the number of organizations classified as k by algorithm i. $Org_{k}^{correct}$ will always be smaller than or equal to Org_{k}^{i} because ML algorithms can hardly predict everything right. For example, $Precision_{B}^{NN} = 0.75$ indicates that 75% of all the organizations classified as "education" by the NN algorithm are correct.

$$Precision_{k}^{i} = \frac{Org_{k}^{correct}}{Org_{k}^{i}} \tag{2}$$

Given a human coder labels an organization as category k, the *recall* measures the chance the classifier i also identifies the organization as k. In Eq. 3, Org_{k}^{hum} is the number of organizations that has been classified as k by human coders. For example, $Recall_{B}^{NN} = 0.80$ denotes that 80% of the organizations classified as "education" by human coders are correctly identified by the NN algorithm.

$$Recall_{k}^{i} = \frac{Org_{k}^{correct}}{Org_{k}^{hum}} \tag{3}$$

3 Results

3.1 Confusion Matrix

References

Barman, E. (2013). Classificatory Struggles in the Nonprofit Sector: The Formation of the National Taxonomy of Exempt Entities, 1969—1987. *Social Science History*, *37*(1), 103–141. 00021. Retrieved July 23, 2018, from http://www.jstor.org/stable/23361114

Carman, J. G., & Fredericks, K. A. (2010). Evaluation Capacity and Nonprofit Organizations: Is the Glass Half-Empty or Half-Full? *American Journal of Evaluation*, *31*(1), 84–104. 00107. doi:10.1177/1098214009352361

Clarke, A. E., & Casper, M. J. (1996). From Simple Technology to Complex Arena: Classification of Pap Smears, 1917-90. *Medical Anthropology Quarterly*, *10*(4), 601–623. 00057. Retrieved September 4, 2018, from https://www.jstor.org/stable/648664

Durkheim, É. (2012). *The Elementary Forms of the Religious Life*. Courier Corporation.

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, *63*(7), 591–601. 00650. doi:10.1037/0003-066X.63.7.591

Fyall, R., Moore, M. K., & Gugerty, M. K. (2018). Beyond NTEE Codes: Opportunities to Understand Nonprofit Activity Through Mission Statement Content Coding. *Nonprofit and Voluntary Sector Quarterly*, *47*(4), 677–701. 00000. doi:10.1177/0899764018768019

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297. 01025. doi:10.1093/pan/mps028

Grønbjerg, K. A. (1994). Using NTEE to classify non-profit organisations: An assessment of human service and regional applications. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *5*(3), 301–328. 00055. doi:10.1007/BF02354038

Hall, P. D. (2006). A Historical Overview of Philanthropy, Voluntary Associations, and Nonprofit Organizations in the United States, 1600–2000. In W. W. Powell & R. Steinberg (Eds.), *The nonprofit sector: A research handbook* (pp. 32–65). 00000. Yale University Press.

Hodgkinson, V. A. (1990). Mapping the non-profit sector in the United States: Implications for research. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *1*(2), 6–32. 00043. doi:10.1007/BF01397436

Hodgkinson, V. A., & Toppe, C. (1991). A new research and planning tool for managers: The national taxonomy of exempt entities. *Nonprofit Management and Leadership*, *1*(4), 403–414. 00028. doi:10.1002/nml.4130010410

Lampkin, L., Romeo, S., & Finnin, E. (2001). Introducing the Nonprofit Program Classification System: The Taxonomy We've Been Waiting for , Introducing the Nonprofit Program Classification System: The Taxonomy We've Been Waiting for. *Nonprofit and Voluntary Sector Quarterly*, *30*(4), 781–793. 00022. doi:10.1177/0899764001304009

National Center for Charitable Statistics. (2006). *Guide to Using NCCS Data*. Urban Institute. 00004. Washington, DC. Retrieved September 4, 2018, from https://nccs-data.urban.org/NCCS-data-guide.pdf

Okten, C., & Weisbrod, B. A. (2000). Determinants of donations in private nonprofit markets. *Journal of Public Economics*, *75*(2), 255–272. 00436. doi:10.1016/S0047-2727(99)00066-3

Roeger, K. L., Blackwood, A. S., & Pettijohn, S. L. (2015). The Nonprofit Sector and Its Place in the National Economy. In J. S. Ott & L. A. Dicke (Eds.), *The Nature of the Nonprofit Sector* (Third edition, pp. 22–37). 00204. Boulder, CO: Westview Press.

Salamon, L. M. [Lester M.], & Anheier, H. K. (1992). In search of the non-profit sector II: The problem of classification. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *3*(3), 267–309. 00332. doi:10.1007/BF01397460

Salamon, L. M. [Lester M], Anheir, H. K., & coaut. (1996). *The international classification of nonprofit organizations ICNPO-Revision 1, 1996*. 00230 OCLC: 760476834. Baltimore, Md: The Johns Hopkins University Institute for Policy Studies.

Sloan, M. F. (2009). The Effects of Nonprofit Accountability Ratings on Donor Behavior. *Nonprofit and Voluntary Sector Quarterly*, *38*(2), 220–236. 00136. doi:10.1177/0899764008316470

Vakil, A. C. (1997). Confronting the classification problem: Toward a taxonomy of NGOs. *World Development*, *25*(12), 2057–2070. 00559. doi:10.1016/S0305-750X(97)00098-3