Automatic Detection of School Closures for Public Health Monitoring

Thomas Manzini[1*], **John Fuini**[1*], **Robert Paul**[1*], **Michael Beaudin**[1*], **Derek Norton**[1] and **Cris Benge**[1]

[1]Microsoft Corp.
first.last@microsoft.com

Abstract

In this paper we describe an approach for automatically detecting the closure of educational facilities. We motivate the need for this type of system through the lens of public health and we discuss the technical stages of our system and further explain the tradeoffs that were made during development having started from a position of limited data. Finally, we report the performance of the system on test data designed to reflect the types of reports that this system would see in the real world and find that the system is able to classify a school as closed with a precision of 0.9167 and a specificity of 0.9778. We discuss the failure cases, the biases, and the drawbacks of the system and finally conclude with the caveats necessary for a team of specialists, who will utilize this system for public health monitoring, to interpret the generated labels.

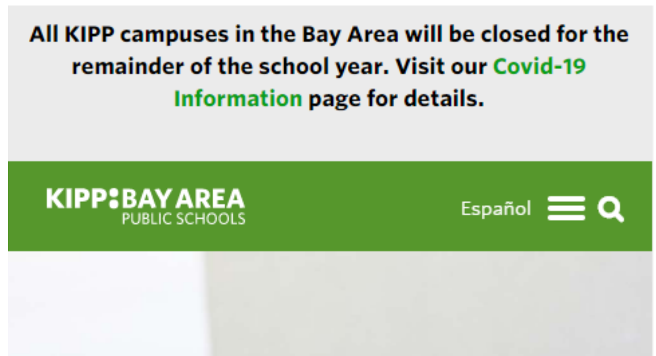

Figure 1: A common method for announcing closures is a banner-style announcement on the websites of educational facilities. These, and similar announcements, were the types of closures statements our system was designed to detect.

1 Introduction

Closing a school is a common tactic used to reduce the spread of contagion in a population[3]. This tactic has been widely applied and studied in the United States during the seasonal flu [16; 11; 12] and has also been applied in response to more complex and wide-sweeping illnesses [28; 8]. Further, the effects of school closures on public health have been extensively studied and quantified [10; 18; 29; 25]. As a result, the decision of whether to close a school is used as a barometer for community health in the area. In the recent past, most school closures have generally been limited to particular geographic areas which can be closely monitored by a specialized team of individuals. However, with the recent rise of COVID-19, school closures, distance learning and other remote lifestyles have become the norm throughout the United States [15]. This recent increase in the number of school closures has prompted the need to automatically detect these closures in order to effectively monitor public health.

In this paper we describe an automated system for detecting a portion of school closures within the United States. We explore common mechanisms that schools use to announce school closures and which types of closure announcements we decided to capture. We detail the computational machinery we employed in automating the process of school closure detection which was informed by needing to make as much use of the limited data that was available. We report the performance of our system on test data and characterize its performance as it would benefit a team of specialists already monitoring school closures. We discuss observations of our system and detail potential shortcomings and biases. Finally, we explore areas for future work in the automatic detection of school closures.

1.1 Research Question

We attempt to determine if it is possible to estimate the closure status of educational facilities based on announcements found on their websites. Further, we aim to present those findings in a manner which is useful in order to augment a team of specialists who are already working in to quantify school closures. To the best knowledge of the authors, the application of machine learning techniques for this task represents novel work which has not yet been explored in the literature.

2 Related Work

To the best knowledge of the authors, there is no published work surrounding the automated detection of school closures. However, in the public health community, a great deal of

*Authors Contributed Equally

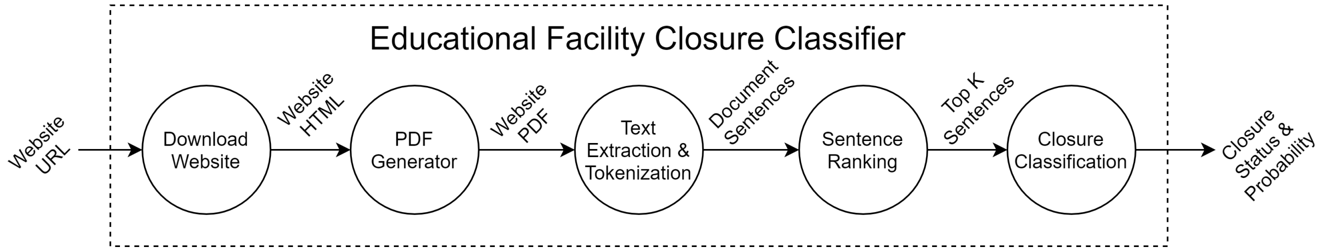

Figure 2: The high level architecture of our school closings classifier which consumes the URL of a website for an educational facility and returns a probability that the educational facility is physically closed.

work has been done in monitoring and effectively responding to events based off of school closures [13; 6]. Other work has also been done in relating school absenteeism with influenza like illnesses [24]. Other authors have developed algorithms to determine when to close a school in response to a public health crisis [26], and other authors have approached the question of when to close a school from an ethical perspective [2].

Within the computer science community, the most similar work is in the development of alerting systems based on school absenteeism. Some school districts have constructed alerting infrastructure to determine when a large number of students fail to attend school and correlate this information with public health data [21; 1], however, it should be noted that the detection of an outbreak is different from the detection of a school closure (though they may be correlated). Some work has been done in the detection of the closure of businesses and other physical spaces. Several process patents have been filed in this space, some making determinations based on business listings [9] and another based on cell phone data [17].

3 School Closures

The act of closing schools is frequently used to prevent the spread of illness and to protect the well being of schoolchildren. From a public health monitoring perspective, it is useful to determine which regions have decided to employ this tactic because it is an indicator for public concern and public health in that geographic area.

School closures can be announced through numerous mediums during a public health event. Non-exhaustively, posts can be made on social media such as Facebook and Twitter, announcements can be made on school websites, students can be informed in class by school staff. Each of these modes of communication presents different challenges in terms of access and detection. Further, there are many announcements made by educational facilities that imply physical closures, but do not explicitly state so. For example, a school district may announce a plan for students to pick up e-learning informational packets while maintaining social distancing, implying to the reader that the educational facility is physically closed, but not explicitly stating so. Such statements must be considered when evaluating if the educational facilities are indeed closed.

After reviewing the different mediums in which closures can be announced we determined that the most effective and accessible task would be to automatically detect announcements of closures and announcements that imply to the reader that the educational facility is closed found on the websites of educational facilities. An example of such an announcement can be found in figure 1.

4 System Details

At a high level, we utilize a retrieve, rank, and extract framework for our system. Approaches like this have been widely popularized by question answering and information extraction systems in numerous domains[5]. Our system starts by first downloading and processing the websites of all United States school Districts as published by the National Center for Educational Statistics (NCES). We process the contents of those websites and extract the textual content using an optical character recognition system (OCR). Following extraction all the textual content is segmented at the sentence level and presented to a ranker scoring. The top k sentences are then selected and presented to a classifier which then makes a final determination about the closure status of the presented educational facility. The overall architecture of our classifier is presented in Figure 2.

In order to provide labels for every school district in the United States, this system must process approximately 13,169 school district websites.[1] Further, as this system must be run on a daily basis to capture the daily trends, we distribute this task across numerous instances of our system and assign multiple websites to each instance.

4.1 Website Retrieval & Content Extraction

The websites for all public school districts or charter schools within the United States and territories thereof are collected from the publications of the NCES. The websites are then rendered through the use of wkhtmltopdf[27], an open source Qt WebKit rendering engine, and exported using the Adobe Portable Document Format (PDF). All the PDFs are temporarily stored for future processing and reference should further inspection be required. The PDFs are the crucial piece of information for this study as they contain the content for each district URL and the schools that reside within each district.

The content of the PDFs is extracted using the out of the box OCR system provided by the Microsoft Azure Cognitive

[1]Another possible approach is to parse URLs at a school level. A decision was made to parse at a district-level and to associate the same closure status with all schools in a district.

Services Toolkit [7]. This OCR system leaves us with spans of text which have been recognized within the PDF. These spans of text are then tokenized at the sentence level using the Punkt Tokenizer [14] as provided by the Natural Language Toolkit [20]. Following the output of these steps, we are left with a set of spans of text separated by the sentence tokenizer and the OCR system.

4.2 Sentence Ranking

Once we retrieve the various spans of text from each document, we then proceed to rank them in order of likelihood of announcing a school closure. In order to do this we trained two Kneeser-Ney Smoothed N-Gram language model [23] on sentences announcing school closure. This model estimates the probability of observing a given sentence under the assumption that the sentence was drawn from the same distribution as the training data—a distribution of sentences announcing school closures. Sentences that appear to resemble school closing sentences are assigned a higher probability by the language model, as they more closely resemble the training data, and sentences that are not about school closings are assigned lower probabilities. We term this stage of our pipeline the Sentence Ranking stage.

We specifically chose this type of model as it is able to make reasonable probability estimates despite having limited data for this task.

This model is used to evaluate the probability of all sentences retrieved from the document. Following the output of this step, we are left with a list of sentences which are ordered based on the language model's probability estimate of how closely they resemble announcements of school closures. We then select the top k sentences, whose associated probability scores are then passed to the next stage, Closure Classification, in order to estimate if the school is in fact closed or not.

4.3 Closure Classification

Having extracted the the top k sentences for each document, we then estimate the probability that a sentence is actually announcing a school closure. To do this, we trained a random forest model [19] sentences announcing school closings as positive examples, and sentences that were highly ranked by the Sentence Ranking stage which did not announce school closures as negative examples. Due to the limited data in this space these training examples selected by manually inspecting and curating the results generated by the Sentence Ranker stage. Sentences were encoded using a series of hand-crafted indicator features triggered by specific words and phrases within each sentence.As was the case with the Sentence Ranker, we specifically chose this type of model and encoding scheme because of limited data that was available in this space. We term this stage of our pipeline the Closure Classification stage.

4.4 Closure Status Reporting

The closure statuses must be organized in a manner convenient to the end-user. To accomplish this, the system saves closure statuses into a designated database that aligns with the school ID. These statuses are also saved alongside the probability of the assigned closure status, and the time the label was estimated. The end-user can then query this database and reorder in any manner suitable for reporting needs. A historical table is also used for tracebility purposes.

	Closure Announced	Closure Not Announced
Closure Predicted	33	3
Closure Not Predicted	26	132

Table 1: The confusion matrix associated with our system measuring performance based on labels extracted from human labeled PDFs (N=194).

Precision	**0.9167**
Recall	0.5593
F1	0.6947
Accuracy	0.8505
Specificity	0.9778

Table 2: The performance metrics we used in evaluating system performance based on labels extracted from human labeled PDFs (N=194).

5 System Performance

This system is designed to support and augment an existing team tasked with monitoring school closures. Specifically, to automate away the "easy" cases so that the team can focus their energy on more complex closure notifications and monitoring tasks. In this setting we have optimized the system for high precision so that we are confident that the educational facilities that are marked as closed by the system require minimal review. Different hyper-parameter settings allow for the different configurations which may alter this dynamic, however, we report results that are designed to maximize precision.

We report our performance on a hand labeled set of school district websites which have been selected randomly from the school district websites published by the NCES and have been held out during training and during system development to minimize developer bias. The confusion matrix associated with out system's performance may be found in Table 1 and the resulting performance metrics can be found in Table 2.

6 Failure Cases

Like many data pipelines, the greatest potential shortcoming is error propagation. Errors can happen at any stage in the pipeline. The URLs from the NCES publication can point to deprecated or no longer maintained websites, the PDF can fail to render the entire page, OCR can misinterpret text in images, the Sentence Ranker may fail to propagate the correct items, and the Status Extractor may fail to correctly classify the sentences. Understanding these different failure modes provides information to the end user about how to interpret the results that they receive from the system.

7 Shortcomings and Biases

Any statistical system expected to operate on noisy data in the real world will have shortcomings and biases[4; 22]. There can be numerous reasons for these behaviors but recognizing and understanding these shortcomings and biases prior to deploying and depending on your machine learning system is crucial.

In the development of the a system which is designed to act on a national scale there are going to be edge cases that are outside of your systems training data. In our setting, one of the most obvious cases is websites that are not in the English language. For example, in parts of the Southwest there are some school districts who do most of their communication with their students in Spanish. At the same time there are school districts throughout the country who serve primarily Native American communities who communicate with students in indigenous languages. As our system has not been trained on no other language data besides English, our system will likely have a disproportionately high error rate when classifying the closure statuses of these educational facilities.

One of the fundamental requirements of our system to classify an educational facility as closed or not is that it must have a website. This again may not always be the case as that not all educational facilities have websites. At the same time, closures may be announced over email, in person, or through mail.

Finally, and potentially most importantly, this system has so far only been trained and evaluated during the time of COVID. As a result there is likely a large distributional bias in the training and testing data that will likely change when COVID is no longer the dominating factor in this space.

All of these factors make it imperative that those interacting with this system have to recognize that they are only being presented with a part of the picture and there may be underlying behaviors that are more complex than what our system can represent.

8 Future Work

We believe that this work represents the initial stages of what is possible in this space after having cold-started a machine learning system on minimal data. Given this initial architecture the immediate next steps are the collection and curation of a large scale training and testing corpus so that the ranking and classification models can be further refined and their performance quantified. Once this corpus is established the application of more complex and potentially performant ranking and classification models becomes possible. This system has been built in such a way that, with interaction from a team of specialists working in this space, the system will be to generate its own training data as it crawls more school websites. The hope is that with time, this system will be able to become more robust and more performant.

In a more practical view, at the moment of writing the vast majority of educational facilities in the United States are physically closed due to the COVID-19 pandemic. As a result, those working in this space are interested in measuring the eventual re-opening of these educational facilities. Augmenting or refining this system in order to capture the reopening of these educational facilities is an obvious next step.

9 Acknowledgements

This work was supported by the efforts of Microsoft Services Disaster Response and the Centers for Disease Control. We would like to acknowledge the efforts and work of Andrew Ditmer (Microsoft), Lewis Curtis (Microsoft), and Joseph Spalviero (Centers for Disease Control) for enabling this work.

10 Conclusion

In this paper we have described an automated system for detecting a portion of school closures within the United States. We have described the technical components used in the development and we have described their interactions. We have quantified the performance of the automated system against a gold standard dataset, and we have described how this system will be used moving forward. We hope that this system will begin to ease the burden of the complex task placed on those monitoring the school closures for the purposes of public health.

References

[1] Atar Baer, Carla V Rodriguez, and Jeffrey S Duchin. An automated system for public health surveillance of school absenteeism. *Journal of Public Health Management and Practice*, 17(1):59–64, 2011.

[2] Benjamin E Berkman. Mitigating pandemic influenza: the ethics of implementing a school closure policy. *Journal of Public Health Management and Practice*, 14(4):372–378, 2008.

[3] Simon Cauchemez, Neil M Ferguson, Claude Wachtel, Anders Tegnell, Guillaume Saour, Ben Duncan, and Angus Nicoll. Closure of schools during an influenza pandemic. *The Lancet infectious diseases*, 9(8):473–481, 2009.

[4] Irene Chen, Fredrik D Johansson, and David Sontag. Why is my classifier discriminatory? In *Advances in Neural Information Processing Systems*, pages 3539–3550, 2018.

[5] Jennifer Chu-Carroll, James Fan, BK Boguraev, David Carmel, Dafna Sheinwald, and Chris Welty. Finding needles in the haystack: Search and candidate generation. *IBM Journal of Research and Development*, 56(3.4):6–1, 2012.

[6] Benjamin J Cowling, Eric HY Lau, Conrad LH Lam, Calvin KY Cheng, Jana Kovar, Kwok Hung Chan, JS Malik Peiris, and Gabriel M Leung. Effects of school closures, 2008 winter influenza season, hong kong. *Emerging infectious diseases*, 14(10):1660, 2008.

[7] Alessandro Del Sole. Introducing microsoft cognitive services. In *Microsoft Computer Vision APIs Distilled*, pages 1–4. Springer, 2018.

[8] Paul V Effler, Dale Carcione, Carolien Giele, Gary K Dowse, Leigh Goggin, and Donna B Mak. Household responses to pandemic (h1n1) 2009–related school closures, perth, western australia. *Emerging infectious diseases*, 16(2):205, 2010.

[9] Andrea Frome, Howard Wellington Trickey, Melanie Clements, Ethan G Russell, Paul Eastlund, Diego Ariel Gertzenstein, Douglas Richard Grundman, and Baris Yuksel. Systems and methods of correlating business information to determine spam, closed businesses, and ranking signals, November 24 2015. US Patent 9,195,987.

[10] Thomas House, Marc Baguelin, Albert Jan Van Hoek, Peter J White, Zia Sadique, Ken Eames, Jonathan M Read, Niel Hens, Alessia Melegaro, W John Edmunds, et al. Modelling the impact of local reactive school closures on critical care provision during an influenza pandemic. *Proceedings of the Royal Society B: Biological Sciences*, 278(1719):2753–2760, 2011.

[11] Charlotte Jackson, Punam Mangtani, Jeremy Hawker, Babatunde Olowokure, and Emilia Vynnycky. The effects of school closures on influenza outbreaks and pandemics: systematic review of simulation studies. *PloS one*, 9(5), 2014.

[12] Charlotte Jackson, Emilia Vynnycky, Jeremy Hawker, Babatunde Olowokure, and Punam Mangtani. School closures and influenza: systematic review of epidemiological studies. *BMJ open*, 3(2):e002149, 2013.

[13] Shoko Kawano and Masayuki Kakehashi. Substantial impact of school closure on the transmission dynamics during the pandemic flu h1n1-2009 in oita, japan. *PloS one*, 10(12), 2015.

[14] Tibor Kiss and Jan Strunk. Unsupervised multilingual sentence boundary detection. *Computational linguistics*, 32(4):485–525, 2006.

[15] Stephen M Kissler, Christine Tedijanto, Marc Lipsitch, and Yonatan Grad. Social distancing strategies for curbing the covid-19 epidemic. *medRxiv*, 2020.

[16] Lisa M Koonin and Martin S Cetron. School closure to reduce influenza transmission. *Emerging infectious diseases*, 15(1):137, 2009.

[17] Adam Leader. Detecting closure of a business, April 9 2015. US Patent App. 13/856,145.

[18] Bruce Y Lee, Shawn T Brown, Philip Cooley, Maggie A Potter, William D Wheaton, Ronald E Voorhees, Samuel Stebbins, John J Grefenstette, Shanta M Zimmer, Richard Zimmerman, et al. Simulating school closure strategies to mitigate an influenza epidemic. *Journal of public health management and practice: JPHMP*, 16(3):252, 2010.

[19] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[20] Edward Loper and Steven Bird. Nltk: the natural language toolkit. *arXiv preprint cs/0205028*, 2002.

[21] Pamela Mann, Erin O'Connell, Guoyan Zhang, Anthoni Llau, Edhelene Rico, and Fermin C Leguen. Alert system to detect possible school-based outbreaks of influenza-like illness. *Emerging infectious diseases*, 17(2):262, 2011.

[22] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. *arXiv preprint arXiv:1908.09635*, 2019.

[23] Hermann Ney, Ute Essen, and Reinhard Kneser. On structuring probabilistic dependences in stochastic language modelling. *Computer Speech & Language*, 8(1):1–38, 1994.

[24] Carla V Rodriguez, Krista Rietberg, Atar Baer, Tao Kwan-Gett, and Jeffrey Duchin. Association between school closure and subsequent absenteeism during a seasonal influenza epidemic. *Epidemiology*, pages 787–792, 2009.

[25] Md Z Sadique, Elisabeth J Adams, and William J Edmunds. Estimating the costs of school closure for mitigating an influenza pandemic. *BMC public health*, 8(1):135, 2008.

[26] Asami Sasaki, Anne Gatewood Hoen, Al Ozonoff, Hiroshi Suzuki, Naohito Tanabe, Nao Seki, Reiko Saito, and John S Brownstein. Evidence-based tool for triggering school closures during influenza outbreaks, japan. *Emerging infectious diseases*, 15(11):1841, 2009.

[27] Jakob Truelsen. wkhtmltopdf–convert html to pdf using webkit (qtwebkit), 2017.

[28] Russell M Viner, Simon J Russell, Helen Croker, Jessica Packer, Joseph Ward, Claire Stansfield, Oliver Mytton, Chris Bonell, and Robert Booy. School closure and management practices during coronavirus outbreaks including covid-19: a rapid systematic review. *The Lancet Child & Adolescent Health*, 2020.

[29] E Vynnycky and WJ Edmunds. Analyses of the 1957 (asian) influenza pandemic in the united kingdom and the impact of school closures. *Epidemiology & Infection*, 136(2):166–179, 2008.