Accident Analysis and Prevention 89 (2016) 111–117

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Computerized "Learn-As-You-Go" classification of traumatic brain injuries using NEISS narrative data

Wei Chen [a], Krista K. Wheeler [b,c], Simon Lin [a], Yungui Huang [a], Huiyun Xiang [b,c,*]

[a] Research Information Solutions and Innovation, The Research Institute at Nationwide Children's Hospital, Columbus, OH, USA
[b] Center for Injury Research and Policy, The Research Institute at Nationwide Children's Hospital, Columbus, OH, USA
[c] Center for Pediatric Trauma Research, Nationwide Children's Hospital, Columbus, OH, USA

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form 8 January 2016
Accepted 21 January 2016
Available online 3 February 2016

Keywords:
Machine learning
Sports injury classification
Computerized classification
Traumatic brain injuries
Unstructured NEISS narratives

ABSTRACT

One important routine task in injury research is to effectively classify injury circumstances into user-defined categories when using narrative text. However, traditional manual processes can be time consuming, and existing batch learning systems can be difficult to utilize by novice users. This study evaluates a "Learn-As-You-Go" machine-learning program. When using this program, the user trains classification models and interactively checks on accuracy until a desired threshold is reached. We examined the narrative text of traumatic brain injuries (TBIs) in the National Electronic Injury Surveillance System (NEISS) and classified TBIs into sport and non-sport categories. Our results suggest that the DUALIST "Learn-As-You-Go" program, which features a user-friendly online interface, is effective in injury narrative classification. In our study, the time frame to classify tens of thousands of narratives was reduced from a few days to minutes after approximately sixty minutes of training.

1. Introduction

The National Electronic Injury Surveillance System (NEISS) dataset consists of a wealth of coded variables (e.g., product, diagnosis, body part) that are important for studying the causes and circumstances of injuries (CPSC, 2015). The NEISS is a national probability sample of hospitals in the U.S. and its territories; it is produced by the Consumer Product Safety Commission to collect information about both consumer product related and other injuries. It is a nonfatal injury data source for the Centers for Disease Control's Web-based Injury Statistics Query and Reporting System. However, the NEISS coded variables do not always allow for the desired specificity, and so the narrative text fields in NEISS are often used to create new variables.

The NEISS narrative text has been used in a variety of ways in recent injury studies including those examining helmet use (Graves et al., 2015), magnets as foreign bodies (Silverman et al., 2013), battery exposures (Sharpe et al., 2012), as well as shopping cart (Martin et al., 2014) and TV-related injuries (De Roo et al., 2013). In most of these studies, manual review processes combined with repeated key word searches were the primary methods used. Despite the flexibility of manual review, this process may be subject to non-standardized procedures and low efficiency; therefore, it can be both error-prone and expensive to repeat (McKenzie et al., 2010; Wellman et al., 2004). Computerized methods are needed to complement, if not replace, this manual work.

Efforts have been made in developing computerized machine learning approaches to classify injury narratives. In a systematic review of that literature, Bayesian probability based methods were found to be the most commonly used (Vallmuur, 2015) in injury research due to its effectiveness and efficiency of classifying text data. Bayesian methods have been successfully applied to various injury classification scenarios including motor vehicle injuries classification (Lehto and Sorock, 1996), claims classification (Bertke et al., 2012) and near-miss causes classification (Taylor et al., 2014) with state-of-the-art accuracy achieved.

However, most of the cited studies used approaches that required labeling significant portions of the data sets when training classifiers. Additionally, the training sets described were collected in batch mode rather than through iterations. Large amounts of training data are sometimes hard to obtain upfront especially in the medical domain (Chen et al., 2015). Generally, the accuracy of a batch learning system improves with increases in the size of the training set (Taylor et al., 2014); however, producing a large training set is time-consuming, and deciding when to stop collecting the training data can be very difficult. Also, traditional batch learning

* Corresponding author at: Center for Injury Research and Policy, The Research Institute at Nationwide Children's Hospital, Columbus, OH, USA.
Fax: +1 614 355 5897.
E-mail address: huiyun.xiang@nationwidechildrens.org (H. Xiang).

http://dx.doi.org/10.1016/j.aap.2016.01.012
0001-4575/

systems were not designed for novice users and therefore require substantial computer science knowledge to achieve the best result.

To make machine-learning programs accessible to novice users, online learning systems (sometimes called active learning) have been developed (Settles, 2012a). These online learning systems are typically characterized by user-friendly web interfaces, and they adopt "learn as you go" methods to iteratively train a classifier. Training or labeling large amounts of data is then less necessary than when using most batch learning systems (Fontenla-Romero et al., 2013). These online learning systems have been recently used in word sense disambiguation (Chen et al., 2013), social media and web content mining (Settles, 2011) and ambient weave computing applications (Schmitt et al., 2008). They have not yet been used frequently in healthcare domains, such as injury research.

In this study, we evaluated an interactive learning program called DUALIST. The program offers a simple web-based interface, and it can be installed on any operating system. It has been previously tested against social network and news data. Our goal was to evaluate this program for use in injury research, and we chose sports injury classification as an example.

2. Methods

The online learning program chosen, DUALIST, offers the combined benefits of interactive machine learning and ease of use (Settles, 2012b). DUALIST, developed by Burr Settles from Carnegie Mellon University, is a result of recent advances in artificial intelligence (Settles, 2011). The DUALIST program is freely available and can be easily installed on any platform (Settles, 2012b). Using DUALIST, users can define a number of classes and labels for each class. Then, they interactively train classifiers through iterations. Each iteration generates a classifier. Multiple classifiers can be generated in one session. Different from ensample learning, classifiers generated in one session are related so later classifiers build upon the previous ones.

2.1. Multinomial naïve Bayes (MNB) model

Each iteration is driven by an underlying probabilistic model called a multinomial naïve Bayes (MNB) model which emulates the human decision making process. When performing a manual review process of injury classification, one relies on lexical clues such as keywords and keyword relations to classify sentences. For example, "fell when playing softball" is considered sports related but "jump in bed" is not, and *fell, playing softball, jump* and *bed* are all key words that contribute to the probability of an injury case belonging to one class or another.

The DUALIST program implemented a specific multinomial naïve Bayes (MNB) model which is known to be "simple, fast, and [one that] work[s] well for several natural language applications."(Muscatello et al., 2005). Naïve Bayes machine learning has also been found to be effective in recent injury classification tasks based on batch learning (Wellman et al., 2004; Lehto et al., 2009). The dualist program implemented both a unigram model (i.e., single word) and a bigram model (i.e., a phrase including two consecutive words) to calculate the probabilities of words and phrases of a sentence. Given the "naïve" assumption, the words of the injury narratives in our cases, are considered to be independent from each other. Each narrative is also considered a mixture of classes, i.e., sports and non-sports in our study. The probability of a word x generated by class y_i can be defined as

$$P_\theta(x|\, y_i) = P(|x|) \prod_k \left(\theta_{jk}\right)^{f_k(x)}$$

where $\theta_{jk} = P(f_k|\, y_i)$ is the probability of generating word f_k given class y_i, and $f_k(x)$ is the frequency count of word f_k in narrative x. The $|x|$ is the document length which is independent of the class, so $P(|x|)$ is a constant. After we dropped $P(|x|)$, we applied the Bayes' rule to the above equation. The posterior probability of a sequence of words belong to a class $P_\theta(y_i|\, x)$ can be defined as

$$P_\theta(y_i|\, x) = \frac{P_\theta(y_i) P_\theta(x|y_i)}{P_\theta(x)}$$

$$\sum_{i \in (1,n)} P_\theta(y_i|\, x) = 1$$

where $P_\theta(y_i)$ is the probability of class y_i and n is the number of classes. The class label is assigned to the narrative based on the largest $P_\theta(y_i|x)$. For example, if the P_θ(Sports|x) is 0.7 and P_θ(Nonsports|x) is 0.3, the instance will be assigned to Sports. The original research paper of Settles provides additional details (Settles, 2011).

2.2. Data

The NEISS dataset is a deidentified dataset freely available on the website of the Consumer Product Safety Commission. Because the data is deidentified and individuals cannot be readily identified with the data, it is not considered human subjects research by our institution. No Institutional Review Board review or an exemption was needed.

We used a NEISS TBI definition recommended in previous research (Xiang et al., 2004; Thompson et al., 2014). NEISS utilizes its own unique coding scheme which includes both body part codes and diagnosis codes. The TBI definition used here includes concussions, fractures to the head, and internal organ injury to the head. We chose to focus on TBIs, because of the research team's prior TBI related research. We obtained 38,933 TBI cases from the 2009 NEISS dataset.

We did not perform specific text preprocessing tasks, such as abbreviation expansion and misspelling corrections, which might be of greater necessity when using other datasets. The medical abbreviations used in our data were fairly standard and most of the class-related features (words or phrases) were spelled out. We did combine the two NEISS text fields to create one narrative sentence for each NEISS record. Each narrative sentence is described below as an *instance* in the language of computer science.

2.3. Case definition

In this study, physical activities that are both competitive and recreational were considered sports (see examples in Table 1). Leisure activities such as *gardening* and *walking* were not sports. This definition has been previously used in the literature, see the referenced article for greater detail (Conn et al., 2003).

We prepared a random subset of 500 cases, and each record was labeled sport or non-sport. These 500 test cases were an ad hoc held-out set used for the purpose of evaluation only. This is not a necessary step when an end user uses the system.

Table 1
Sports categories.

Sports category	Examples
Competitive sports	Basketball, cycling, exercising, football, baseball/softball, soccer, ice skating/skateboarding/snowboarding, gymnastics/cheerleading, playground equipment, etc.
Recreational sports	Bowling, fishing, hiking, biking, jogging, horse riding, all-terrain vehicle riding.

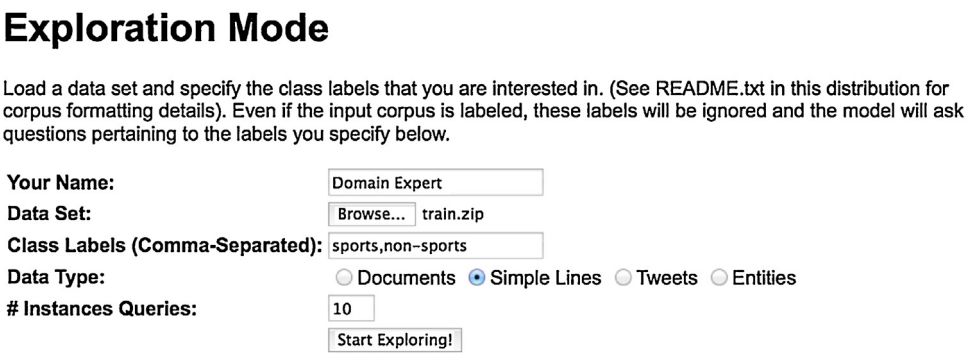

Fig. 1. DUALIST online learning experiment configuration.

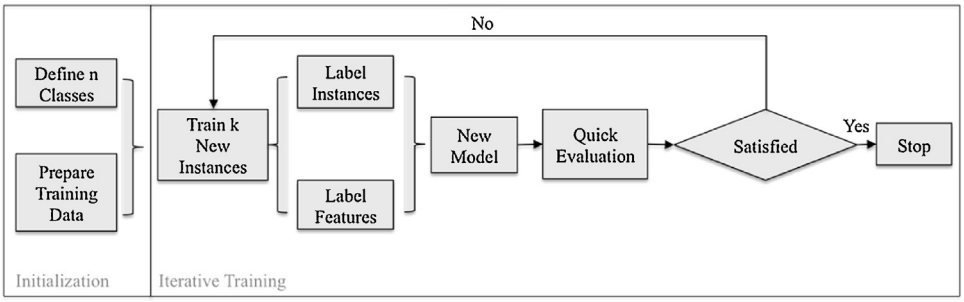

Fig. 2. Training workflow.

We first experimented with the binary classification task of sport and non-sport injuries to introduce active learning to injury researchers using a simple example and to provide a full evaluation of its effectiveness through detailed experiments. We then used the program for a multi-level classification task using the same dataset. Though this task can be potentially more challenging, it is still doable, with unbalanced input data (Settles, 2012a). In the multi-level class experiment, our rater assigned each injury narrative to one of the three classes: sport injury involving only one person (the injured subject), sport injury involving more than one person (e.g., injured subject collides with a team player), and non-sport injury. Our rater went through 500 randomly selected training cases to generate the classification model, which was evaluated against a held-out test set including 500 randomly selected new cases.

2.4. DUALIST workflow

The DUALIST program has a user friendly initialization interface. We configured the software for the binary experiment as shown in Fig. 1. We chose the class labels (sports, non-sports). For each iteration, a fixed but configurable number of new instances can be trained, and we chose 10 instances per iteration.

After this initial configuration, the user enters the classifier training phase. DUALIST, as implied by its name, implements a dual mode of classifier training. A user can label either the instances (i.e., the complete narrative sentence for each NEISS record) or the features (i.e., words and phrases extracted from the sentence), or both. The training process is carried out through iterations. After each iteration, a new classification model is generated. The performance of each classification model can then be evaluated against the test dataset of labeled cases. When the performance is deemed satisfactory, the trained computerized classifier can be applied to the remaining unlabeled cases. Fig. 2 shows the workflow of training a classifier.

Fig. 3 shows the interactive and iterative training interface. A *Submit* button on the top saves the current model and then populates the container with a new set of instances. After a model is generated, the user can use a command provided by the DUALIST program to apply the model to the test set and evaluate the test's accuracy. At any point, a user can choose to close the program because of satisfaction with accuracy of the classification model.

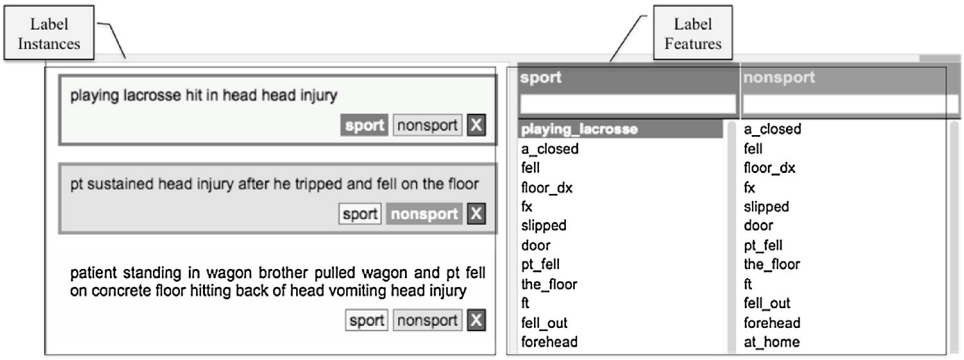

Fig. 3. DUALIST interactive and iterative training interface.

We chose to complete 50 iterations, 10 instances each before ending the training process.

2.5. Experimental design

As described above, a random subset of 500 cases were labeled as a test set and 38,433 unlabeled cases remained for "Learn-As-You-Go" training. The label of each narrative in the test set was based on inter-user agreement between the two local domain experts (injury researchers). The overall agreement rate was over 90%. For those ambiguous cases not agreed upon between the injury experts, a third project oversight person decided the label.

Our domain experts used the DUALIST program to train sport injury classifiers interactively through multiple iterations with 10 cases trained during each iteration. The online learning program does not require users to go through all 38,433 training cases but allows the user to stop at any point s/he is satisfied with the classification accuracy. The classification accuracy can be checked at any point by running a command line against the defined test set provided by the DUALIST program.

After training 500 instances, 50 iterations with 10 instances each, the user stopped when achieving accuracy of 90%. DUALIST produces a classification model with each iteration; thus, the user created 50 classifiers which contributed to the final model during this training process. We use the notation *train* [*n*] to denote the last classifier generated by training the first *n* instances to date. Previous research has shown that the dual mode—annotating both instances (the full narrative text) and features (keywords or phrases within the narrative text) —is the most effective (Settles, 2011), and this was our approach.

2.6. Evaluation

In order to find out at what point the program started to perform well, each trained classifier was applied to the test data set and then evaluated. True positive rates (TPR, percentage of correctly classified sports injuries of all sports injuries) and false positive rates (FPR, percentage of incorrectly classified non-sport injuries of all non-sport injuries) were calculated. ROC and AUC were used as metrics to measure the performance of the binary classifier.

For each of the 50 classifiers, we varied the probability threshold of the label sports from 0 to 1 by an increment of 0.01 in deciding the predicted label, resulting in 100 (TPR, FPR) data points of an ROC of a classifier. Based on TPR and FPR, ROCs can be drawn as FPR vs. TPR. The AUC of each ROC was calculated by estimating an accumulated area under the ROC curve. In the literature, a model with AUC between 0.8 and 0.9 is generally considered good, and one with AUC between 0.9 and 1.0 is considered excellent (Metz, 1978).

The TBI dataset was unbalanced with respect to the number of *sports* vs. *non-sports* cases by an approximate 2:1 ratio. The ROC and AUC are considered stable measures regardless of the balancing of the data (Fawcett, 2004). The last classifier obtained was applied later to classify the entire dataset. The total time of automated classification was recorded to contrast with the manual process. We also plotted the percentage of correctness as a measure of accessing the accuracy of each classifier using 0.5 as the threshold for assigning sports class.

3. Results

It took our domain expert(s) approximately 1 h to train the first 500 instances in the training set before s/he stopped. The randomly selected 500 TBI test cases were about 2:1, sports versus non-sports related. The last model created 4291 features among which 169 were user labeled features. Features were mainly noun or noun phrases such as *football*, *football_player* and *bike*, as well as verb phrases such as *playing_football* and *riding_bike*.

The user experiment results demonstrated that our overall model accuracy consistently improved with more samples being trained through the iterative online learning process. This is demonstrated in the converging trend of the ROCs of classifiers at different training stages, as shown in Fig. 4. However, the biggest improvement was found to be between the first 50 and the first 200 instances. The AUC for train50, train100, train200, train400 and train500 were 0.76, 0.79, 0.88, 0.93, and 0.95 respectively.

As indicated by the AUCs, our injury classifiers were found to start performing well after training with as few as the first 400 cases in the training set. The TPR quickly climbed to above 0.85 within a narrow range of FPR of 0 to 0.15 after training with 200 instances. Overall, the ROC curves showed that the DUALIST program was working effectively across different classification thresholds chosen.

In Fig. 5, we plotted the percentage of correctness for all 50 classification models, all with the default threshold of 0.5. It was shown that the accuracy quickly converged to near 0.9 after as few as 20 iterations, which is equivalent to 24 min of manual labeling of 200 instances. Using the best classifier model obtained, the last one, it took the computer about 3 min to classify the remaining 38,433 unlabeled instances. By contrast, this same task would take a person approximately 78 dedicated hours to finish. This 78 h time estimate is based upon the number of TBI cases in our database and the fact that it took raters an average of 24 min to label 200 instances, but it does not take into account the use of key word searches or other time saving methods. Therefore; we achieved about 1560 times efficiency improvement by using a computerized classifier trained by a novice user using the DUALIST program.

In a separate experiment, evaluating our multi-level classification task, we found that our interactively constructed multi-level classifier achieved a level of accuracy comparable to that seen in the binary classification task: 84% accuracy, and 0.90 overall AUC. The overall AUC was calculated as a weighted average of the AUC of three classes based on class distributions. The AUC of "sport injury with only one person involved" class was the lowest, AUC = 0.87. The AUC of "sport injury with more than one person involved" class was the highest, AUC = 0.96. The AUC of "non-sport class" was 0.89.

4. Discussion

In this study, we evaluated a computerized online learning program for the binary classification task of sports vs. nonsports related TBIs. Our efforts were among the first to use and evaluate an online learning program like DUALIST in a real world application for injury research. We assessed its effectiveness and performance using the ROC, AUC and accuracy measures. Our experiment generated good classifier models after training as few as 500 narratives by a novice user.

Our binary classifier accuracy outperformed previous experiments using the same DUALIST system: 90% accuracy (ours) vs. 70% accuracy (average accuracy in the other research) (Settles and Zhu, 2016). Our performance improvement may be due to several reasons. First, our raters were trained injury researchers who were knowledgeable in the subject matter to ensure consistent labeling. Second, our classification task was simpler than the task of sentiment classification in the other study. However, it should be noted that the previous experiment used a different narrative dataset and was conducted in a multi-category situation (Wellman et al., 2004).

In our separate multi-level experiment, we achieved comparable accuracy in a classification task involving three classes: 84% accuracy and 0.90 AUC overall. In sum, we found the DUALIST program capable of producing state-of-the-art accuracy in both

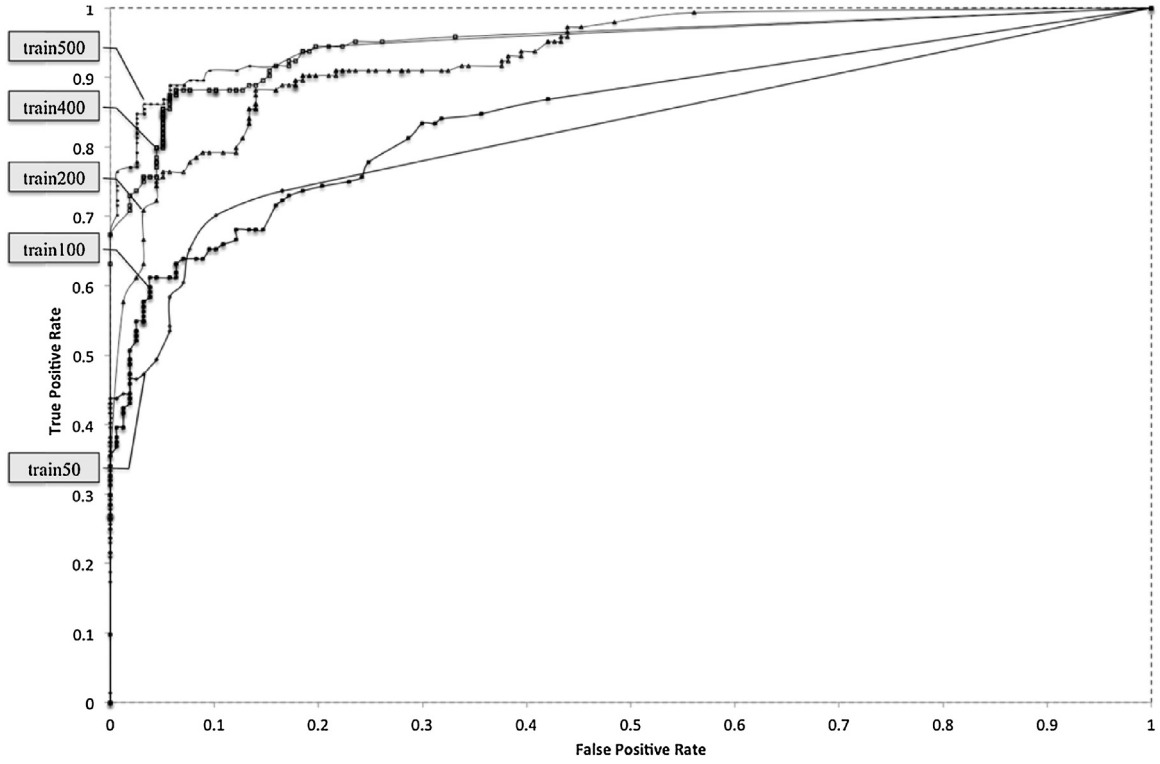

Fig. 4. Improvement of classification accuracy through iterative training.

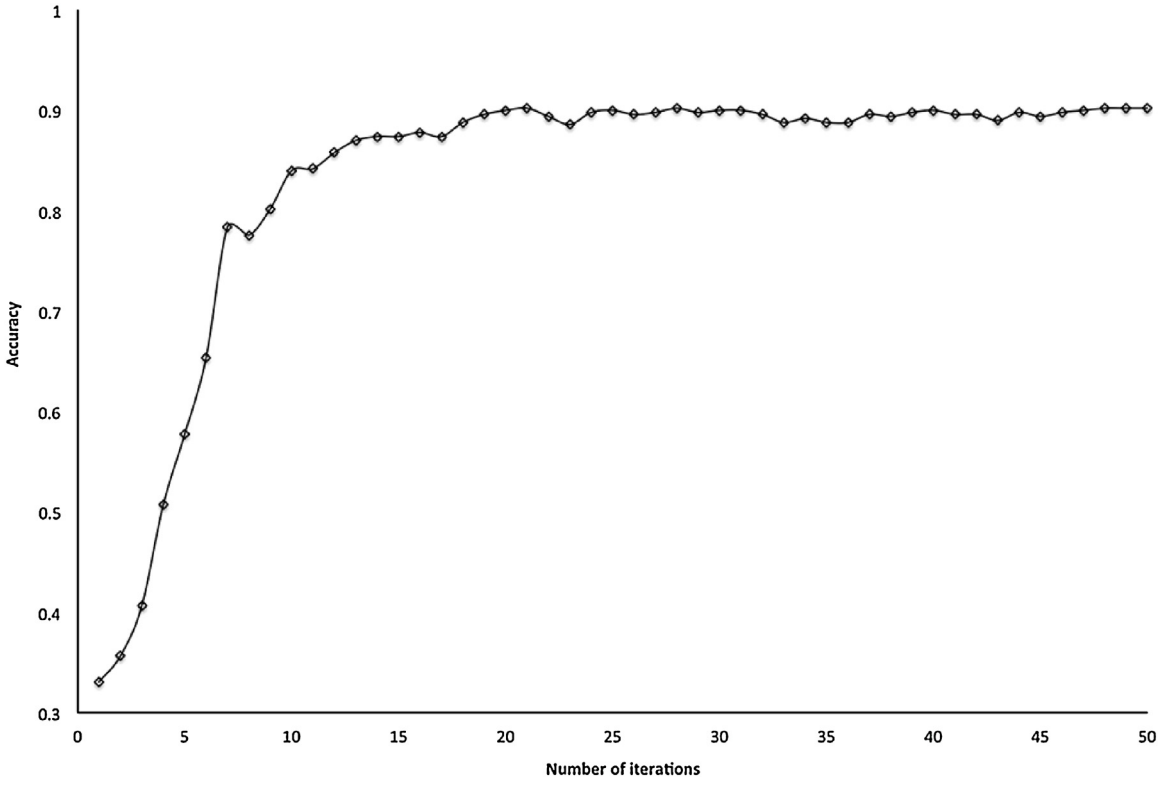

Fig. 5. Classification accuracy improvement through iterative training.

binary and multi-class situations requiring a minimum of effort from novice researchers. Our conclusions on the usability and performance of the DUALIST program were consistent with previous research (Settles, 2011; Settles, 2012b).

5. Limitations

The DUALIST program was designed to be used by one user at a time. We have not done experiments on classification tasks that

involve user collaborations. Additionally, the sport and non-sport classification schema is relatively simple. Other applications may have more complicated classification scenarios. In those situations, one may have to go through more narratives during the interactive training phase in order to have a good coverage of the samples from each class. While our classification accuracy quickly converged to a satisfactory level after training as few as 200 cases, this may not true for multi-class applications, or with different topics, or when dealing with unbalanced datasets (Chen et al., 2011; Sahare and Gupta, 2012). We also have not yet experimented with classification using a more heterogeneous narrative dataset. Our input data were pre-filtered TBI injuries. Given a general dataset including all types of injuries, the signal in terms of sports injury key words may not be as strong because sports injury would be a much smaller injury mechanism category in the entire injury database. Therefore, for classification tasks involving a specific subcategory of injury, we recommend first sub-setting the data to create more a focused input set.

Finally, similar to other Bayes-based machine learning systems, the DUALIST program classifies cases by keywords rather than context and semantics which could be more useful but latent in the text. For example, if an object that caused the injury was from a previous sports activity, it is possible for the DUALIST program to still classify it as a sports-injury rather than as an existing injury.

6. Conclusions

To our knowledge, the DUALIST program is the first readily available and free online learning program for novice users with machine learning-based classification needs. It has been previously shown to effectively classify various types of corpus including news, twitter data, and reviews among others; it has worked well across application domains (Settles, 2012a; Settles, 2012b; Settles and Zhu, 2016). Our effort is among the first to test it with novice users from the injury domain as part of a larger effort to promote active learning systems.

Based on our evaluation, this program was found to be easy and flexible to use for first time users. It provided an interactive interface that was not available in traditional machine learning systems. One trains as he/she goes and can flexibly stop at any time point. The ability to "train-as-you-go" using an interactive learning program like DUALIST substantially lowers the technical threshold for users new to a machine learning system.

The DUALIST program is able to save labor cost dramatically without compromising the accuracy. In our case, to generate a good classifier required no more than an hour of manual work. The interactively generated reusable classifier could dramatically improve the efficiency of classification tasks which have previously been done mainly through manual means.

Competing interests

There are no competing interests.

Contributors

This work was carried out in collaboration between all authors. Authors HX and SML designed the study and obtained funding. Author WC participated in the study design, performed the analysis, and wrote the first draft of the manuscript. Author SML and YH supervised the analyses of the study. Author KW experimented with the system. All authors contributed to the writing and analysis. All authors read and approved the final manuscript.

What is already known on the subject

- Injury classification can be a labor-intensive task in injury research.
- Traditional batch learning systems, which are designed for classification purposes, are not accessible to novice users.

What this study adds

- The "Learn-As-You-Go" machine learning program DUALIST was evaluated by novice users to classify sports injuries using National Electronic Injury Surveillance System (NEISS) data.
- The quantitative evaluation of the program using machine learning-based measures suggested that the effectiveness of this program achieved near-the-state-of-art accuracy.
- The efficiency of classification was improved by nearly a thousand fold using a computerized automatic classifier when compared to traditional manual methods.

Acknowledgments

The project described was supported by the National Center for Advancing Translational Sciences, National Institutes of Health (PI: Rebecca Jackson, Grant #: UL1TR001070), Agency for Healthcare Research and Quality (PI: Huiyun Xiang, Grant #:1R03HS022277), and Centers for Disease Control and Prevention (PI: Huiyun Xiang, Grant #: 1R49CE002106). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, AHRQ, and CDC.

We thank Dr. Cheng Chen at the Center for Injury Research and Policy, The Research Institute at Nationwide Children's Hospital for coding the multi-level classification data. We also thank Dr. Junxin Shi and Kathryn Coxe for participation in the initial testing of DUALIST system.

References

Bertke, S., Meyers, A., Wurzelbacher, S., Bell, J., Lampl, M., Robins, D., 2012. Development and evaluation of a Naïve Bayesian model for coding causation of workers' compensation claims. J. Safety Res. 43 (5), 327–332.

CPSC. National Electronic Injury Surveillance System (NEISS), 2015. Available from:http://www.cpsc.gov/en/Research–Statistics/NEISS-Injury-Data/.

Chen, E., Lin, Y., Xiong, H., Luo, Q., Ma, H., 2011. Exploiting probabilistic topic models to improve text categorization under class imbalance. Infor. Process. Manag. 47 (2), 202–214.

Chen, Y., Cao, H., Mei, Q., Zheng, K., Xu, H., 2013. Applying active learning to supervised word sense disambiguation in MEDLINE. J. Am. Med. Inform. Assoc., amiajnl-2012-001244.

Chen, W., Kowatch, R., Lin, S., Splaingard, M., Huang, Y., 2015. Interactive cohort identification of sleep disorder patients using natural language processing and i2b2. Appl. Clin. Inf. 6 (2), 345–363.

Conn, J., Annest, J.L., Gilchrist, J., 2003. Sports and recreation related injury episodes in the US population, 1997–99. Inj. Prev. 9 (2), 117–123.

De Roo, A.C., Chounthirath, T., Smith, G.A., 2013. Television-related injuries to children in the United States, 1990–2011. Pediatrics 132 (2), 267–274.

Fawcett, T., 2004. ROC graphs: notes and practical considerations for researchers. Mach. Learn. 31, 1–38.

Fontenla-Romero, Ó., Guijarro-Berdiñas, B., Martinez-Rego, D., Pérez-Sánchez, B., Peteiro-Barral, D., 2013. Online machine learning. Effic. Scalability Methods Comput. Intellect, 27.

Graves, J.M., Whitehill, J.M., Hagel, B.E., Rivara, F.P., 2015. Making the most of injury surveillance data: using narrative text to identify exposure information in case-control studies. Injury 46 (5), 891–897.

Lehto, M.R., Sorock, G.S., 1996. Machine learning of motor vehicle accident categories from narrative data. Methods Inf. Med. 35 (4–5), 309–316.

Lehto, M., Marucci-Wellman, H., Corns, H., 2009. Bayesian methods: a useful tool for classifying injury narratives into cause groups. Inj. Prev. 15 (4), 259–265.

Martin, K.J., Chounthirath, T., Xiang, H., Smith, G.A., 2014. Pediatric shopping-cart-related injuries treated in US emergency departments, 1990–2011. Clin. Pediatrics 53 (3), 277–285.

McKenzie, K., Scott, D.A., Campbell, M.A., McClure, R.J., 2010. The use of narrative text for injury surveillance research: a systematic review. Accident Anal. Prev. 42 (2), 354–363.

Metz, C.E., 1978. Basic principles of ROC analysis. Seminars in Nuclear Medicine 8 (4), 283–298, Epub 1978/10/01.
Muscatello, D.J., Churches, T., Kaldor, J., Zheng, W., Chiu, C., Correll, P., et al., 2005. An automated, broad-based, near real-time public health surveillance system using presentations to hospital Emergency Departments in New South Wales, Australia. BMC Public Health 5 (1), 141.
Sahare, M., Gupta, H., 2012. A review of multi-class classification for imbalanced data. Int. J. Adv. Comput. Res. 2 (3), 160–164.
Schmitt, J., Hollick, M., Roos, C., Steinmetz, R., 2008. Adapting the user context in realtime: tailoring online machine learning algorithms to ambient computing. Mobile Netw. Appl. 13 (6), 583–598.
Settles, B., Zhu, X., 2012. Behavioral factors in interactive training of text classifiers. Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 563–567.
Settles, B., 2011. Closing the loop: Fast, interactive semi-supervised annotation with queries on features and instances. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 1467–1478.
Settles, B., 2012a. Active learning. Synth. Lect. Artif. Intell. Mach. Learn. 6 (1), 1–114.
Settles B., 2012. Dualist Interactive machine learning for text analysis. Available from: https://github.com/burrsettles/dualist.
Sharpe, S.J., Rochette, L.M., Smith, G.A., 2012. Pediatric battery-related emergency department visits in the United States, 1990–2009. Pediatrics 129 (6), 1111–1117.
Silverman, J.A., Brown, J.C., Willis, M.M., Ebel, B.E., 2013. Increase in pediatric magnet-related foreign bodies requiring emergency care. Ann. Emerg. Med. 62 (6), 604–608, e1.
Taylor, J.A., Lacovara, A.V., Smith, G.S., Pandian, R., Lehto, M., 2014. Near-miss narratives from the fire service: A Bayesian analysis. Accident Anal. Prev. 62, 119–129.
Thompson, M.C., Wheeler, K.K., Shi, J., Smith, G.A., Groner, J.I., Haley, K.J., et al., 2014. Surveillance of paediatric traumatic brain injuries using the NEISS: choosing an appropriate case definition. Brain Inj. 28 (4), 431–437.
Vallmuur, K., 2015. Machine learning approaches to analysing textual injury surveillance data: a systematic review. Accident Anal. Prev. 79, 41–49.
Wellman, H.M., Lehto, M.R., Sorock, G.S., Smith, G.S., 2004. Computerized coding of injury narrative data from the National Health Interview Survey. Accident Anal. Prev. 36 (2), 165–171.
Xiang, H., Stallones, L., Smith, G.A., 2004. Downhill skiing injury fatalities among children. Inj Prev. 10 (2), 99–102.