Co-training over Domain-independent and Domain-dependent Features for Sentiment Analysis of an Online Cancer Support Community

Prakhar Biyani[1], Cornelia Caragea[2], Prasenjit Mitra[1], Chong Zhou[1], John Yen[1], Greta E. Greer[3], Kenneth Portier[3]

[1]College of Information Sciences and Technology, The Pennsylvania State University

[2] Department of Computer Science and Engineering, University of North Texas

[3] American Cancer Society, Inc., University of North Texas

Email: pxb5080@ist.psu.edu, ccaragea@unt.edu, {pmitra, cwz5091, jyen}@ist.psu,edu, {greta.greer, kenneth.portier}@cancer.org

Abstract—Sentiment analysis has been widely researched in the domain of online review sites with the aim of getting summarized opinions of product users about different aspects of the products. However, there has been little work focusing on identifying the polarity of sentiments expressed by users in online health communities such as cancer support forums, etc. Online health communities act as a medium through which people share their health concerns with fellow members of the community and get social support. Identifying sentiments expressed by members in a health community can be helpful in understanding dynamics of the community such as dominant health issues, emotional impacts of interactions on members, etc. In this work, we perform sentiment classification of user posts in an online cancer support community (Cancer Survivors Network). We use Domain-dependent and Domain-independent sentiment features as the two complementary views of a post and use them for post classification in a semi-supervised setting using the co-training algorithm. Experimental results demonstrate effectiveness of our methods.

Keywords—Sentiment analysis, co-training, indirect emotional support, direct emotional support, online health community.

I. INTRODUCTION

Many online health communities (OHCs) exist today on general health issues (e.g., SPARKPEOPLE[1] and Everyday Health[2]) as well as on specific diseases such as cancer and AIDS (e.g., Cancer Survivors Network (CSN)[3] and WebMD[4]). In these communities, millions of people discuss their health-related issues, ask questions about their ailments, the symptoms they experience, medications, side-effects, and share their health concerns to get emotional support [1], [2]. Although it may seem odd at first to share and discuss such important issues with unknown people, or to consult health-related websites, there is substantial value in doing so. For example, one can learn what to expect in the context of a disease by aggregating the views of people who went through the exact same symptoms, and thus, being able to deal with the disease in a better way. Previous research has shown that people feel much better and change to positive attitudes if they talk to other people after or during a traumatic event such as a disease [3].

The online sharing of views on health-related issues has resulted into the generation of a huge amount of user-generated data in the form of discussions in the archives of health communities. Analyzing these conversational data can help the network managers in obtaining crucial information about the dynamics of these communities, e.g., identifying dominant health issues in the community, effects of interactions on emotional states of individual members, etc. Understanding emotional impacts of online participation on members can help in improving the features of the community portals to enable facilitation of emotional support to the network members more effectively.

In this paper, we analyze the sentiment of user messages (or posts) of an online cancer support community, the Cancer Survivors' Network (CSN) of the American Cancer Society. We identify whether a post is *positive* or *negative* based on the polarity of the emotion expressed in it and model the task of sentiment analysis as a binary classification problem.

A. Why Sentiment Analysis of CSN Posts?

Sentiment analysis of CSN posts can help to understand the dynamics of the network. In CSN, members interact to get social support of which emotional support is an important and major part. Often, members post messages seeking emotional support when they feel depressed, sad about their health conditions, cancer reports, family problems, etc. A post is indicative of the emotional state of a poster. Analyzing sentiment of posts of a member over a period of time can be helpful in seeing the emotional effects of interactions on that member. A member may show a positive change in sentiment by getting replies from other members. A negative change of sentiment may be indicative of not getting replies, getting discouraging replies from some members, or some other factors. Also, there are certain members in every community who influence the community in a positive or a negative way. In the context of CSN, an important influence can be in terms of sentiment change, i.e., to see if there are members who have significant and consistent role in positively changing the sentiments of other members by interacting with them. Interactions between different support seekers and support providers can be studied to calculate the change in sentiment of posts of seekers before and after getting the support.

---

[1]http://sparkpeople.com/

[2]http://www.everydayhealth.com/

[3]http://www.csn.cancer.org

[4]http://exchanges.webmd.com/

Common approaches for learning sentiment classifiers are based on supervised methods which rely on the availability of large amounts of labeled data. However, because of the high cost and effort in labeling the data, the amount of labeled data is very small compared to the amount of unlabeled data. Against this background, one question that can be raised is: *Can we effectively use the information contained in the unlabeled data to improve the sentiment classification in a health domain?* The research that we describe here addresses specifically this question.

We present a semi-supervised learning approach to sentiment classification of CSN posts that is able to incorporate information from the unlabeled data into the models. We use two types of sentiment features for classification: domain-independent (DI) and domain-dependent (DD) features. DI features represent polarity clues [4], [5], emoticons, punctuation marks (among others), which are used to express sentiments on online social media in general. DD features, on the other hand, are specific to a particular community. We use various DD features such as n-grams and POS tags, extracted from the CSN posts, and thus, specific to CSN. Previous works on sentiment analysis have used these features separately in supervised learning settings [6], [3]. To the best of our knowledge, we are the first to combine these two types of features in a health related domain using co-training, a well-known semi-supervised learning approach. Specifically, we investigate if the general and the domain-specific information about sentiment expressions can be effectively combined in a co-training approach to improve sentiment classification. We first train two supervised classifiers using DI and DD features and then combine them in such a way that one classifier can "guide" the other to minimize the number of mistaken examples. Experimental results show that the domain-dependent and domain-independent views can be used successfully in a co-training setting to improve sentiment classification in a health domain. Our classifiers that incorporate information from unlabeled data achieve an F-1 score of up to 0.843 outperforming strong baselines.

II. RELATED WORK

Sentiment analysis has been a highly active research area due to its important applications in mining, analyzing and summarizing user opinions from online sites such as product review sites, forums, Facebook, and Twitter [4], [7], [8], [9]. It essentially deals with identifying the polarity (positive or negative) of a piece of text (often with respect to a particular target). Here, we survey some of the sentiment analysis works.

Pang et al. [6] use supervised machine learning algorithms for sentiment analysis of movie reviews. They train their models using lexical features such as unigrams, bigrams, POS tags, etc. In their later work, they improve the sentiment classification by considering only the subjective sentences and applying polarity classifiers (developed in their previous work) on those sentences [10]. McDonald et. al [11] use joint models based on sequence labeling for sentiment classification at sentence and document level for product reviews. Wan et al. [12] perform sentiment classification of Chinese product reviews using co-training. They used machine translation to obtain the training data from labeled English reviews. For a Chinese review, they used its Chinese features and translated English features as the two independent views and used them to train their classifiers in the co-training setting.

Recently, there has been a growing interest in analyzing sentiments about various topics/themes/issues talked in social media. Pal et al. [13] detect sentiments about a set of pre-defined themes in blogs. Stavrianou et al. [14] propose an opinion-oriented graphical model for extracting information about opinions expressed in online forums. They focus on certain types of opinion information (such as opinion flow and attitudes of users), which cannot be extracted in the social network graphical model. Jiang et al. [9] perform target-dependent sentiment classification of tweets. They do polarity classification on the subjective content of tweets and finally take into account the contextual information of tweets (replies and retweets) using graphical models to improve sentiment classification.

Similar to our work, Qiu et al. [3] perform sentiment classification of CSN posts. They use sentiment features, e.g., sentiment clues, sentiment strength, punctuation marks, and two content features: *name* and *slang* to train their classifiers in a supervised setting. In contrast, we use semi-supervised methods and additional features. Specifically, we use domain-specific features such as unigrams, bigrams, and their POS tags. Also, we extract some of the features for post classification from a post and its last sentence.

III. APPROACH

In this section, we state our problem, describe the features used to classify the sentiment and explain our classification model.

A. Problem Formulation

CSN is a dynamic online community of cancer survivors, their families and friends. The structure of the CSN is similar to that of online forums, where users interact by posting messages in discussion threads. In CSN, members discuss cancer related issues such as cancer medications, side-effects, procedures, and cancer experiences. An interesting aspect of these discussions is the presence of emotions in the messages. Often, members start a discussion thread for seeking emotional support from the community by posting updates about their cancer reports, health conditions, etc. Other members provide emotional support by sympathizing, talking about their own experiences to make the (support) seeker feel better. These posts are indicative of the *dynamic* emotional states of the posters. Analyzing sentiments expressed in these posts can be very helpful in understanding the dynamics of the community .We identify the polarity of emotions/sentiments contained in a post by labeling the post as either positive or negative and model the task of sentiment analysis as a binary classification problem. **Problem Statement**: Given a post of a discussion thread in CSN, classify it into one of the two classes: *Positive* (denoted by $+$) and *Negative* (denoted by $-$).

B. Feature Engineering

In this section, we describe various features used in our classification model and our intuition behind using them. Table I lists and defines the various features used in our classification model.

Feature Name	Description
Polarity clues (extracted from post and its last sentence)	
PosDensity	Number of positive polarity clues in the post normalized by the number of words in the post.
NegDensity	Number of negative polarity clues in the post normalized by the number of words in the post.
PosVsNeg	Number of positive sentiment words per negative sentiment word in the post. Calculated as (No. of positive polarity clues+1)/(No. of negative polarity clues+1)
Sentiment strength features (extracted from post and its last sentence.)	
PosStrength	Positive sentiment strength of the post as given by SentiStrength algorithm.
NegStrength	Negative sentiment strength of the post as given by SentiStrength algorithm
PosVsNegStrength	PosStrength divided by NegStrength
Punctuation Marks	
numQues	Number of question marks in the post.
isQues	Whether the post has a question mark or not (1 for yes, 0 for no)
NumExclaim	Number of exclaimation marks in the post.
isExclaim	Whether the post has a exclaimation mark or not (1 for yes, 0 for no)
Domain-dependent features (for a sentence containing three words W_i, W_{i+1}, W_{i+2}.)	
BoW	W_i, W_{i+1}, W_{i+2}

TABLE I.   DESCRIPTION OF VARIOUS FEATURES USED FOR SENTIMENT CLASSIFICATION.

1) Domain-independent Features: We extract the following domain-independent sentiment features from the posts of CSN:

(a) **Polarity Clues:** These are the words/phrases/symbols used to express polarity of opinions/emotions in speech or written text. Polarity clues are a good indicator of the polarity of a piece of text and have been used extensively in sentiment analysis [4], [15]. Also, in online interactions, emoticons (such as ":)", ":(", ":-D", etc.) are widely used to express emotional states. We expect the distribution of these positive and negative polarity clues to be different in positive and negative posts. For a post, we use the frequency of occurrences of these clues in the post as features for classification. We used the list of positive and negative words created by Hu et. al. [4], and used the list of emoticons available from Wikipedia [5]. We extract three features PosDensity, NegDensity and PosVsNegDensity from a post and its last sentence (See Section III-C for details related to the last sentence). PosDensity is the number of positive polarity clues (positive words and positive emoticons) normalized by the number of words in the post. NegDensity is computed in a similar way. PosVsNegDensity is the number of positive polarity clues per negative polarity clues and is calculated by dividing $(PosDensity + 1)$ by $(NegDensity + 1)$.

(b) **Punctuation Marks:** In online interactions, punctuation is commonly used to show the intensity of emotions. For example, *I like it!* and *I like it !!!!!!!!* are two different ways of writing the same text, but with different intensities of emotion. The intensity of positive emotion is higher in the latter case. We extract question marks and exclamation marks from a post and used their frequency and occurrence in the post as features. We calculate four punctuation features for a post: numQuestion (number of question marks), isQuestion (whether a post contains a question mark or not), NumExclaim (number of exclamation marks) and isExclaim (whether a post contains an exclamation mark or not).

(c) **Sentiment Strength:** These features capture the strength of the sentiments expressed in posts. To calculate sentiment strength, we used the SentiStrength algorithm [16].

The algorithm is specifically designed to calculate sentiment strength of short informal texts in online social media. For a piece of text, the algorithm computes a positive sentiment value and a negative sentiment value. Using SentiStrength, we compute three features: PosStrength (positive sentiment strength), NegStrength (negative sentiment strength) and PosVsNegStrength (PosStrength/NegStrength). We extract these features from a post and its last sentence.

2) Domain-dependent Features: There are specific terms in every domain that are expressive of or are associated with a particular sentiment. These terms may or may not occur in other domains or, if they occur, then they may be associated with different sentiments. For example, the term *positive* is generally associated with positive sentiment but in the context of cancer reports it has negative sense, e.g., "the test is *positive*". Similarly, *unpredictable* is a positive attribute for a movie plot but negative for effects of a medication. To find these domain-dependent terms, we used the words of the posts as features. We removed the stop words using a general stop words list [6] and set the minimum document frequency *df* of a term to 3 (we experimented with document frequencies 3, 5 and 10 and found that $df = 3$ gave the best results). We call these features bag-of-words and denote them by BoW.

Post	Support type	Sentiment
Will be undergoing double mastectomy next week. Feeling quite nervous as I am sure you all can relate to! Name_A	Seeking emotional support.	Negative
Will keep you in prayers. Good luck!	Providing direct emotional support.	Positive
I undergone a bi-lateral mastectomy with Trampflap reconstruction. It wasn't easy for me to prepare myself and was feeling very nervous. Finally, I thank god as it went well and I found the reconstructed breast acceptable. Good luck Name_A !	Providing indirect emotional support.	Positive

TABLE II.   AN EXAMPLE THREAD SHOWING POSTS CONTAINING DIRECT AND INDIRECT EMOTIONAL SUPPORT ALONGWITH THEIR SENTIMENTS.

C. Last Sentence Effect

In this sub-section, we explain our reasoning behind extracting some of the features mentioned above from last sentences of the posts. In CSN, members provide emotional support to the support seekers mainly in two ways: Direct Emotional Support (DES) and Indirect Emotional Support (IES). In DES, they provide support *directly* by giving encouragement, positive feedback, sympathy, or other sentimental support *without* talking about their own experiences. In IES, they encourage the seeker *indirectly* by talking about their own stories, situations, experiences, and struggles that are similar to those of the seeker to relate themselves to the seeker in an effective manner and, finally, talking about their success in dealing with those situations to provide encouragement. This difference between DES and IES results in structural and lexical differences between the posts providing the two types of support. DSE posts usually have a positive tone throughout and, hence, contain positive sentiment words. In contrast, IES posts either have neutral tone (due to description of treatments, medications, etc.) or have negative tone (due to description of feelings, struggles related to the treatments) in the major part of the post and positive tone in the end, when the support providers talk about their successes in dealing with situations and say encouraging words. Hence, IES posts contain factual and negative sentiment words in their major part and positive words in the end (usually in the last sentence). To capture these differences, we extract Polarity Clues and Sentiment Strength features from the entire post and its last sentence, separately. Table II shows a thread where the thread starter seeks emotional support and the second post and the third post of the thread provides DES and IES, respectively. In many cases, members only write a few words, e.g., *good luck, thank you, bye*, or their names at the end of the post. To deal with these situations, we use a threshold of 3 words for a sentence to be the last sentence of a post.

Algorithm 1 Co-training algorithm

    Input: L, U, K, 's'
    $j \leftarrow 0$
    while $j \leq K$ && $U \neq \phi$ **do**
        Use L_I to train classifier C_I.
        Use L_D to train classifier C_D.
        $S \leftarrow \phi$
        Sample 's' posts from U and move them to S.
        $U \leftarrow U \setminus S$
        $H^I \leftarrow GetHighConfPost(C_I, S)$
        $H^U \leftarrow GetHighConfPost(C_U, S)$
        $L \leftarrow L \cup H^I \cup H^U$
        $j \leftarrow j + 1$
    end while
    Output: Classifiers C_I, C_D.

D. Model Training

Annotating large amounts of data for training high-accuracy classifiers is time-consuming and costly. It seems more plausible to have a small amount of annotated data to provide initial supervision to a classifier and then use the unannotated data (which is readily available in large amounts) to improve a classifier's performance. For this setting, semi-supervised learning such as co-training is a natural approach.

Co-training was originally developed by Blum and Mitchell [17] and applied to webpage classification. One of the requirements for co-training to work is that the data can be represented using two independent views. However, recent works show that the independence criteria can be relaxed [18] without much impact on the performance. For the webpage classification, the content of a webpage and its hyperlink are used as two independent views of a webpage. Two separate classifiers are trained on the two feature sets corresponding to the two views of the data (e.g., the words extracted from the content of a webpage and the words extracted from its hyperlink). The two classifiers then predict labels of unlabeled instances. The unlabeled instances predicted with highest confidence by both classifiers are moved to the labeled data and removed from the unlabeled set, ensuring that conflicting instances (i.e., instances predicted with high confidence by both classifiers, however, in opposite classes) are discarded. This process is repeated until the number of iterations reaches a particular threshold or all the unlabeled data are used up. The idea is that the two classifiers "teach" each other by re-training each classifier at each iteration on the data enriched with labeled instances predicted with high confidence by the other classifier.

In our case, we use domain-independent (DI) and domain-dependent (DD) sentiment features as the two views of a post for sentiment classification. The co-training algorithm is shown in Algorithm 1. L and U denote the labeled and the unlabeled datasets, respectively. I and D correspond to the domain-independent and domain-dependent views, respectively. Classifiers C_I and C_D are trained on the two views, L_I and L_D, of L respectively and make predictions on unlabeled instances in set S that is created by sampling s instances from U. Instances whose labels are predicted with high confidence (H^I and H^U) are selected using the method $getHighConfPost$ for both the classifiers (C_I and C_D) and are added to L. The method adds n highest confidence negative and p highest confidence positive examples from U to L, for different values of n and p The process is repeated K times, until the unlabeled data is exhausted. During the testing phase, the final prediction (on a test instance) is computed by taking the product of the predictions of the two (trained) classifiers as in [17].

IV. EXPERIMENTS AND RESULTS

We now describe our data and the experimental setting, and present our results.

A. Data

Our data comes from a popular online cancer support community, the Cancer Survivors' Network (CSN), developed and maintained by the American Cancer Society. CSN is an online community for cancer patients, cancer survivors, their families and friends. The features of CSN are similar to many online forums with dynamic interactive medium such as chat rooms, discussion boards, etc. Members of CSN post messages seeking information about cancer related issues, emotional support and other social support such as celebration of events and success stories. To conduct our experiments, we used posts from discussion boards of CSN between June, 2000 to June, 2012. A dataset of 786,000 posts from 75,867 threads was used in this study. For the labeled data for sentiment classification, we used the same subset of annotated posts as in the work of Qiu et al. [3]. This labeled data is a random sample of 293 posts from the discussion boards of CSN. Each post

is annotated as positive or negative based on the sentiment expressed by the poster. There are 201 posts labeled as positive and 92 posts labeled as negative.

B. Experimental Setting

We experimented with various machine learning algorithms (Naive Bayes, Support Vector Machines, Logistic Regression, Bagging, Boosting, etc.) to conduct our classification experiments. We first build classification models on the two types of features (DI and DD) using supervised learning algorithms. Logistic regression was found to give the best classification performance with DI sentiment features and Naive Bayes Multinomial with the DD sentiment features. We used the Weka data mining toolkit [19] to train and test the supervised learning algorithms. We used our own implementation of co-training. To evaluate the performance of our classifiers, we used macro-averaged precision, recall and F-1 score.

For co-training, we used 2000 unlabeled instances in U, sampled from all the unlabeled data, and 100 instances in S, sampled from U (see Algorithm 1). We experimented with different values for the number of iterations (K) and the size of unlabeled data (p, n) added to the labeled data after each iteration. Here, we report the results of the best performing setting, i.e, $K = 90$ and $(p, n) = (2, 1)$ where p and n are the number of examples from positive and negative classes added from the unlabeled data to the labeled data after each iteration respectively.

Model	Precision	Recall	F-1 score
BoW	0.714	0.724	0.717
DI^{nl}	0.782	0.788	0.783
DI^{l}	0.79	0.795	0.792
Qiu et al. [3]	0.781	0.788	0.781
Co-training $(BoW + DI^{l})(p = 2, n = 1, K = 90)$	**0.844**	**0.843**	**0.844**

TABLE III.   PERFORMANCE OF DIFFERENT CLASSIFICATION MODELS.

C. Baselines

We use the following classification models as baselines:

1) **DD**: Classifiers trained on domain-dependent features
2) **DI**: Classifiers trained on domain-independent features.
3) **Qiu et al.**: Classifiers developed by Qiu et al. [3].

D. Experimental Results

Table III presents results of comparison of the proposed co-training approach with the different classification baselines described above. As can be seen in the table, co-training outperforms all the other models with F-1 score of 0.843. Next, we have results of domain-independent features (DI). To see the impact of last sentence on the sentiment classification, we first build a model using DI features extracted only from posts and not separately from their last sentences [7] (denoted as DI^{nl}) and then we build the DI^{l} model in which we extract the DI features from posts and their last sentences separately. Including the last sentence improves the performance over the DI^{nl} model. In fact, DI^{l} model is the second best performing model (after co-training) outperforming the other two baselines. The model proposed by Qiu et al.[3] performs third best outperforming the BoW model.

---

[7]That is we do not extract those features separately from the last sentence. We do not discard the last sentence and extract the features from the remaining sentences of the post.

V. CONCLUSION AND FUTURE WORK

In this paper, we performed sentiment classification of user posts in an online health community using co-training. We use domain-specific and general information about sentiment expressions and combine them in the co-training setting. Our experiments show that co-training is an effective way to combine the two information with respect to sentiment classification performance. Also, we find that the last sentences of the posts play an important role in the sentiment classification. In future, we plan to use feature selection algorithms to extract high quality DD and DI features and use them to improve the sentiment classificationn and explore other learning methods such as active learning and unsupervised learning.

REFERENCES

[1] K. J. Petrie and J. A. Weinman, *Perceptions of health and illness: Current research and applications*, 1997, vol. 1.

[2] K. P. Davison, J. W. Pennebaker, and S. S. Dickerson, "Who talks," *American Psychologist*, vol. 55, no. 2, pp. 205–217, 2000.

[3] B. Qiu, K. Zhao, P. Mitra, D. Wu, C. Caragea, J. Yen, G. Greer, and K. Portier, "Get online support, feel better – sentiment analysis and dynamics in an online cancer survivor community," in *SocialComm' 11*, 2011, pp. 274–281.

[4] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *SIGKDD '04*, ser. KDD '04, 2004, pp. 168–177.

[5] V. Stoyanov, C. Cardie, and J. Wiebe, "Multi-perspective question answering using the opqa corpus," in *HLT-EMNLP '05*, ser. HLT '05. Stroudsburg, PA, USA: ACL, 2005, pp. 923–930.

[6] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *ACL-02*, ser. EMNLP '02, Stroudsburg, PA, USA, 2002, pp. 79–86.

[7] D. Ly, K. Sugiyama, Z. Lin, and M. Kan, "Product review summarization from a deeper perspective," in *JCDL*, 2011, pp. 311–314.

[8] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and T. By, "Sentiment analysis on social media," in *ASONAM' 12*, 2012, pp. 919–926.

[9] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in *HLT '11*, vol. 1, 2011, pp. 151–160.

[10] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *ACL '04*, 2004, p. 271.

[11] R. McDonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar, "Structured models for fine-to-coarse sentiment analysis," in *ACL '07*, vol. 45, 2007, p. 432.

[12] X. Wan, "Co-training for cross-lingual sentiment classification," in *ACL '09*, 2009, pp. 235–243.

[13] J. Kumar Pal and A. Saha, "Identifying themes in social media and detecting sentiments," in *ASONAM'10*, 2010, pp. 452–457.

[14] A. Stavrianou, J. Velcin, and J.-H. Chauchat, "Definition and measures of an opinion model for mining forums," in *ASONAM'09*. IEEE, 2009, pp. 188–193.

[15] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *ACL '02*, ser. ACL '02, 2002, pp. 417–424.

[16] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment strength detection for the social web," *Journal of the American Society for Information Science and Technology*, 2012.

[17] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *COLT '98*, ser. COLT' 98. New York, NY, USA: ACM, 1998, pp. 92–100.

[18] M.-F. Balcan, A. Blum, and Y. Ke, "Co-training and expansion: Towards bridging theory and practice," *Computer Science Department*, p. 154, 2004.

[19] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The weka data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.