Sexual predator detection in chats with chained classifiers

Hugo Jair Escalante
LabTL, INAOE
Luis Enrique Erro No. 1,
72840, Puebla, Mexico
hugojair@inaoep.mx

Esaú Villatoro-Tello*
Universidad Autónoma Metropolitana
Unidad Cuajimalpa
Mexico City, Mexico
villatoroe@inaoep.mx

Antonio Juárez
LabTL, INAOE
Luis Enrique Erro No. 1,
72840, Puebla, Mexico
antjug@inaoep.mx

Luis Villaseñor
LabTL, INAOE
72840, Puebla, Mexico
villasen@inaoep.mx

Manuel Montes-y-Gómez
LabTL, INAOE
72840, Puebla, Mexico
mmontesg@inaoep.mx

Abstract

This paper describes a novel approach for sexual predator detection in chat conversations based on sequences of classifiers. The proposed approach divides documents into three parts, which, we hypothesize, correspond to the different stages that a predator employs when approaching a child. Local classifiers are trained for each part of the documents and their outputs are combined by a chain strategy: *predictions of a local classifier are used as extra inputs for the next local classifier*. Additionally, we propose a ring-based strategy, in which the chaining process is iterated several times, with the goal of further improving the performance of our method. We report experimental results on the corpus used in the first international competition on sexual predator identification (PAN'12). Experimental results show that the proposed method outperforms a standard (global) classification technique for the different settings we consider; besides the proposed method compares favorably with most methods evaluated in the PAN'12 competition.

1 Introduction

Advances in communications' technologies have made possible to any person in the world to communicate with any other in different ways (e.g., text, voice, and video) regardless of their geographical locations, as long as they have access to internet. This undoubtedly represents an important and highly needed benefit to society. Unfortunately, this benefit also has brought some collateral issues that affect the security of internet users, as nowadays we are vulnerable to many threats, including: cyber-bullying, spam, fraud, and sexual harassment, among others.

A particularly important concern has to do with the protection of children that have access to internet (Wolak et al., 2006). Children are vulnerable to attacks from paedophiles, which "groom" them. That is, adults who meet underage victims online, engage in sexually explicit text or video chat with them, and eventually convince the children to meet them in person. In fact, one out of every seven children receives an unwanted sexual solicitation online (Wolak et al., 2006). Hence, the detection of cyber-sexual-offenders is a critical security issue that challenges the field of information technologies.

This paper introduces an effective approach for sexual predator detection (also called sexual predator identification) in chat conversations based on chains of classifiers. The proposed approach divides documents into three parts, with the hypothesis that different parts correspond to the different stages that predators adopt when approaching a child (Michalopoulos and Mavridis, 2011). Local classifiers are trained for each part of the documents and their outputs are combined by a chaining strategy. In the chain-based approach the predictions of a local classifier are used as extra inputs for the next local classifier. This strategy is inspired from chain-based classifiers developed for the task of multi-label classification (Read et al., 2011). A ring-based approach is proposed, in which the generation of chains of classifiers is iterated several times. We report experimental results in the corpus used in the first international competition on sexual predator identification (PAN-2012) (Inches and Crestani,

*Esaú Villatoro is also external member of LabTL at INAOE.

Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 46–54,
Atlanta, Georgia, 14 June 2013.

2012). Experimental results show that chain-based classifiers outperform standard classification methods for the different settings we considered. Furthermore, the proposed method compares favorably with alternative methods developed for the same task.

2 Sexual predator detection

We focus on the detection of sexual predators in chat rooms, among the many cyber-menaces targeting children. This is indeed a critical problem because most sexually-abused children have agreed voluntarily to met with their abuser (Wolak et al., 2006). Therefore, anticipatively detecting when a person attempts to approach a children, with malicious intentions, could reduce the number of abused children.

Traditionally, a term that is used to describe malicious actions with a potential aim of sexual exploitation or emotional connection with a child is referred as "Child Grooming" or "Grooming Attack" (Kucukyilmaz et al., 2008). Defined in (Harms, 2007) as: *"a communication process by which a perpetrator applies affinity seeking strategies, while simultaneously engaging in sexual desensitization and information acquisition about targeted victims in order to develop relationships that result in need fulfillment"* (e.g. physical sexual molestation).

The usual approach[1] to catch sexual predators is through police officers or volunteers, whom behave as fake children in chat rooms and provoke sexual offenders to approach them. Unfortunately, online sexual predators always outnumber the law enforcement officers and volunteers. Therefore, tools that can automatically detect sexual predators in chat conversations (or at least serve as support tool for officers) are highly needed.

A few attempts to automate processes related to the sexual predator detection task have been proposed already (Pendar, 2007; Michalopoulos and Mavridis, 2011; RahmanMiah et al., 2011; Inches and Crestani, 2012; Villatoro-Tello et al., 2012; Bogdanova et al., 2013). The problem of detecting conversations that potentially include a sexual predator approaching a victim has been approached, for example, by (RahmanMiah et al., 2011; Villatoro-Tello et al., 2012; Bogdanova et al., 2013). RahmanMiah et al. discriminated among child-exploitation, adult-adult and general-chatting conversations using a text categorization approach and psychometric information (RahmanMiah et al., 2011). Recently, Bogdanova et al. approached the same problem, the authors concluded that standard text-mining features are useful to distinguish general-chatting from child-exploitation conversations, but not for discriminating between child-exploitation and adult-adult conversations (Bogdanova et al., 2013). In the latter problem, features that model behavior and emotion resulted particularly helpful. N. Pendar approached the problem of distinguishing predators from victims within chat conversations previously confirmed as containing a grooming attack (Pendar, 2007). The author collapsed all of the interventions from each participant into a document and approached the problem as a standard text categorization task with two classes (victim vs. predator).

A more fine grained approximation to the problem was studied by (Michalopoulos and Mavridis, 2011). The authors developed a probabilistic method that classifies chat interventions into one of three classes: 1) *Gaining Access:* indicate predators intention to gain access to the victim; 2) *Deceptive Relationship:* indicate the deceptive relationship that the predator tries to establish with the minor, and are preliminary to a sexual exploitation attack; and 3) *Sexual Affair:* clearly indicate predator's intention for a sexual affair with the victim. These categories correspond to the different stages that a sexual offender adopt when approaching a child. As (Pendar, 2007), (Michalopoulos and Mavridis, 2011) approached this problem as one of text categorization (equating interventions to short-documents). They removed stop words and applied a spelling correction strategy, their best results were obtained with a Naïve Bayes classifier, reaching performance close to 96%. Thus giving evidence that the three categories can be recognized reasonably well. Which in turn gives evidence that modeling the three stages could be beneficial for recognizing sexual predators; for example, when it is not known whether a conversation contains or not a grooming attack. This is the underlying hypothesis behind the proposed method. We aim to use local classifiers, specialized in the different stages a predator approaches a

[1] Adopted for example by the Perverted Justice organization, http://www.perverted-justice.com/

child. Then, we combine the outputs of local classifiers with the goal of improving the performance on sexual predator detection in conversations including both: grooming attacks and well-intentioned conversations.

Because of the relevance of the problem, and of the interest of several research groups from NLP, it was organized in 2012 the first competition of sexual predator identification (Inches and Crestani, 2012). The problem approached in the competition was that of identifying sexual predators from conversations containing both: grooming attacks and well-intentioned conversations. The organizers provided a large corpus divided into development and evaluation data. Development (training) data were provided to participants for building their sexual-predator detection system. In a second stage, evaluation (testing) data were provided to participants, whom had to apply their system to that data and submit their results. Organizers evaluated participants using their predictions on evaluation data (labels for the evaluation data were not provided to participants during the competition).

Several research groups participated in that competition, see (Inches and Crestani, 2012). Some participants developed tailored features for detecting sexual predators (see e.g., (Eriksson and Karlgren, 2012)), whereas other researchers focused on the development of effective classifiers (Parapar et al., 2012). The winning approach implemented a two stage formulation (Villatoro-Tello et al., 2012): in a first step suspicious conversations where identified using a two class classifier. Suspicious conversations are those that potentially include a sexual predator (i.e., a similar approach to (RahmanMiah et al., 2011)). In a second stage, sexual predators were distinguished from victims in the suspicious conversations identified in the first stage (a similar approach to that of (Pendar, 2007)). For both stages a standard classifier and a bag-of-words representation was used.

The methods proposed in this paper were evaluated in the corpus used in the first international competition on sexual predator detection, PAN'12 (Inches and Crestani, 2012). As explained in the following sections, the proposed method uses standard representation and classification methods, therefore, the proposed methods can be improved if we use tailored features or learning techniques for sexual predator detection.

3 Chain-based classifiers for SPD

Chain-based classifiers were first proposed to deal with multi-label classification (Read et al., 2011). The goal was to incorporate dependencies among different labels, which are disregarded by most multi-label classification methods. The underlying idea was to increase the input space of classifiers with the outputs provided by classifiers trained for other labels. The authors showed important improvements over traditional methods.

In this paper, we use chain-based classifiers to incorporate dependencies among local classifiers associated to different segments of a chat conversation. The goal is building an effective predator-detection model made of a set of local models specialized at classifying certain segments of the conversation. Intuitively, we would like to have a local model associated to each of the stages in which a sexual predator approaches a child: *gaining access*, *deceptive relationship* and *sexual affair* (Michalopoulos and Mavridis, 2011). We associate a segment of the conversation to each of the three stages. The raw approach proposed in this work consists of dividing the conversation into three segments of equal length. The first, second and third segments of each conversation are associated to the first, second and third stages, respectively. Although, this approach is too simple, our goal was to determine whether having local classifiers combined via a chaining strategy could improve the performance on sexual predator detection.

We hypothesize that as the vocabulary used in different segments of the conversation is different, specialized models can result in better performance for classifying these local segments. Since local classifiers can only capture local information, it is desirable to somehow connect these classifiers in order to make predictions taking into account the whole conversation. One way to make local classifiers dependent is thought the chain-based methodology, where the outputs of one local classifier are feed as inputs for the next local classifier; the final prediction for the whole conversation can be obtained in several ways as described below.

The proposed approach is described in Figure 1. Since our goal is to detect sexual predators from chat conversations directly, we model each user (well-intentioned user, victim or sexual predator) by their set of interventions. Thus, we generate a single conversation for each user using their interventions, keeping the order in which such interventions happened. The approached problem is to classify these conversations into sexual-predator or any-other-type-of-user. In the following we call simply conversations to the generated per-user conversations.

Chat conversations are divided into three (equally-spaced) parts. Next, one local-classifier is trained for each part of the document according to a predefined order[2], where two out of the three classifiers (second and third) are not independent. Let p_1, p_2, and p_3 denote the segments of text that will be used for generating the first, second and third classifiers. The triplet $\{p_1, p_2, p_3\}$ can be any of the six permutations of 3 segments, this tripled determines the order in which classifiers will be built. Once that a particular order has been defined, a first local-classifier, f_1, is trained using the part p_1 from all of the training documents ($p_1 \in \{first, second, third\}$). Next, a second local-classifier, f_2, is trained by using the part p_2 from all of the training documents. f_2 is built by using both attributes extracted from part p_2 of conversations and the outputs of the first classifier over the training documents. Thus, classifier f_2 depends on classifier f_1, through the outputs of the latter model. A third local-classifier, f_3, is trained using attributes extracted from part p_3 from all conversations, the input space for training f_3 is augmented with the predictions of classifiers f_2 and f_1 over the training documents. Hence, the third classifier depends on the outputs of the first and second classifiers.

Once trained, the chain of local-classifiers can be used to make predictions for the whole conversation in different ways. When a test conversation needs to be classified it is also split into 3 parts. Part p_1 is feeded to classifier f_1, which generates a prediction for f_1. Next, part p_2 from the test document, together with the prediction for p_1 as generated by f_1 are feeded to classifier f_2. Likewise, the outputs of f_2 and f_1, together with part p_3 from the document are used as inputs for classifier f_3. Clearly, since we have predictions for the test document at the three stages of the chain (from $f_{1,2,3}$) we can make a prediction at any stage. The prediction from classifier f_3 is called *chain-prediction* as it is the outcome of the dependent local-classifiers.

[2]We hypothesize that building a chain of classifiers using different orders results in different performances, we evaluate this aspect in Section 4.

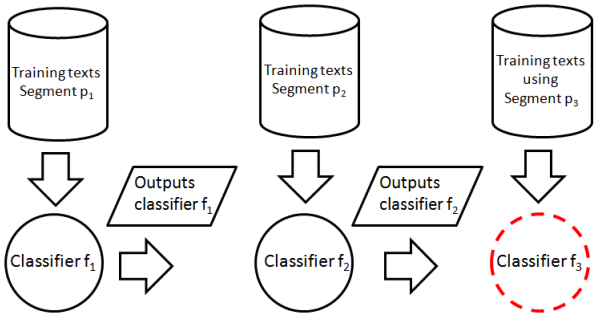

Figure 1: General diagram of the chain-based approach.

Additionally to local and chain-prediction, we propose a ring-like structure for chain-based classifiers in which the outputs of the third local-classifier are used again as inputs for another local model, where the order can be different to that used in the previous iteration. This process is iterated for a number of times, where we can make predictions at every link (local-classifier) of the ring. In addition, after a number of iterations we can make predictions by combining the outputs (like in an ensemble) generated by all of the classifiers considered in the ring up to that iteration. The underlying idea is to explore the performance of the chain as more local-models, that can use short and long term dependencies with other classifiers, are incorporated. Our hypothesis is that after incorporating a certain number of local-dependent-models, the predictions for the whole conversations will be steady and will improve the performance of the straight chain approach.

Algorithm 1 describes the proposed ring-based classifier. $\mathcal{E}$ denotes the set of extra inputs that have to be added to individual classifiers, which are the cumulative outputs of individual classifiers. $\mathcal{P}$ is a set of predefined permutations from which different orders can be took from, where $\mathcal{P}_i$ is the i^{th} permutation. We denote with $\texttt{atts}\ (p_i, \mathcal{E})$ to the pro-

cess of extracting attributes from documents' part p_i and merging them with attributes stored in $\mathcal{E}$. `atts` generates the representation that a classifier can use. `train` $[f(X)]$ denotes the process of training classifier f using inputs X. $\mathcal{M}_c$ stores the models trained through the ring process.

Algorithm 1 Ring-based classifier.

Require: g : # iterations; $\mathcal{P}$: set of permutations;
  $\mathcal{E} = \{\}$
  $i = 0; c = 1;$
  while $i \leq g$ **do**
    $i++;$
    $\{p_1, p_2, p_3\} \leftarrow \mathcal{P}_i;$
    for $j = 1 \rightarrow 3$ **do**
      $X \leftarrow$ `atts` $[p_j, \mathcal{E}]$
      $f_j^* \leftarrow$ `train` $[f_j(X)];$
      $\mathcal{M}_c \leftarrow f_j^*;$
      $\mathcal{E} \leftarrow \mathcal{E} \cup f_j^*(p_j, \mathcal{E});$
      $c++;$
    end for

  end while
  return $\mathcal{M}_c$: trained classifiers (ring-based approach);

When a test conversation needs to be labeled, the set of classifiers in $\mathcal{M}$ are applied to it using the same order in the parts that was used when generating the models. Each time a model is applied to the test instance, the prediction of such model is used to increase the input space that is to be used for the next model. We call the prediction given by the last model $\mathcal{M}_g$, *ring-prediction*. One should note that, as before, we can have predictions for the test conversation from every model $\mathcal{M}_i$. Besides, we can accumulate the predictions for the whole set of models $\mathcal{M}_{1,...,g}$. Another alternative is to combine the predictions of the three individual classifiers in each iteration of the ring (every execution of the for-loop in Algorithm 1); this can be done, e.g., by weight averaging. In the next section we report the performance obtained by all these configurations.

4 Experiments and results

For the evaluation of the proposed approach we considered the data set used in the first international competition on sexual predator identification[3] (PAN-2012) (Inches and Crestani, 2012). Table 1 presents some features from the considered data set. The data set contains both chat conversations including sexual predators approaching minors and (authentic) conversations between users (which can or cannot be related to a sexual topic). The data set provided by the organizers contained too much noisy information that could harm the performance of classification methods (e.g., conversations with only one participant, conversations of a few characters long, etc.). Therefore, we applied a preprocessing that aimed to both remove noisy conversations and reducing the data set for scalability purposes. The filtering preprocessing consisted of eliminating: conversations with only one participant, conversations with less than 6 interventions per each participant, conversations that had long sequences of unrecognized characters (images, apparently). The characteristics of the data set after filtering are shown within parentheses in Table 1. It can be seen that the size of the data set was reduced considerably, although a few sexual predators were removed, we believe the information available from them was insufficient to recognize them.

Table 1: Features of the data set considered for experimentation (Inches and Crestani, 2012). We show the features of the raw data and in parentheses the corresponding features after applying the proposed preprocessing.

Feature	Development	Evaluation
# Convers.	$66,928$ $(6,588)$	$155,129$ $(15,330)$
# Users	$97,690$ $(11,038)$	$218,702$ $(25,120)$
# Sexual Pr.	$148(136)$	254 (222)

Conversations were represented using their bag-of-words. We evaluated the performance of different representations and found that better results were obtained with a Boolean weighting scheme. No stop-word removal nor stemming was applied, in fact, punctuation marks were conserved. We proceeded this way because we think in chat conversations every character conveys useful information to characterize users, victims and sexual predators. This is because of the highly unstructured and informal language used in chat conversations, as discussed in related works (Kucukyilmaz et al., 2008; RahmanMiah et al., 2011; Rosa and Ellen, 2009).

For indexing conversations we used the TMG toolbox (Zeimpekis and Gallopoulos, 2006). The re-

[3] http://pan.webis.de/

sultant vocabulary was of $56,964$ terms. For building classifiers we used a neural network as implemented in the CLOP toolbox (Saffari and Guyon, 2006). Our choice is based on results from a preliminary study.

4.1 Performance of local classifiers

We first evaluate the performance of global and local classifiers separately. A global classifier is that generated using the content of the whole conversation, it resembles the formulation from (Pendar, 2007). Local classifiers were generated for each of the segments. Table 2 shows the performance of the global and local models. We report the average (of 5 runs) of precision, recall and F_1 measure for the positive class (sexual predators).

Table 2: Performance of global (row 2) and local classifiers (rows 3-6).

Setting	Precision	Recall	F_1 Measure
Global	95.14%	49.91%	65.42%
Segment 1	**96.16%**	**59.20%**	**73.23%**
Segment 2	96.25%	48.82%	64.72%
Segment 3	93.43%	51.87%	66.68%

It can be seen from Table 2 that the performance of the global model and that obtained for segments 2 and 3 are comparable to each other in terms of the three measures we considered. Interestingly, the best performance was obtained when the only the first segment of the conversation was used for classification. The difference is considerable, about 11.93% of relative improvement. This is a first contribution of our work: *using the first segment of a conversation can improve the performance obtained by a global classifier.* Since the first segment of conversations (barely) corresponds to the *gaining access* stage, the result provides evidence that sexual predators can be detected by the way they start approaching to their victims. That is, the way a well-intentioned person starts a conversation is somewhat different to that of sexual predators approaching a child. Also, it is likely that this makes a difference because for segments 2 and 3, conversations containing grooming attacks and well-intentioned conversations can be very similar (well-intentioned conversations can deal sexual thematic as well).

4.2 Chain-based classifiers

In this section we report the performance obtained by different settings of chain based classifiers. We first report the performance of the chain-prediction strategy, see Section 3. Figure 2 shows the precision, recall and F_1 measure, obtained by the chain-based classifier for the different permutations of the 3 segments (i.e., all possible orders for the segments). For each order, we report the initial performance (that obtained with the segment in the first order) and the chain-prediction, that is the prediction provided by the last classifier in the chain.

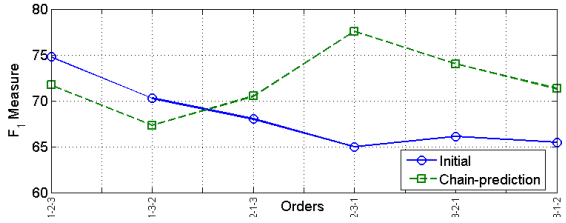

Figure 2: F_1 measure by the initial and chain-based classifier for different orders.

From Figure 2 it can be observed that the chain-prediction outperformed the initial classifier for most of the orders in terms of F_1 measure. For orders starting with segment 1 (1-2-3 and 1-3-2) chain-based classifiers worsen the initial performance. This is due to the high performance of local classifier for segment 1 (see Table 2), which cannot be improved with successive local classifiers. However, the best performance overall was obtained by the chain-based classifier with the order 2-3-1. The relative improvement of this configuration for the chain-based method over the global classifier (the one using the whole conversations) was of 18.52%. One should note that the second-best performance was obtained with the order 3-2-1. Hence, putting the most effective classifier (that for segment 1) at the end seems to have a positive influence in the chain-based classifier. We have shown evidence that chain-based classifiers outperform both the global classifier and any of the local methods. Also, the order of classifiers is crucial for obtaining acceptable results with the chain technique: *using the best classifier in the last position yields better performance; and, putting the best classifier at the beginning would lead the chain to worsen initial performance.*

4.3 Ring-based classifiers

In this section we report experimental results on sexual predator detection obtained with the ring-based strategy. Recall a ring-based classifier can be seen as a chain that is replicated several times with different orders, so we can have predictions for each of the local classifiers at each node of the ring/chain. Besides, we can obtain periodical/cumulative predictions from the chain and predictions derived from combining predictions from a subset of local classifiers in the chain. We explore the performance of all of these strategies in the rest of this section.

We implement ring-based classifiers by successively applying chain-based classifiers with different orders. We consider the following alternatives for detecting predators with ring-based classifiers:

- **Local.** We make predictions with *local classifiers* each time a local classifier is added to the ring (no dependencies are considered). We report the average performance (*segments avg.*) and the maximum performance (*segments max.*) obtained by local classifiers in each of the orders tried.

- **Chain-prediction.** We make predictions with *chain-based classifiers* each time a local classifier is added to the ring. We report the average performance (*chain-prediction avg.*) and the maximum performance (*chain-prediction max.*) obtained by chain-based classifiers per each of the orders tried.

- **Ensemble of chain-based classifiers.** We combine the outputs of the three *chain-based classifiers* built for each order; this method is referred to as *LC-Ensemble*.

- **Cumulative ensemble.** We combine the outputs (via averaging) of all the *chain-based classifiers* that have been built each time an order is added to the ring; we call this method *Cumulative-Ensemble*.

Besides reporting results for these approaches we also report the performance obtained by the global classifier (*Whole conversations*), see Table 2.

We iterated the ring-based classifier for a fixed number of orders. We tried 24 orders, repeating the following process two times: we tried the permutations of the 3 segments in lexicographical order, followed by the same permutations on inverted lexicographical order. So a total of 24 different orders were evaluated. Figure 3 shows the results obtained by the different settings we consider for a typical run of our approach.

Several findings can be drawn from Figure 3. With exception of the average of local classifiers (*segments avg.*), all of the methods outperformed consistently the global classifier (*whole conversations*). Thus confirming the competitive performance of local classifiers and that of chain-based variants. The best local classifier from each order (*segments max.*) achieved competitive performance, although it was outperformed by the average of chain-based classifiers (*chain-prediction avg.*). Since local classifiers are independent, no tendency on their performance can be observed as more orders are tried. On the contrary, the performance chain-based methods (as evidenced by the avg. and max of chain-predictions) improves for the first 8-9 orders and then remains steady. In fact, the best (per-order) chain-prediction (*chain-prediction max.*) obtained performance comparable to that obtained by ensemble methods. One should note, however, that in the *chain-prediction max.* formulation we report the best performance from each order tried, which might correspond to different segments in the different orders. Therefore, it is not clear how the select the specific order to use and the specific segment of the chain that will be used for making predictions, when putting in practice the method for a sexual-predator detection system. Notwithstanding, stable average predictions can be obtained when more than 6-8 orders are used (*chain-prediction avg.*), still the performance of this approach is lower than that of ensembles.

Clearly, the best performance was obtained with the ensemble methods: *chain-ensemble* and *cumulative-ensemble*. Both approaches obtained similar performance, although the *chain-ensemble* slightly outperformed *cumulative-ensemble*. The chain-ensemble considers dependencies within each order and not across orders, thus its performance after trying the 6 permutations of 3 segments did not vary significantly. This is advantageous as only 6 orders have to be evaluated to obtain competitive performance. Unfortunately, as with single chain-classifiers it may be unclear how to select the particular order to use to implement a sexual-predator detection system.

On the other hand, the *cumulative-ensemble* ob-

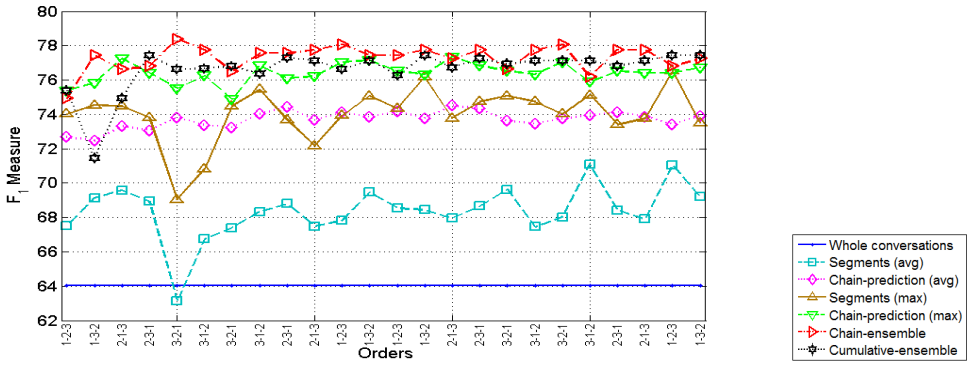

Figure 3: Performance of the different variants of ring-based classifiers for sexual predator detection.

tained stable performance after $\approx$ 12 orders were considered. Recall this method incorporates dependencies among the different orders tried. Although it requires the evaluation of more orders than the *chain-ensemble* to converge, this method is advantageous for a real application: *after a certain number of orders it achieves steady performance, and since it averages the outputs of all of the chain-classifiers evaluated up to a certain iteration, its performance does not rely on selecting a particular configuration.* In consequence, we claim the cumulative-ensemble offers the best tradeoff between performance, stability and model selection.

4.4 Comparison with related works

Table 3 shows a comparison of the configuration *cumulative-ensemble* against the top-ranked participants in the PAN'12 competition. We show the performance of the top-5 participants as described in (Inches and Crestani, 2012), additionally we report the average performance obtained by the methods of the 16 participating teams. We report, F_1 and $F_{0.5}$ measures, and the rank for each participant. We report $F_{0.5}$ measure because that was the leading evaluation measure for the PAN'12 competition.

From Table 3 it can be observed that the proposed method is indeed very competitive. *The results obtained by our method outperformed significantly the average performance (row 7) obtained by all of the participants in all of the considered measures.* In terms of F_1 measure our method would be ranked in the fourth position, while in terms of the $F_{0.5}$ measure our method would be ranked third.

Table 3: Comparison of the proposed method with related works evaluated in the PAN'12 competition (Inches and Crestani, 2012).

Participant	F_1	$F_{0.5}$	**Rk.**
(Villatoro-Tello et al., 2012)	87.34	93.46	1
(Inches and Crestani, 2012)	83.18	91.68	2
(Parapar et al., 2012)	78.16	86.91	3
(Morris and Hirst, 2012)	74.58	86.52	4
(Eriksson and Karlgren, 2012)	87.48	86.38	5
(Inches and Crestani, 2012)	49.10	51.06	-
Our method	78.98	89.14	-

5 Conclusions

We introduced a novel approach to sexual-predator detection in which documents are divided into 3 segments, which, we hypothesize, could correspond to the different stages in that a sexual predator approaches a child. Local classifiers are built for each of the segments, and the predictions of local classifiers are combined through a strategy inspired from chain-based classifiers. We report results on the corpus used in the PAN'12 competition, the proposed method outperforms a global approach. Results are competitive with related works evaluated in PAN'12. Future work includes applying the chain-based classifiers under the two-stage approach from Villatoro et al. (Villatoro-Tello et al., 2012).

Acknowledgments

This project was supported by CONACYT under project grant 134186. The authors thank INAOE, UAM-C and SNI for their support.

References

D. Bogdanova, P. Rosso, and T. Solorio. 2013. Exploring high-level features for detecting cyberpedophilia. In *Special issue on on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2012), Computer Speech and Language (accepted)*.

G. Eriksson and J. Karlgren. 2012. Features for modelling characteristics of conversations. In P. Forner, J. Karlgren, and C. Womser-Hacker, editors, *Working notes of the CLEF 2012 Evaluation Labs and Workshop*, Rome, Italy. CLEF.

C. Harms. 2007. Grooming: An operational definition and coding scheme. *Sex Offender Law Report*, 8(1):1–6.

G. Inches and F. Crestani. 2012. Overview of the international sexual predator identification competition at PAN-2012. In P. Forner, J. Karlgren, and C. Womser-Hacker, editors, *Working notes of the CLEF 2012 Evaluation Labs and Workshop*, Rome, Italy. CLEF.

T. Kucukyilmaz, B. Cambazoglu, C. Aykanat, and F. Can. 2008. Chat mining: predicting user and message attributes in computer-mediated communication. *In Information Processing and Management*, 44(4):1448–1466.

D. Michalopoulos and I. Mavridis. 2011. Utilizing document classification for grooming attack recognition. In *Proceedings of the IEEE Symposium on Computers and Communications*, pages 864–869.

C. Morris and G. Hirst. 2012. Identifying sexual predators by svm classification with lexical and behavioral features. In P. Forner, J. Karlgren, and C. Womser-Hacker, editors, *Working notes of the CLEF 2012 Evaluation Labs and Workshop*, Rome, Italy. CLEF.

J. Parapar, D. E. Losada, and A. Barreiro. 2012. A learning-based approach for the identification of sexual predators in chat logs. In P. Forner, J. Karlgren, and C. Womser-Hacker, editors, *Working notes of the CLEF 2012 Evaluation Labs and Workshop*, Rome, Italy. CLEF.

N. Pendar. 2007. Toward spotting the pedophile telling victim from predator in text chats. In *Proceedings of the IEEE International Conference on Semantic Computing*, pages 235–241, Irvine California USA.

M. W. RahmanMiah, J. Yearwood, and S. Kulkarni. 2011. Detection of child exploiting chats from a mixed chat dataset as text classification task. In *Proceedings of the Australian Language Technology Association Workshop*, pages 157–165.

J. Read, B. Pfahringer, G. Holmes, and E. Frank. 2011. Classifier chains for multi-label classification. *Machine Learning Journal*, 85(3):333–359.

K. D. Rosa and J. Ellen. 2009. Text classification methodologies applied to micro-text in military chat. In *Proceedings of the eight IEEE International Conference on Machine Learning and Applications*, pages 710–714.

A. Saffari and I Guyon. 2006. Quick start guide for CLOP. Technical report, Graz-UT and CLOPINET, May.

E. Villatoro-Tello, A. Juárez-González, H. J. Escalante, M. Montes-Y-Gómez, and L. Villaseñor-Pineda. 2012. A two-step approach for effective detection of misbehaving users in chats. In P. Forner, J. Karlgren, and C. Womser-Hacker, editors, *Working notes of the CLEF 2012 Evaluation Labs and Workshop*, Rome, Italy. CLEF.

J. Wolak, K. Mitchell, and D. Finkelhor. 2006. Online victimization of youth: Five years later. Bulleting 07-06-025, National Center for Missing and Exploited Children, Alexandia, Alexandria, VA.

D. Zeimpekis and E. Gallopoulos, 2006. *Grouping Multidimensional Data: Recent Advances in Clustering*, chapter TMG: A MATLAB toolbox for generating term-document matrices from text collections, pages 187–210. Springer.