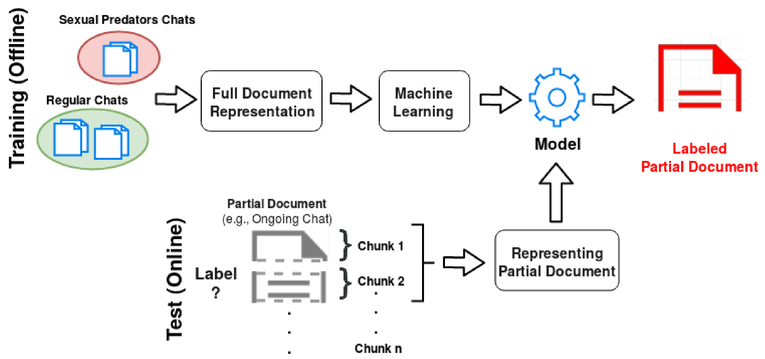
# Early Text Classification using Multi-Resolution Concept Representations

A. Pastor López-Monroy[1,2], Fabio A. González[3]
Manuel Montes-y-Gómez[4], Hugo Jair Escalante[4], Thamar Solorio[2]
[1] *Mathematics Research Center CIMAT,* [2] *Department of Computer Science, University of Houston, USA*
[3] *Systems and Computer Engineering Department, Universidad Nacional de Colombia*
[4] *Department of Computer Science, INAOE, México*

## 1. Introduction

- Detecting risks on Social Media data with as much anticipation as possible is crucial for prevention.
- **Early Text Classification (ETC)** is an emerging field where the goal is to anticipate the prediction by using as little text as possible.

**Early Classification on Social Media**




- We need to model very short length documents on early stages.
- Late stages require to exploit the additional evidence as much as possible to make accurate predictions.

## 2. Word Representation

**Temporal Variation of Terms (TVT)**

| | $p_1$ | . | . | . | $p_i$ |
|---|---|---|---|---|---|
| $t_1$ | $wtp_{11}(p_1, t_1)$ | . | . | . | $wtp_{i1}(p_i, t_1)$ |
| . | . | . | | | |
| . | . | | . | | |
| . | . | | | . | |
| $t_j$ | $wtp_{1j}(p_1, t_j)$ | | | | $wtp_{ij}(p_i, t_j)$ |



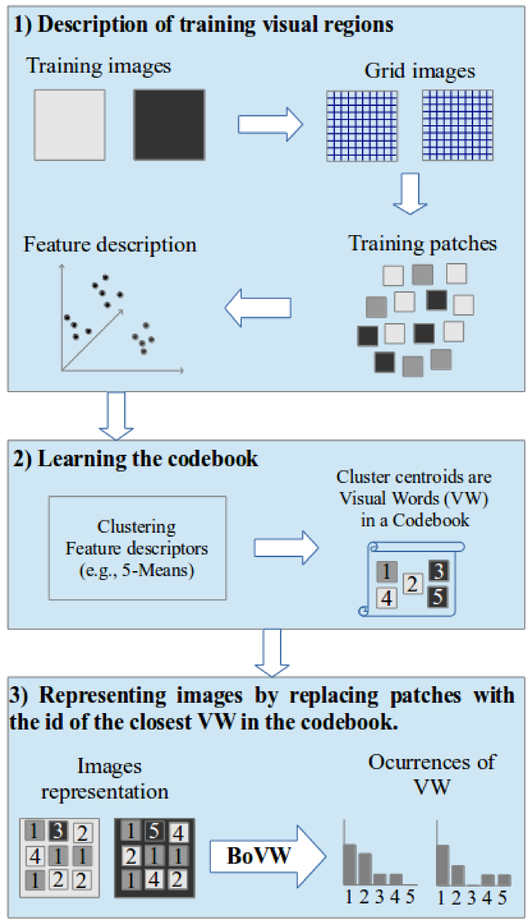

The representation builds term vectors $\mathbf{t_j} = \langle tp_{1j}, \ldots, tp_{ij} \rangle$, where $tp_{ij}$ is a value representing the relationship of the term $t_j$ with the class $p_i$.

$$wtp_{ij} = \sum_{k:d_k \in P_i} \log_2 \left(1 + \frac{tf_{kj}}{len(d_k)}\right)$$

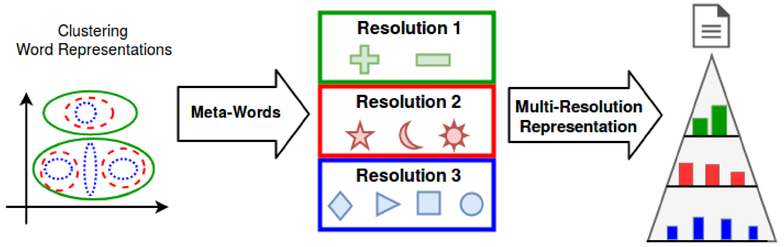
These word vectors are built at specific chunks in training.

## 3. Multi Resolution Representation (MulR)

**Single Resolution is analogous to the BoVW**

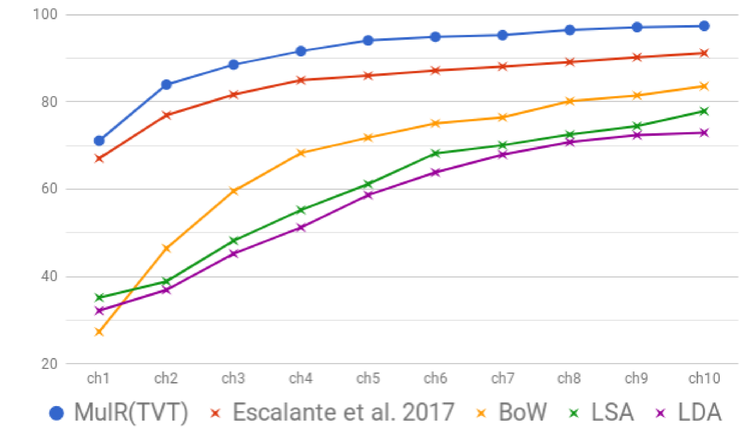



**Intuitive idea of Multi-Resolutions**




## 4. Evaluation

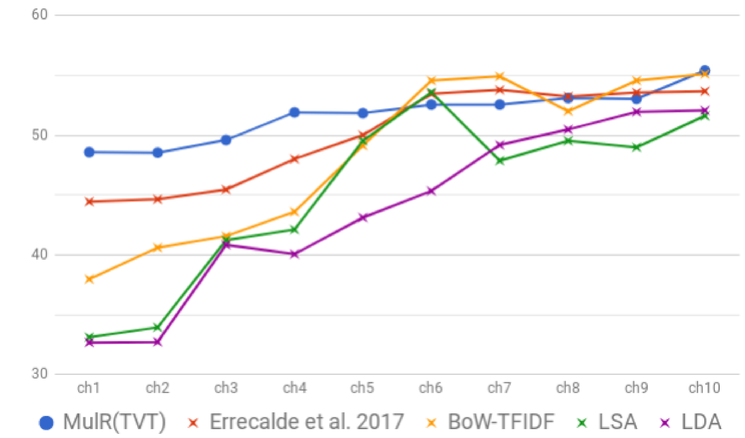
**Sexual Predator Detection**




1. The best word vector representation was TVT for ETC.
2. MulR is effective for all early chunks.
3. As more text is available, methodologies significantly improve.

| Resolutions: Sexual Predators | | | | | |
|---|---|---|---|---|---|
| **Test set** | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ |
| chunk-1 | 4 | 3 | 2 | 1 | 0 |
| chunk-2 | 3 | 2 | 3 | 2 | 0 |
| chunk-3 | 3 | 2 | 2 | 2 | 1 |
| chunk-4 | 3 | 2 | 3 | 1 | 1 |
| chunk-5 | 3 | 2 | 3 | 1 | 1 |
| chunk-6 | 3 | 1 | 4 | 1 | 1 |
| chunk-7 | 3 | 1 | 3 | 2 | 1 |
| chunk-8 | 2 | 2 | 1 | 2 | 3 |
| chunk-9 | 3 | 1 | 1 | 2 | 3 |
| chunk-10 | 3 | 0 | 2 | 2 | 3 |

**Depression Detection**




1. MulR(TVT) improves between ≈5 % and ≈2 % in chunks 1 to 4.
2. Depression Detection problem is much harder than SPD.
3. $F_1$ under ≈60 % and highly unbalanced dataset.