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)

The representation builds term vectors \( t_j = (t_{p_{j1}}, \ldots, t_{p_{jN}}) \), where \( t_{p_{j1}} \) is a value representing the relationship of the term \( t_j \) with the class \( p_i \).

\[
 w(t_{p_{j1}}) = \sum_{k:d_{kj}\in P_i} \log_2 \left( 1 + \frac{f_{kj}}{\text{len}(d_k)} \right)
\]

These word vectors are built at specific chunks in training.

3. Multi Resolution Representation (MulR)

Single Resolution is analogous to the BoVW

1) Description of training visual regions

2) Learning the codebook

3) Representing images by replacing patches with the id of the closest VW in the codebook.

Intuitive idea of Multi-Resolutions

4. Evaluation

Sexual Predator Detection

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

Depression Detection

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