Towards Automatic Ranking App Risks via Heterogeneous Privacy Indicators

[1]Deguang Kong, [2]Lei Cen and [1]Hongxia Jin

[1]Samsung Research America, [2]Purdue University

05/07/2016

Background

- Mobile Applications (Apps)
  - The number of mobile apps has increased dramatically
    - Google Play: over 1 million Apps, over 50 billion downloads in July 2013; over 1.2 million Apps in June 2014
  - Apps have played an important role with the popularity of smart phones

Background

- Severe threats to cyber security
  - Macfee: 82% of the apps track user’s information; 80% of the apps collect location information
  - G DATA: on Android devices, 440,267 new malware samples in the first quarter of 2015

Motivation

□ How to identify the security and privacy risks of mobile apps?

Solutions

Google (User’s responsibility)

- Users approve permissions for security
- Bounce (static/dynamic analysis on malicious apps)

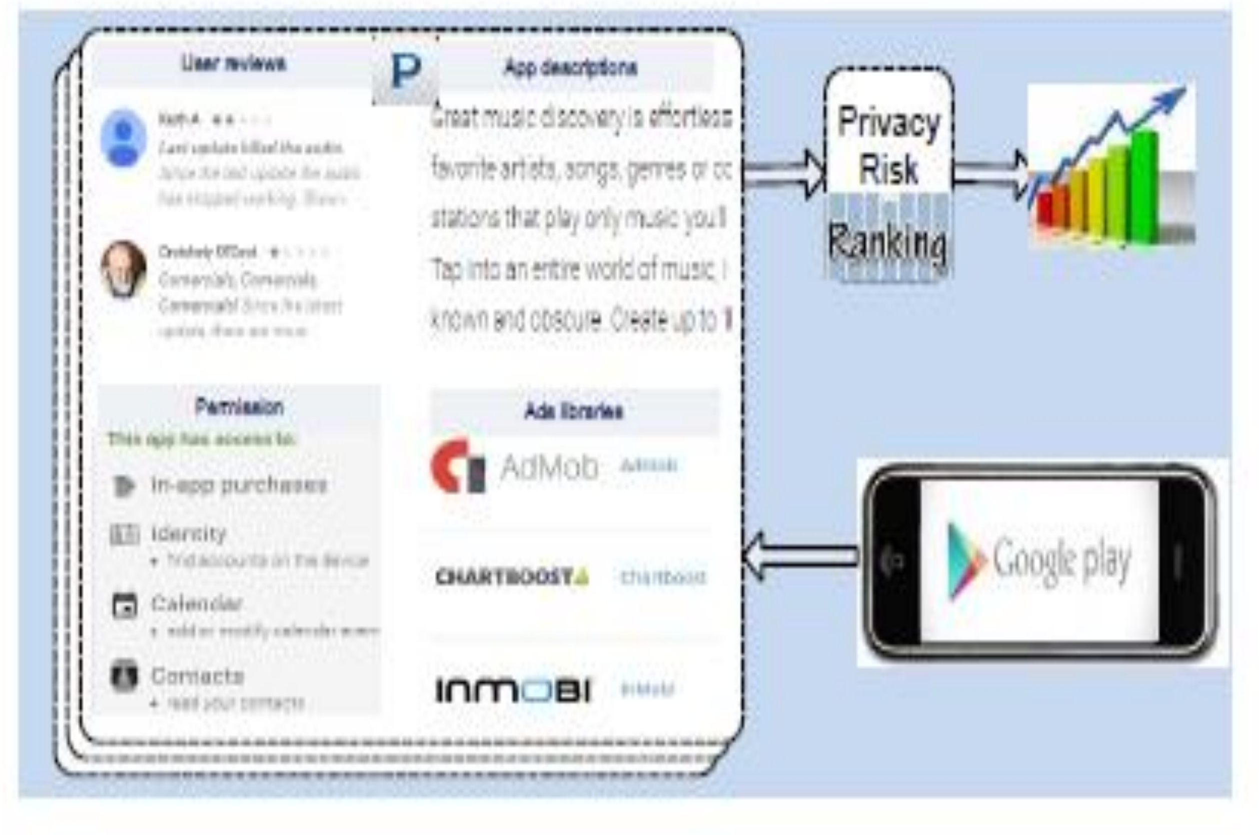
Apple (Market’s responsibility)

- Apple performs manual inspection

- Not enough security/privacy awareness
- Not user-friendly

Method

- How to identify the security and privacy risks of mobile apps?


User reviews
P
App descriptions
Great music discovery is effortless
favorite artists, songs, genres or co
stations that play only music you'll
Tap into an entire world of music, l
known and obscure. Create up to 1
Permission
This app has access to:
In-app purchases
Identity
Calendar
Contacts
Ads libraries
AdMob
CHARTBOOST
InMOBI
Privacy
Risk
Ranking
Google play

- Ranking the risks of mobile apps using app meta data
  - description,
  - user review
  - permission access
  - ads library.
- A ranking model is proposed to capture the relations between the ranking score and privacy indicators.

Approach

- Key idea: ranking the apps from labeled apps to unlabeled apps based on label propagation

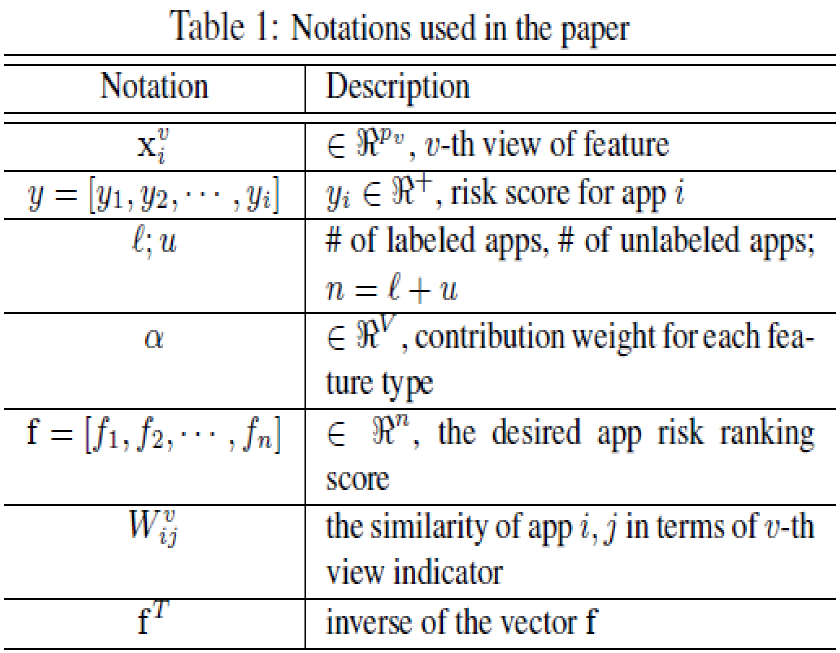
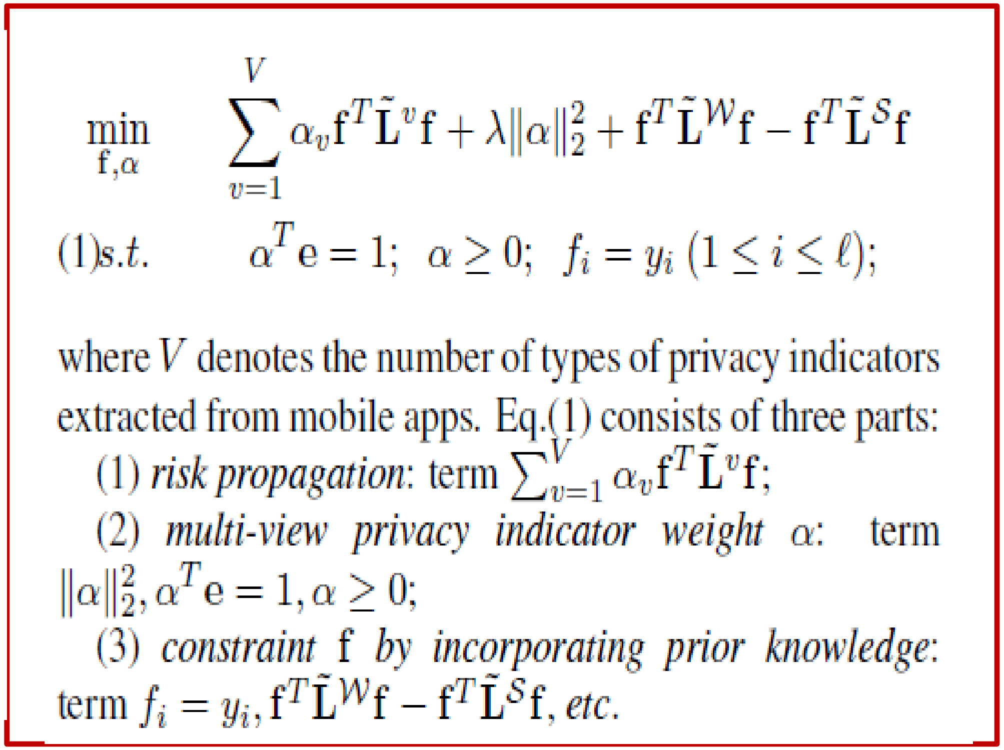


Table 1: Notations used in the paper

Notation	Description
$\mathbf{x}_i^v$	$\in \Re^{p_v}$, v-th view of feature
$y = [y_1, y_2, \cdots, y_i]$	$y_i \in \Re^+$, risk score for app i
$\ell; u$	# of labeled apps, # of unlabeled apps; $n = \ell + u$
α	$\in \Re^V$, contribution weight for each feature type
$\mathbf{f} = [f_1, f_2, \cdots, f_n]$	$\in \Re^n$, the desired app risk ranking score
W_{ij}^v	the similarity of app i, j in terms of v-th view indicator
$\mathbf{f}^T$	inverse of the vector f



$$\min_{\mathbf{f},\alpha} \quad \sum_{v=1}^{V} \alpha_v \mathbf{f}^T \tilde{\mathbf{L}}^v \mathbf{f} + \lambda \|\alpha\|_2^2 + \mathbf{f}^T \tilde{\mathbf{L}}^W \mathbf{f} - \mathbf{f}^T \tilde{\mathbf{L}}^S \mathbf{f}$$

$$(1) s.t. \quad \alpha^T \mathbf{e} = 1; \;\; \alpha \geq 0; \;\; f_i = y_i \; (1 \leq i \leq \ell);$$

where V denotes the number of types of privacy indicators extracted from mobile apps. Eq.(1) consists of three parts:

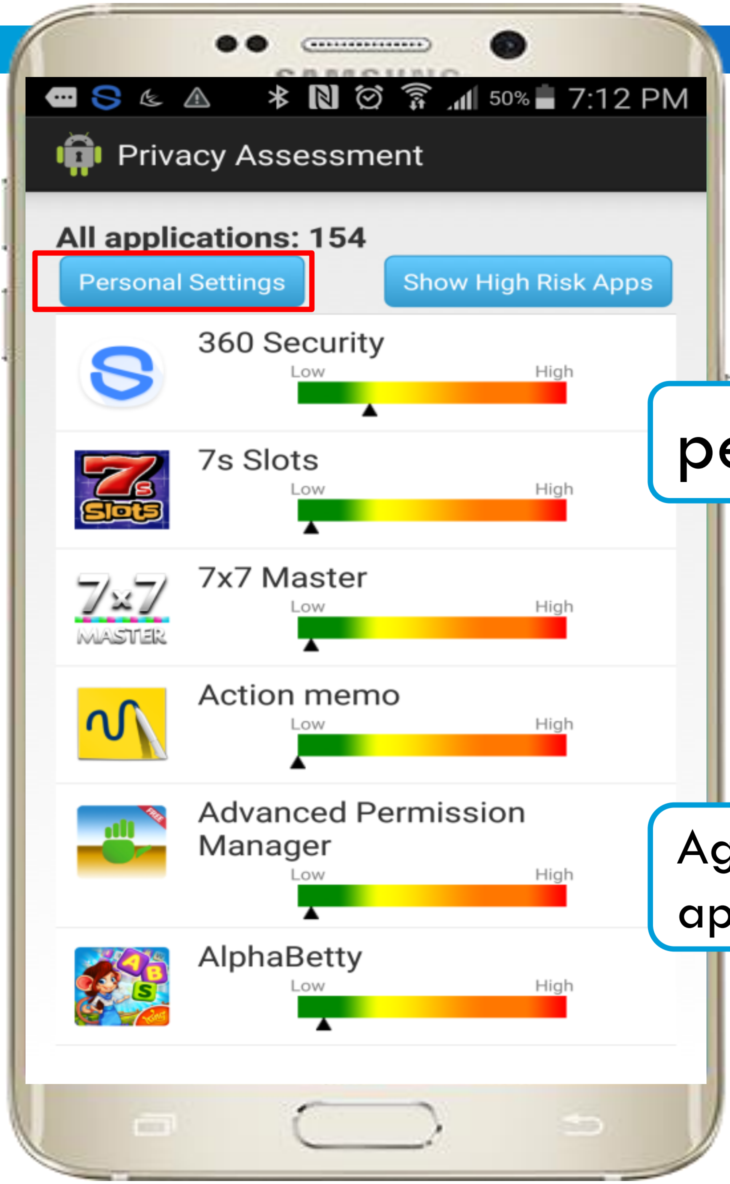
(1) *risk propagation*: term $\sum_{v=1}^{V} \alpha_v \mathbf{f}^T \tilde{\mathbf{L}}^v \mathbf{f}$;

(2) *multi-view privacy indicator weight* α: term $\|\alpha\|_2^2, \alpha^T \mathbf{e} = 1, \alpha \geq 0$;

(3) *constraint* f *by incorporating prior knowledge*: term $f_i = y_i, \mathbf{f}^T \tilde{\mathbf{L}}^W \mathbf{f} - \mathbf{f}^T \tilde{\mathbf{L}}^S \mathbf{f}$, *etc.*

Demo

Demo



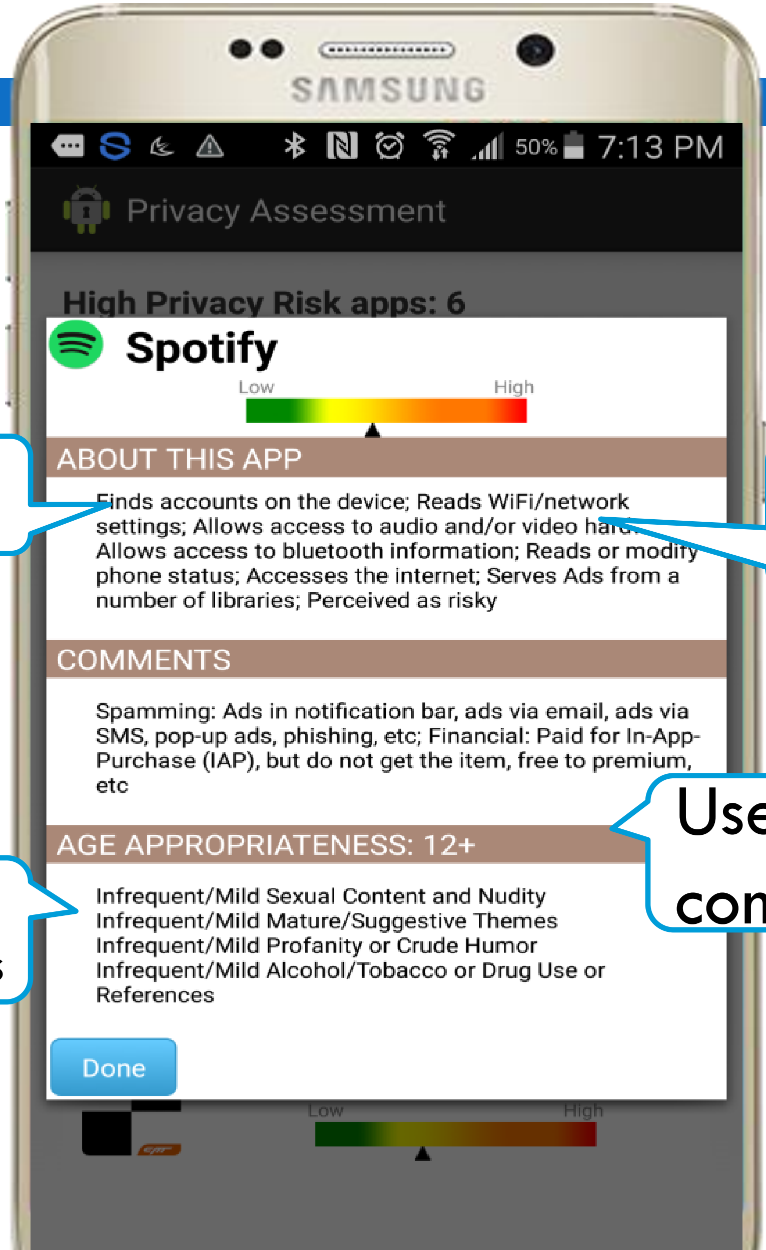
Privacy Assessment

All applications: 154

- 360 Security
- 7s Slots
- 7x7 Master
- Action memo
- Advanced Permission Manager
- AlphaBetty

permission

Age appropriateness



Privacy Assessment

High Privacy Risk apps: 6

Spotify

ABOUT THIS APP

Finds accounts on the device; Reads WiFi/network settings; Allows access to audio and/or video har Allows access to bluetooth information; Reads or modify phone status; Accesses the internet; Serves Ads from a number of libraries; Perceived as risky

COMMENTS

Spamming: Ads in notification bar, ads via email, ads via SMS, pop-up ads, phishing, etc; Financial: Paid for In-App-Purchase (IAP), but do not get the item, free to premium, etc

AGE APPROPRIATENESS: 12+

Infrequent/Mild Sexual Content and Nudity
Infrequent/Mild Mature/Suggestive Themes
Infrequent/Mild Profanity or Crude Humor
Infrequent/Mild Alcohol/Tobacco or Drug Use or References

Ads library

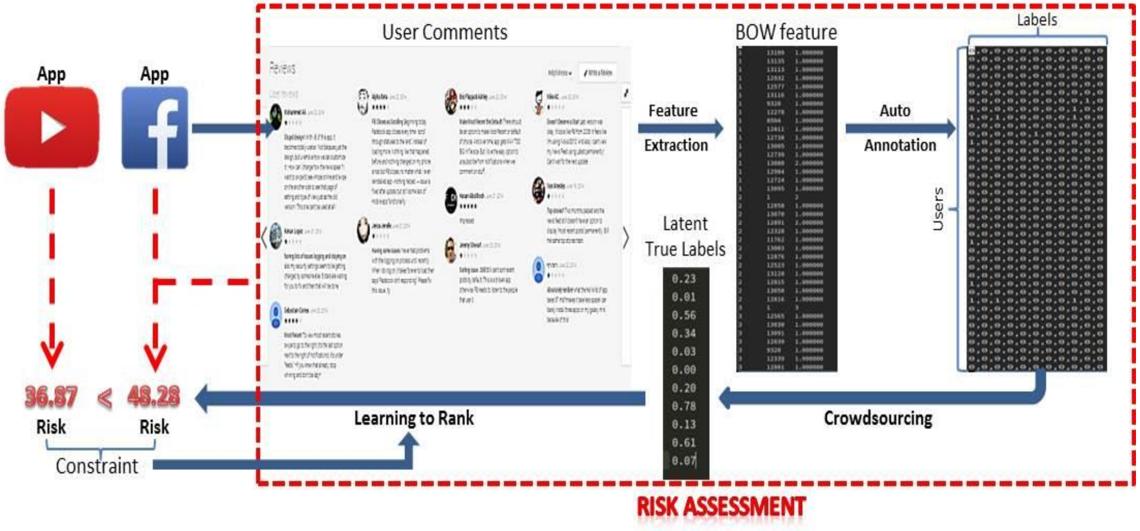
User comments

Other Related Works

Paper title	Venues
Protecting Your Children from Inappropriate Content in Mobile Apps: An Automatic Maturity Rating Framework	ACM CIKM’2015
AUTOREB: Automatically Understanding the Review-to-Behavior Fidelity in Android Applications	ACM CCS’2015
Mobile App Security Risk Assessment: A Crowdsourcing Ranking Approach from User Comments	SIAM DM’2015
Towards Permission Request Prediction on Mobile Apps via Structure Feature Learning	SIAM DM’2015
Personalized Mobile App Recommendation: Reconciling App Functionality and User Privacy Preference	ACM WSDM’2015
PinPlace: associate semantic meanings with indoor locations without active fingerprinting	ACM Ubicomp’2015

Mobile App Security Risk Assessment: A Crowdsourcing Ranking Approach from User Comments (SDM’15)

- Motivation
  - How to rank the privacy risks of mobile apps?
- Our approach
  - Use crowdsourcing to accumulate user comments into app-level features (“feature extraction” → “auto annotation” → “crowdsourcing”)
  - Use “learning to rank” model to predict risk scores by utilizing these latent features while enforcing pairwise constraints



Personalized Mobile App Recommendation: reconciling app functionality and user privacy preferences (WSDM'15)

Motivation

- Mobile app recommendation for users by considering apps' privacy concerns

Our method

- Quantify the tradeoff between App's functionality and user's privacy preference
- Leveraging Poisson Matrix Factorization for recommendation tasks

User i's overall preference for App j

$$g_{i,j} = \underbrace{g_{\text{func},i,j}}_{\text{functionality match score}} + \underbrace{\lambda g_{\text{privacy},i,j}}_{\text{privacy respect score}}$$

Privacy Risk

	App 1	App 2	App …	App m
User 1	2	?	4	2
User 2	3	5	?	?
User …	?	1	3	?
User n	4	?	?	3

Privacy Concern

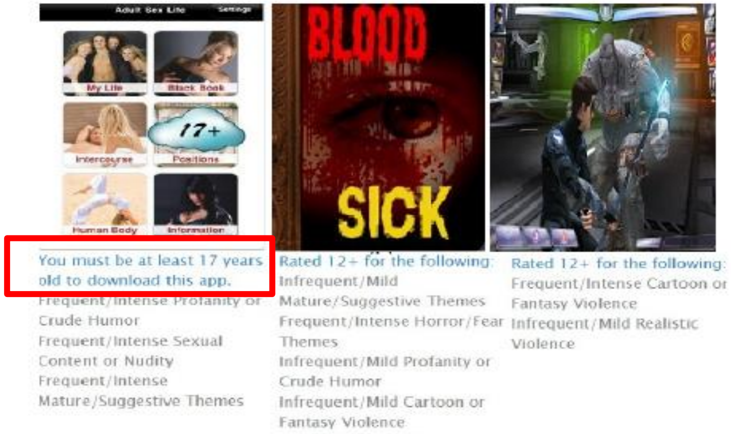
Protecting Your Children from Inappropriate Content in Mobile Apps: An Automatic Maturity Rating Framework (CIKM’15)

Motivation

- Maturity contents such as violence, drug use, etc. may harm children or adolescents
- Predict maturity levels for mobile Apps and the associated reasons with high accuracy and low cost

Our approach

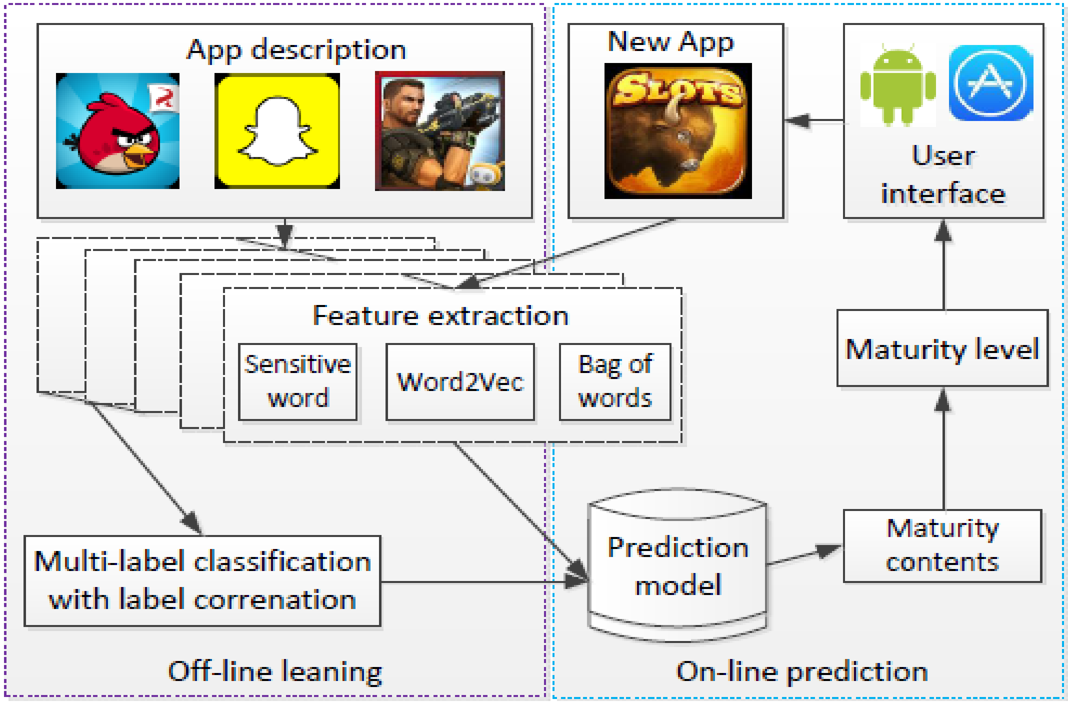
- Feature learning
- Predictive modeling



You must be at least 17 years old to download this app.
Frequent/Intense Profanity or Crude Humor
Frequent/Intense Sexual Content or Nudity
Frequent/Intense Mature/Suggestive Themes

Rated 12+ for the following:
Infrequent/Mild Mature/Suggestive Themes
Frequent/Intense Horror/Fear Themes
Infrequent/Mild Profanity or Crude Humor
Infrequent/Mild Cartoon or Fantasy Violence

Rated 12+ for the following:
Frequent/Intense Cartoon or Fantasy Violence
Infrequent/Mild Realistic Violence



Thank you

- Thanks to all the contributors from SRA.