香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

OASIS: Prioritizing Static Analysis Warnings for Android Apps Based on App User Reviews

Lili Wei

Yepang Liu

Shing-Chi Cheung

Hong Kong University of Science and Technology

September 8th, 2017

Static Analyzers

Lint: Android Studio built-in static analyzer recommended by Google

- Anki-Android (36,801 items)
  - **Android** (11 items)
  - **Android > Lint > Accessibility** (6 items)
  - **Android > Lint > Correctness** (161 items)
  - **Android > Lint > Correctness > Messages** (576 items)
  - **Android > Lint > Internationalization** (3 items)
  - **Android > Lint > Internationalization > Bidirectional Text** (69 items)
  - **Android > Lint > Performance** (1,710 items)
  - **Android > Lint > Security** (2 items)
  - **Android > Lint > Usability** (10 items)
  - ⋮

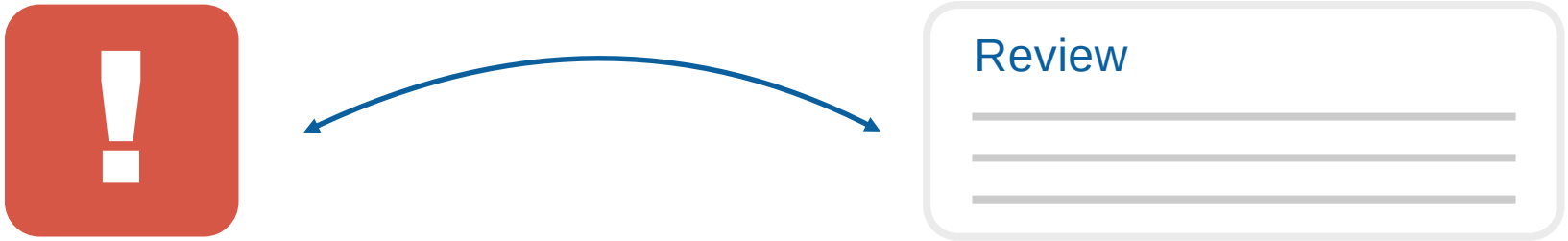
Which of them are positive?

Identifying Positive Warnings

OASIS: pri**O**ritize w**A**rnings based on u**S**er rev**I**ew**S**

OASIS - Key Idea

Warnings and user reviews are intrinsically correlated



Warnings

Possible problems of Android apps

Structured

User Reviews

Comments on problems of Android apps

Unstructured

- A warning should be ranked higher if its described issue can cause user-perceivable problems

How to link the structured warnings with unstructured app user reviews?

- Estimate the user-perceivable problems caused by a warning by recovering links between the warnings and user reviews

Motivating Example

Warning

File: AndroidManifest.xml
Line: 221
Severity: *WARNING*

Description:

Use of `android.hardware.action.NEW_PICTURE` is deprecated for all apps starting with the N release independent of the target SDK.

Review 1

Automatic uploads always fail

Instant upload not working after updating to v7.0…

Review 2

Instant upload doesn't work in nougat

Review 3

…the most important part of the app is instant download and it doesn't work on Android N...

The warning and the reviews are extracted from OwnCloud (https://github.com/owncloud/android)

Motivating Example

➢ Links between warnings and user reviews

Warning

File: AndroidManifest.xml
Line: 221
Severity: *WARNING*

Description:

Use of `android.hardware.action.NEW_PICTURE` is deprecated for all apps starting with the N release independent of the target SDK.

Review 1

Automatic uploads always fail

Instant upload not working after updating to v7.0…

Review 2

Instant upload doesn't work in nougat

Review 3

…the most important part of the app is instant download and it doesn't work on **Android N**..

The warning and the reviews are extracted from OwnCloud (https://github.com/owncloud/android)

Motivating Example

➢ Augment warning descriptions

Warning

File: AndroidManifest.xml
Line: 221
Severity: *WARNING*

Description:

Use of `android.hardware.action.NEW_PICTURE` is deprecated for all apps st
independent of the target

InstantUpload

```
219. <receiver
         android:name=".broadcastreceivers.InstantUploadBroadcastReceiver">
220.     <intent-filter>
221.         <action android:name="android.hardware.action.NEW_PICTURE" />
222.         <data android:mimeType="image/*" />
223.     </intent-filter>
```

Review 1

Automatic uploads always fail

Instant upload not working after updating to v7.0…

Review 2

Instant upload doesn't work in nougat

Review 3

…the most important part of the app is instant download and it doesn't work on Android N...

Other contextual information:

➢ Enclosing method and its callers
➢ Corresponding components

The warning and the reviews are extracted from OwnCloud (https://github.com/owncloud/android)

Motivating Example

➢ Handling unstructured user reviews

Warning

File: AndroidManifest.xml
Line: 221
Severity: *WARNING*

Description:

Use of `android.hardware.action.NEW_PICTURE` is deprecated for all apps st
independent of the target

InstantUpload

```
219. <receiver
         android:name=".broadcastreceivers.InstantUploadBroadcastReceiver">
220.     <intent-filter>
221.         <action android:name="android.hardware.action.NEW_PICTURE" />
222.         <data android:mimeType="image/*" />
223.     </intent-filter>
```

Review 1

Automatic uploads always fail

Instant upload not working after updating to v7.0…

Review 2

Instant upload doesn't work in nougat

Review 3

…the most important part of the app is **instant download** and it doesn't work on Android N...

The warning and the reviews are extracted from OwnCloud (https://github.com/owncloud/android)

Motivating Example

```
219. <receiver
         android:name=".broadcastreceivers.InstantUploadBroadcastReceiver">
220.     <intent-filter>
221.         <action android:name="android.hardware.action.NEW_PICTURE" />
222.         <data android:mimeType="image/*" />
223.     </intent-filter>
```

InstantUpload

Review 3

…the most important part of the app is **instant download** and it doesn't work on Android N...

Knowledge Base

- Concept similarity: recover semantic links between warnings and user reviews
  - Map the documents for warnings and user reviews into the concept space
  - Calculate similarity in the concept space
- Microsoft Concept Graph
  - Large public-available knowledge graph
  - Mapping words to their concept categories

The warning and the reviews are extracted from OwnCloud (https://github.com/owncloud/android)

Warning Prioritization

Warning

File: AndroidManifest.xml
Line: 221
Severity: *WARNING*
Description:

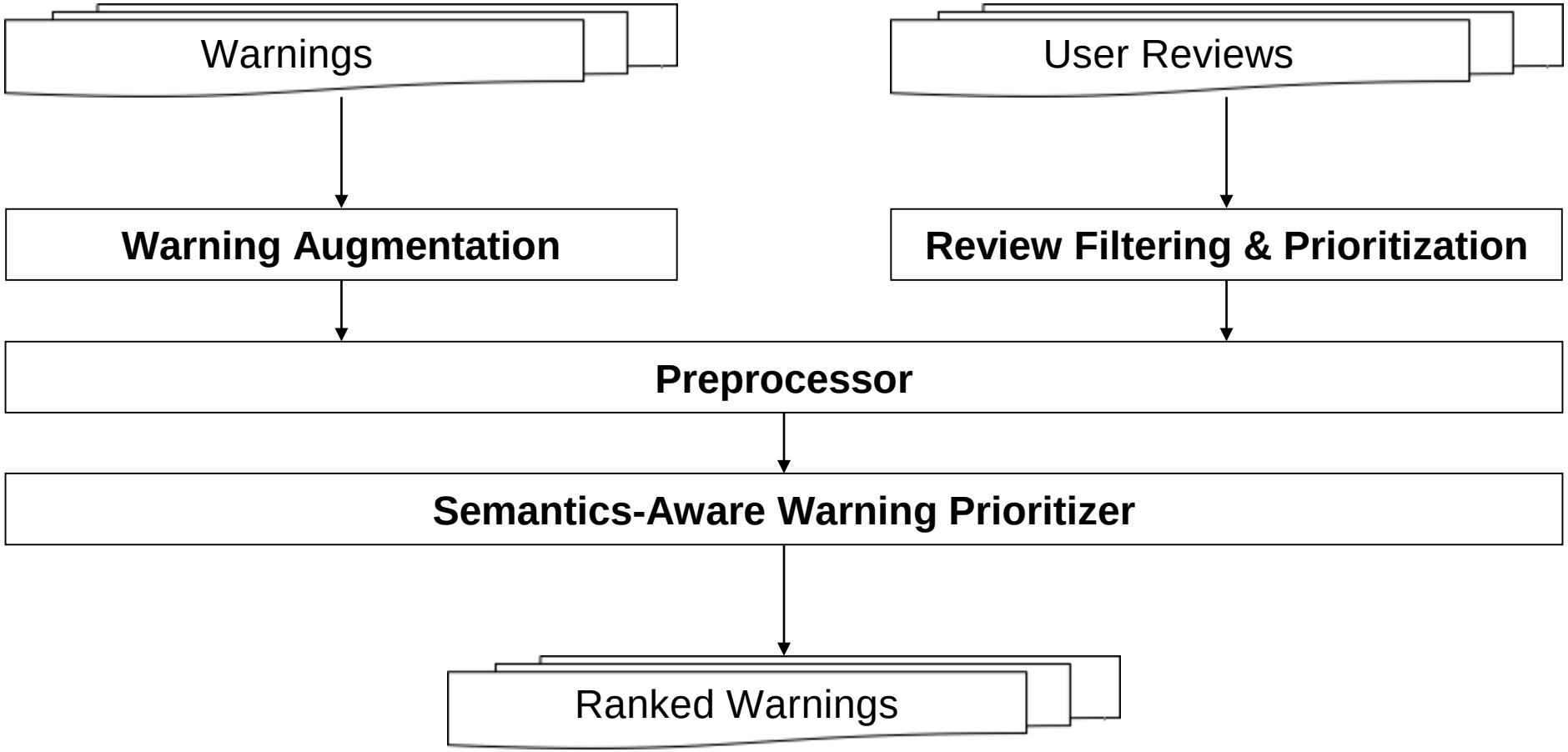
Use of `android.hardware.action.NEW_PICTURE` is deprecated for all apps starting with the N release independent of the target SDK.

- Ranking score of a warning

$$S(w,R) = \sum_{r_i \in R} f(r_i) \times Similarity(w, r_i)$$

  - w: warning
  - r_i: the i-th user review
  - $f(r_i)$: weight of r_i
  - $Similarity(w,r_i)$: textual similarity between w and r_i

OASIS - Overview



Experimental Setup

- Experimental Subjects
  - Large-scale and popular
  - Diversified in category and number of user reviews

App Name	Category	KLOC	Downloads	Rating	# Warnings	# Reviews
AnkiDroid	Education	60.8	1M - 5M	4.5	5,404	3,621
c:geo	Entertainment	83.3	1M - 5M	4.4	10,775	4,871
K-9 Mail	Communication	92.8	5M - 10M	4.2	5,751	8,181
ownCloud	Productivity	53.6	100K - 500K	3.7	2,462	948
TransDroid	Tools	29.9	100K - 500K	4.3	1,508	495
WordPress	Social	144.4	5M - 10M	4.2	8,195	8,851

Experimental Setup

- Evaluation metric
  - *Precision@N:* the percentage of positive warnings among the top N (N = 1, 5, 10. . .) warnings in each ranked list.
- Ground truth
  - Manually inspected the top 50 warnings in each ranked warning list
    - Positive warning
      - True positive
      - Causing user-perceivable problems

Evaluation

- **RQ1: Usefulness of user reviews**
  - Can the use of app user reviews facilitate the identification of positive warnings and improve the usefulness of Lint's static analysis?

Baseline I: Ranking Lint warnings by its **default issue severity settings**

- Warnings of the same severity level are shuffled five times
- The evaluation metrics for Baseline I are calculated as the average of the five shuffled results

RQ1 - Usefulness of User Reviews

Precision@N of OASIS and Baseline I

N	AnkiDroid		K-9 Mail		ownCloud		TransDroid		WordPress		c:geo	
	OASIS	Baseline I	OASIS	Baseline I	OASIS	Baseline I	OASIS	Baseline I	OASIS	Baseline I	OASIS	Baseline I
1	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%	0.0%	0.0%	40.0%	0.0%	0.0%
5	40.0%	8.0%	40.0%	0.0%	80.0%	12.0%	100.0%	0.0%	40.0%	16.0%	0.0%	0.0%
10	50.0%	4.0%	40.0%	0.0%	50.0%	6.0%	70.0%	0.0%	30.0%	14.0%	0.0%	0.0%
20	60.0%	2.0%	25.0%	0.0%	45.0%	9.0%	70.0%	0.0%	45.0%	18.0%	0.0%	0.0%
30	50.0%	2.0%	26.7%	0.0%	50.0%	8.7%	53.3%	0.0%	50.0%	17.3%	0.0%	0.0%
40	47.5%	1.7%	30.0%	0.0%	45.0%	9.0%	52.5%	0.0%	55.0%	18.0%	2.5%	0.0%
50	44.0%	2.8%	28.0%	0.0%	46.0%	9.2%	48.0%	0.0%	54.0%	18.8%	4.0%	0.4%

Evaluation

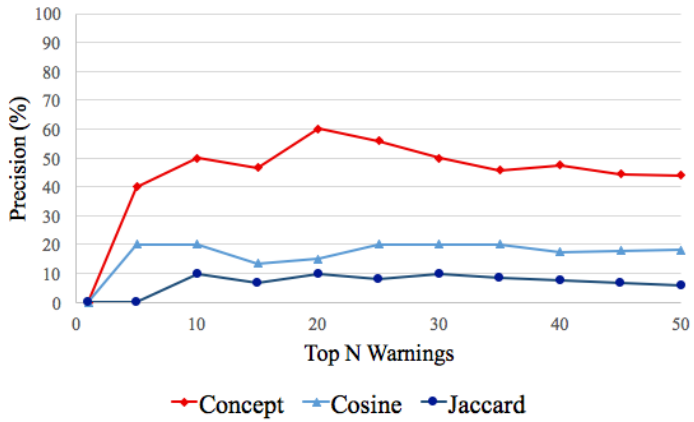
- **RQ2: Effectiveness of concept similarity**
  - Can our proposed concept similarity contribute to OASIS's performance of prioritizing positive warnings?

Baseline II: Ranking Lint warnings by linking warnings with reviews by conventional **token-based similarities** (Jaccard and Cosine)

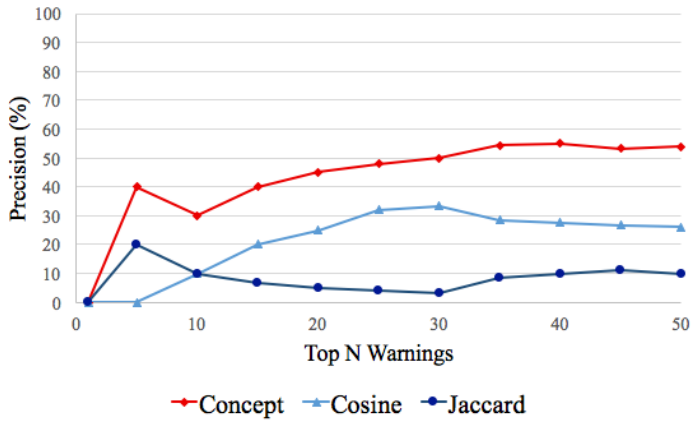
- Steps other than similarity calculation follow those of OASIS

RQ2 - Effectiveness of Concept Similarity

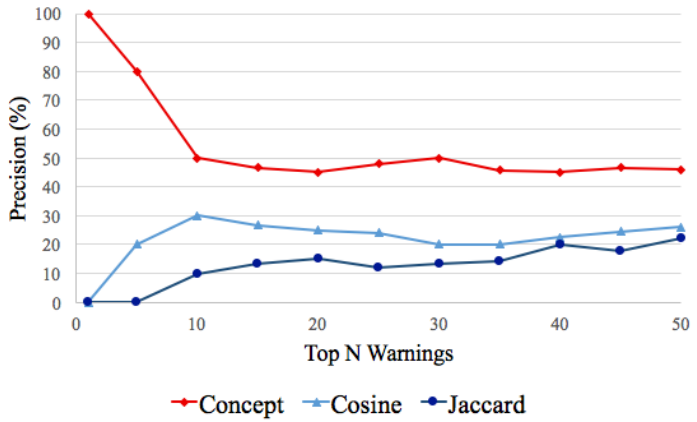
Precision@N of OASIS and Baseline II



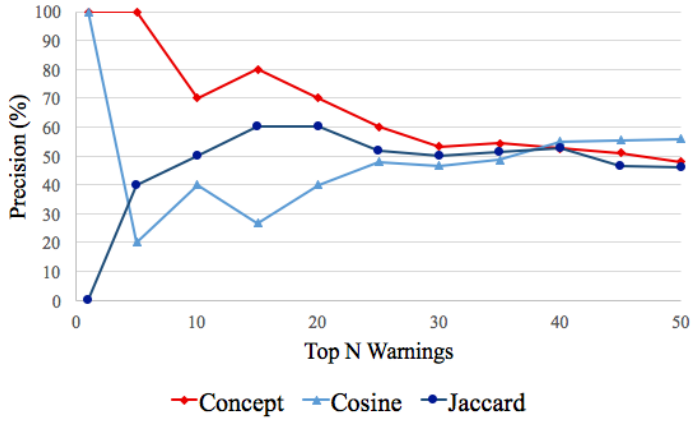
AnkiDroid



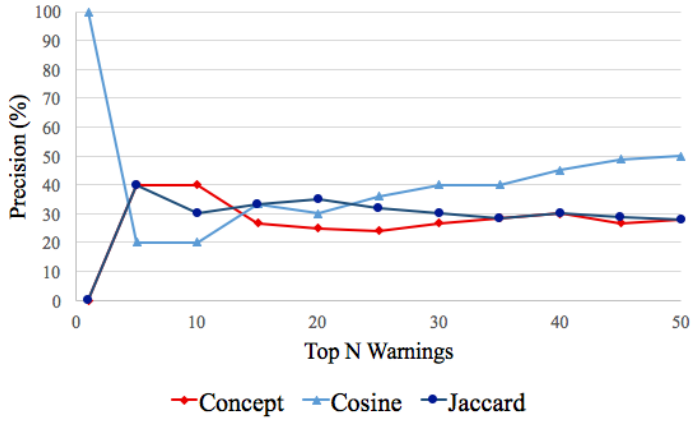
WordPress



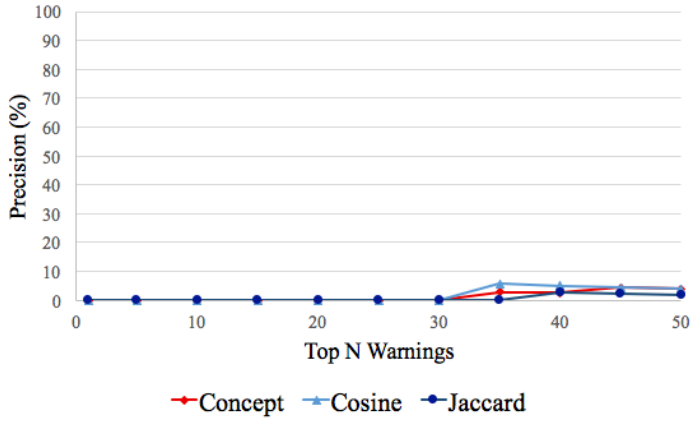
ownCloud



TransDroid



K-9 Mail



c:geo

RQ2 - Effectiveness of Concept Similarity

➢ Top-ranked K-9 Mail warnings

Warning

Description:

“org.apache.http.HttpResponse” is deprecated

Context:

Callers: doInitial**Connection**, **sendMessages**
markServerMessagesRead, etc

Review 1

…Sometimes clicking '**Mark** as **read**' has no effect…

Review 2

Doesn't **connect**…cannot find in and outgoing **servers**…

Review 3

Get a lot of 'Failed to **send** no SMTP'…when **message** fails to **send** it doesn't stay in outbox to retry…

Conclusion

- First technique, OASIS, to prioritize static analysis warnings based on app user reviews
- User reviews are useful for identifying positive warnings
- Leveraging semantics information can improve the performance of OASIS
- Future work
  - Better augment warnings
  - Improve the similarity calculation between warnings and user reviews

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Thank you

Tool and dataset:

http://sccpu2.cse.ust.hk/oasis/

CASTLE