

User Review Mining for Assisting App Development

GAO, Cuiyun

Supervisor: Prof. Michael R. Lyu

Aug. 27, 2018



香港中文大學

The Chinese University of Hong Kong

User Reviews



Instagram

Instagram Social



Contains Ads

Editors' Choice

★★★★☆ 66,599,591



★★★★☆ August 7, 2018



Number of people flagged as helpful

Rating and post date

App is absolute trash. Crashes every three seconds on a flagship stock Pixel 2 XL with zero problems with any other quality app.

Review text

User Reviews are Important

3,800,000 apps



Google Play

2,000,000 apps

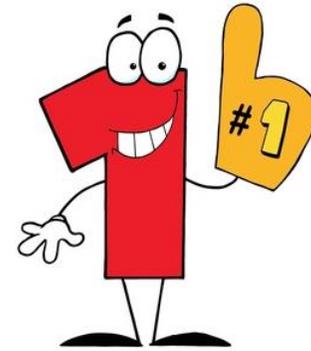


App Store

669,000 apps



Windows Store



\$80,000



Ranking = (# of installs weighted for the past few hours) + (# of installs weighted for the past few days) + **REVIEWS (star rating + number of reviews)** + Engagement (# of times app opened etc.) + Sales (\$)

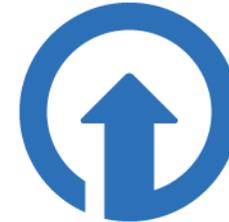
User Review Analysis is Important



Users



App Install
58%



App Update
33%



App Purchase
36%



Developers

User Reviews in App Development



“Optimism is an occupational hazard of programming: **Feedback** is the treatment”
--Kent Beck



- App Bugs
- App Features



Google 用户

★★★★★ 2018年5月14日

Today my timeline is broken. I don't see posts of my friends just some random stuff from unknown sources. Fix it!

App Reviews

User Review Analysis is Challenging

- Massive review **quantity**



60,000 reviews/day



7,000+ reviews/day



2,000+ reviews/day

User Review Analysis is Challenging

- Massive review **quantity**
- Many **non-informative** reviews
- **Short** length and limited context
- Many **noisy** words
- **Detailed app issues** hard to be predefined



Rate	Date	Review Text
1	August 7, 2018	Cant upload stories, itsl just stuck at posting, really bad app
5	August 7, 2018	Lol
4	August 7, 2018	Awsmmmmmm
1	August 7, 2018	Please fix

Traditional Study

- Keyword-based review retrieval
- Filtering out non-informative reviews
- Classifying reviews to predefined topics
- User sentiment prediction
- Device compatibility issue
- ...

Required predefined rules

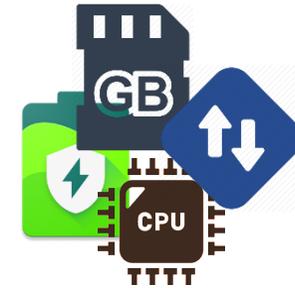
Required labeled reviews

Reduce the manual power!



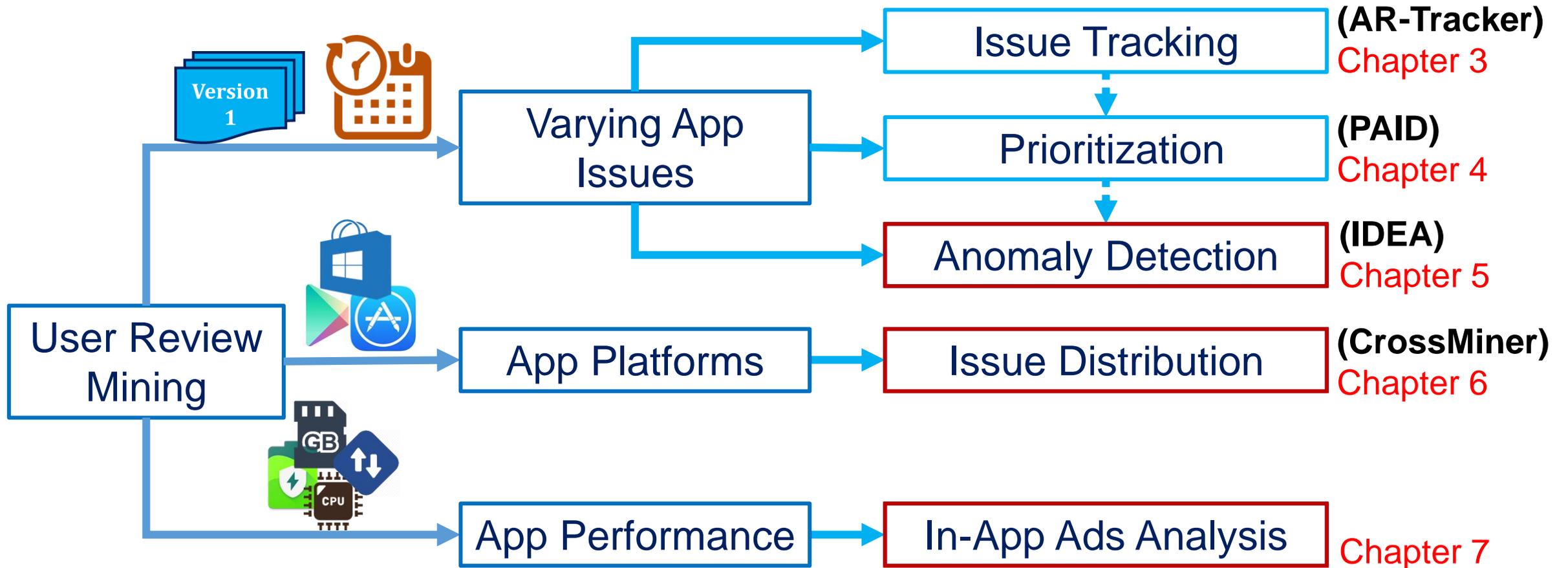
What Traditional Study Ignores

Focus on **static collection**, and ignore the **multiple dimensions** of reviews



★★★★☆ 3.0	Untitled Brief description of the updates would be good. Instead of " A few minor updates to make Twitter an even better place "	Aug 07, 2018	English	7.56.1.1050
★★★★★ 5.0	Untitled A great app overall	Aug 07, 2018	English	7.56.0.1048
★★★★★ 5.0	Untitled Goood	Aug 07, 2018	English	7.51.0
★★★★★ 5.0	Untitled My grandma got a face tatt	Aug 07, 2018	English	7.57.0.1051

Thesis Contributions



Outline

- Topic 1: Online app review analysis for **detecting emerging app issues**
- Topic 2: Issue Prioritization **across different app platforms**
- Topic 3: Exploration on the effects of **in-app ads** on user experience
- Conclusion and future work

Outline

- Topic 1: Online app review analysis for **detecting emerging app issues**
- Topic 2: Issue Prioritization across different app platforms
- Topic 3: Exploration on the effects of in-app ads on user experience
- Conclusion and future work

Outline

- Topic 1: Online app review analysis for **detecting emerging app issues**
 - Motivation
 - Framework of detecting emerging app issues
 - Evaluation
 - Summary

Outline

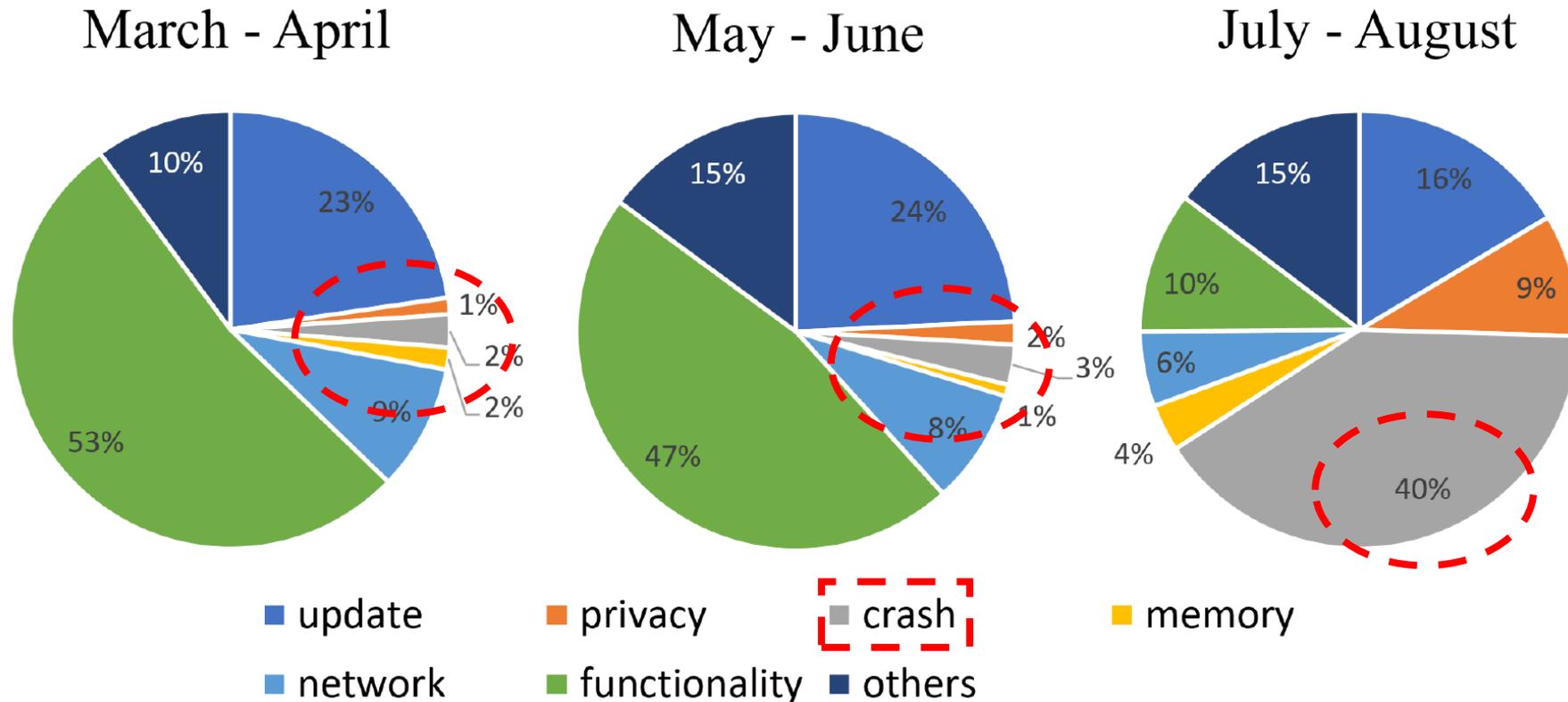
- Topic 1: Online app review analysis for **detecting emerging app issues**
 - Motivation
 - Framework of detecting emerging app issues
 - Evaluation
 - Summary



Emerging Issues

Definition:

An issue **rarely appears** in previous slice but is mentioned by a **significant proportion** of user reviews in current slice.





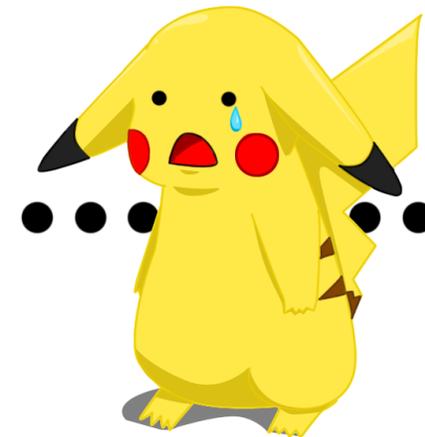
A Real Case for Emerging Issues



From Vocativ.com

Update

Serious!





Reviews can Help

 April 19, 2017
..Unable to download lessons for **offline viewing**....Not being able to view lectures **offline** is very disappointing.

 April 18, 2017
★★★★★
...In this version there is so much confusion about navigation, specifically in **enrollment section**....

WHAT'S NEW

Version 3 of your Udacity App has arrived.

... gets a brand-new overview of your **enrollment**, and streamlined navigation...

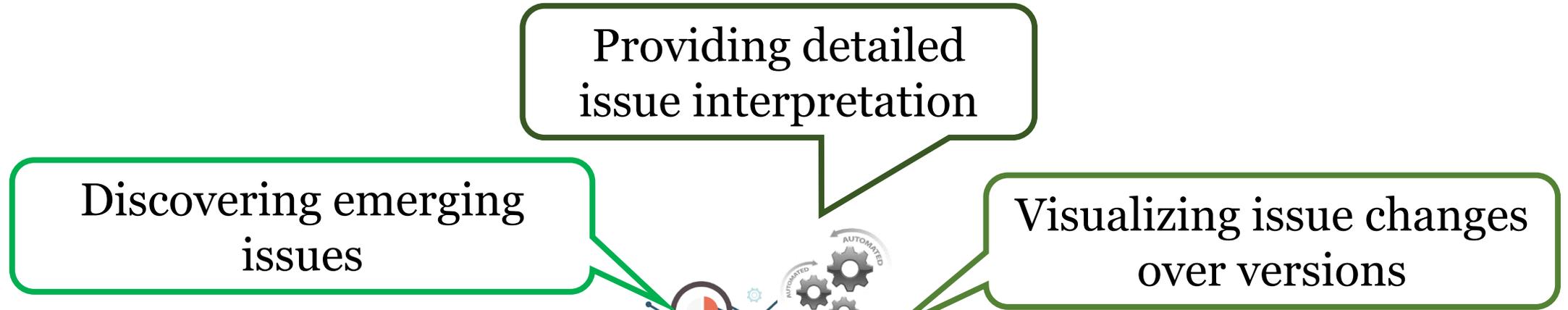
V3.1.x

- **Improved offline experience**

- ...

Changelog

Our Task



Google 用户

★★★★★ 2018年5月14日

Today my timeline is broken. I don't see posts of my friends just some random stuff from unknown sources. Fix it!

App Reviews

IDEA
(**I**dentifying **E**merging **A**pp issues)

WHAT'S NEW

- Version 3 of your Udacity App has arrived!
- Now students have full access to their materials. In addition everyone gets a brand-new overview of your enrollments, and streamlined navigation within your mobile classroom
- v 3.2.0
 - Bug fixes and performance improvements
- v 3.1.x
 - New on-boarding experience
 - Improved offline experience
 - Push notifications enabled for mentors and mentorship's students

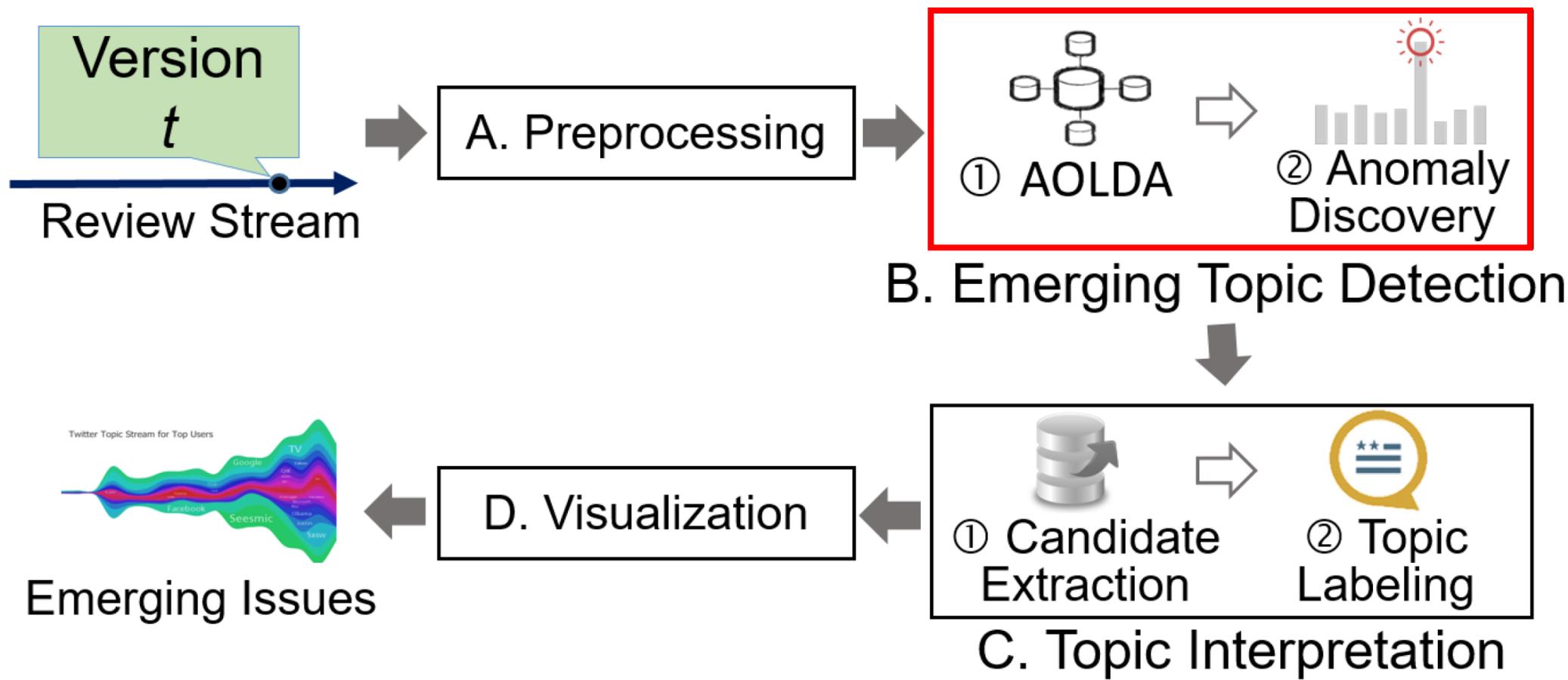
Changelog!

App Updates

Outline

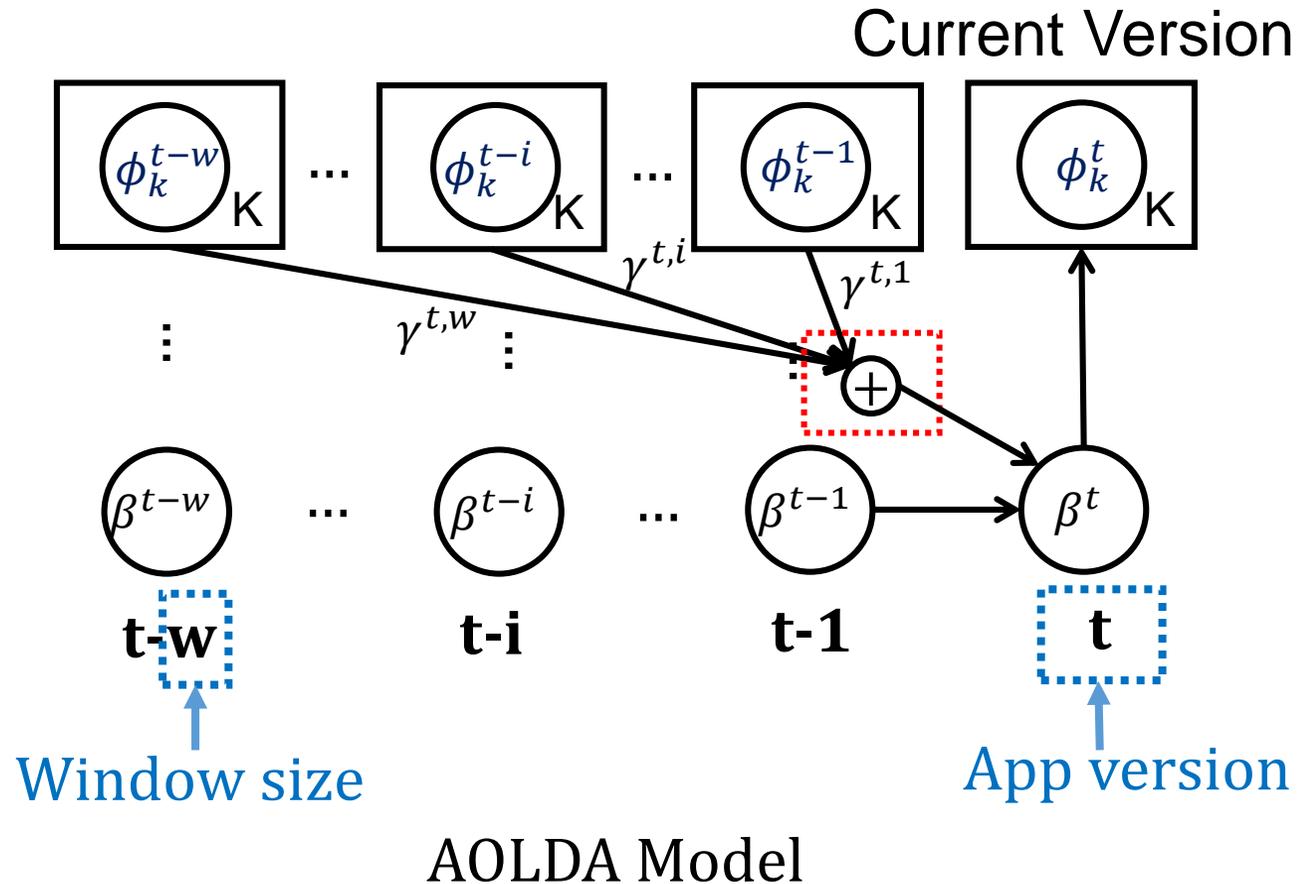
- Topic 1: Online app review analysis for detecting emerging app issues
 - ❑ Motivation
 - ❑ Framework of detecting emerging app issues
 - ❑ Evaluation
 - ❑ Summary

IDEA Framework



Adaptive On-Line LDA

Soft alignment of previous ϕ

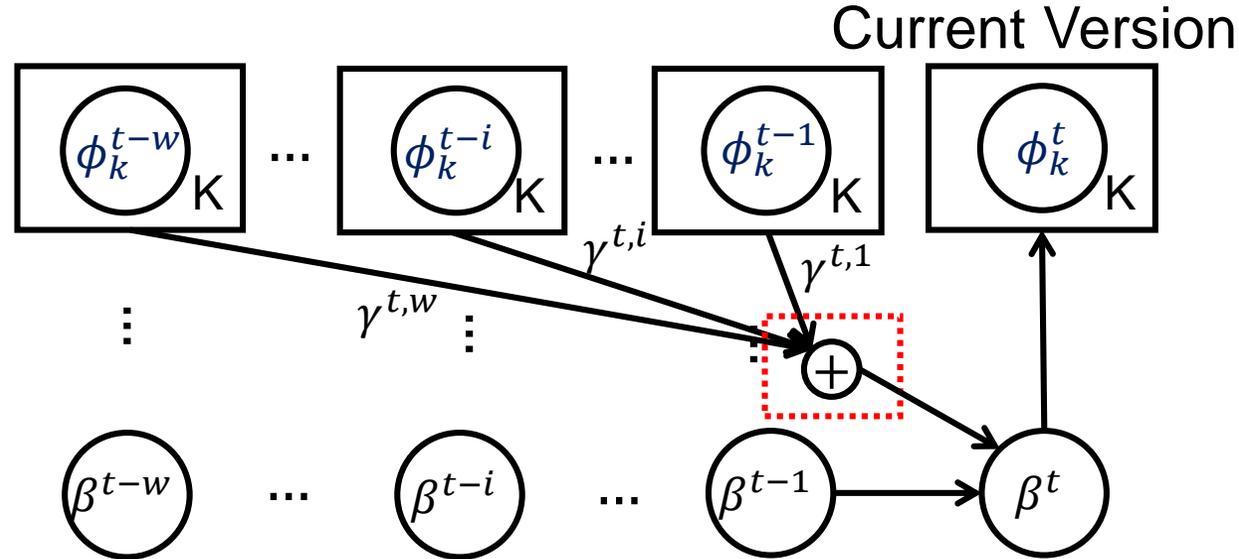


Topic prior Topic distribution

$$\beta_k^t = \sum_{i=1}^w \gamma_k^{t,i} \phi_k^{t-i}$$

$$\gamma_k^{t,i} = \frac{\exp(\phi_k^{t-i} \cdot \beta_k^{t-1})}{\sum_{j=1}^w \exp(\phi_k^{t-j} \cdot \beta_k^{t-1})}$$

Anomaly Discovery



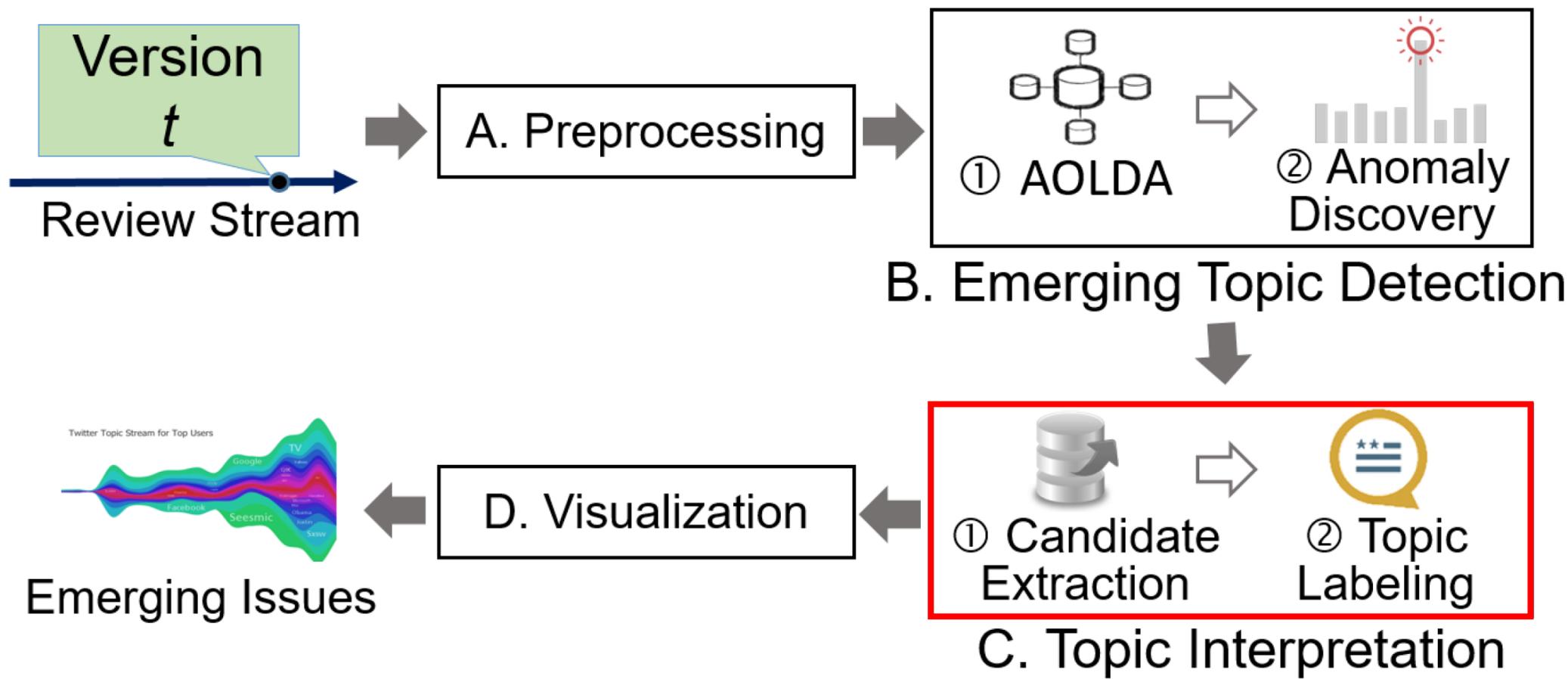
Jensen-Shannon Divergence:

$$D_{JS}(\phi_k^t || \phi_k^{t-1}) = \frac{1}{2} D_{KL}(\phi_k^t || M) + \frac{1}{2} D_{KL}(\phi_k^{t-1} || M)$$

ϕ_k^{t-1}
↓
Topic distribution

$$M = \frac{1}{2} (\phi_k^t + \phi_k^{t-1})$$

IDEA Framework



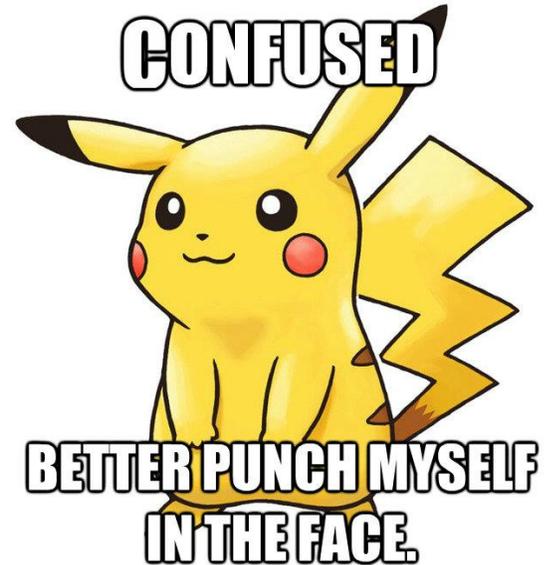


Topic Interpretation

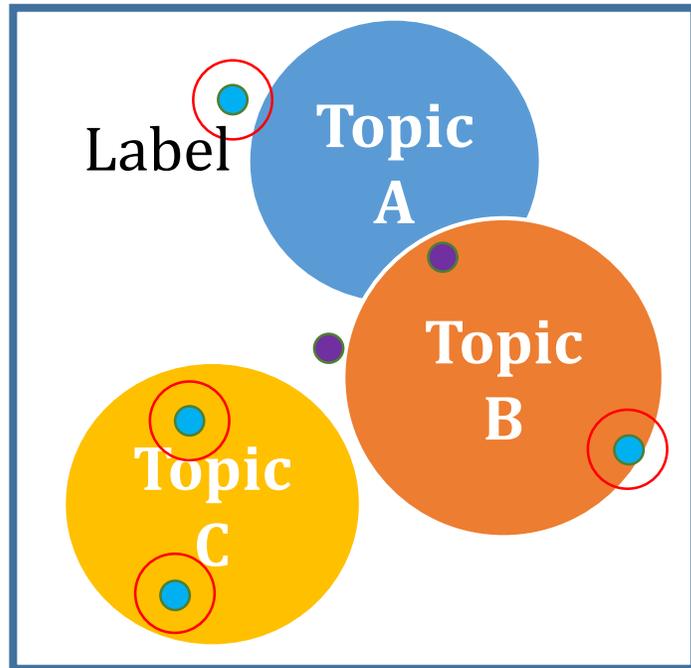
Top words of topic distribution

What is the issue behind those words?

Topic 1	Topic 2	Topic 3	Topic 4
comment	link	back	load
say	video	also	video
reply	open	button	even
try	work	change	work
error	description	go back	take
...



Topic Labeling



Semantic Score

$$Score_{sem}(l, \phi_k^t) = sim(l, \phi_k^t) - \frac{\mu}{K-1} \sum_{j \neq k} sim(l, \phi_j^t)$$

Penalty (points to μ)
Topic distribution (points to ϕ_j^t)

Sentiment Score

$$Score_{sen}(l) = \exp\left(\frac{-r_l}{\log(h_l)}\right)$$

Length (points to h_l)

Total Score

$$Score(l, \phi_k^t) = Score_{sem}(l, \phi_k^t) + \lambda \cdot Score_{sen}(l)$$

Outline

- Topic 1: Online app review analysis for **detecting emerging app issues**
 - ❑ Motivation
 - ❑ Framework of detecting emerging app issues
 - ❑ Evaluation
 - ❑ Summary

Experiment Datasets

App Name	Category	Crawled Platform	#Reviews	#Versions
NOAA Radar	Weather	App Store	8,363	16
YouTube	Multimedia	App Store	37,718	33
Viber	Communication	Google Play	17,126	8
Clean Master	Tools	Google Play	44,327	7
Ebay	Shopping	Google Play	35,483	9
Swiftkey	Productivity	Google Play	21,009	16
Total:			164,026	89

Ground Truth



Version	Date	Changelog
3.6	10-Sep-15	(1) Performance & stability fixes for flawless location search .
3.7	16-Sep-15	This update has a bunch of iOS 9 -specific features for your convenience! (1) Get straight to the weather in your favorite locations as they now appear in Spotlight search results on your device. (2)

Word2Vec [Mikolov et al. 2013]

Sim \geq 0.6 \Rightarrow Hit

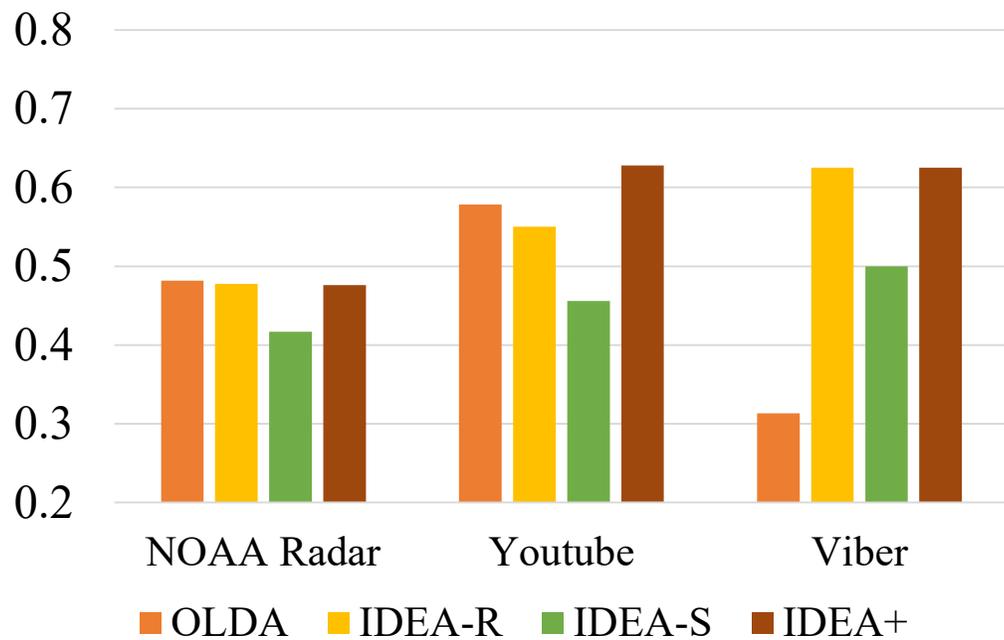
Metric

$$Precision_{Emerge} = \frac{Emerge \cap Changelog}{Emerge}$$

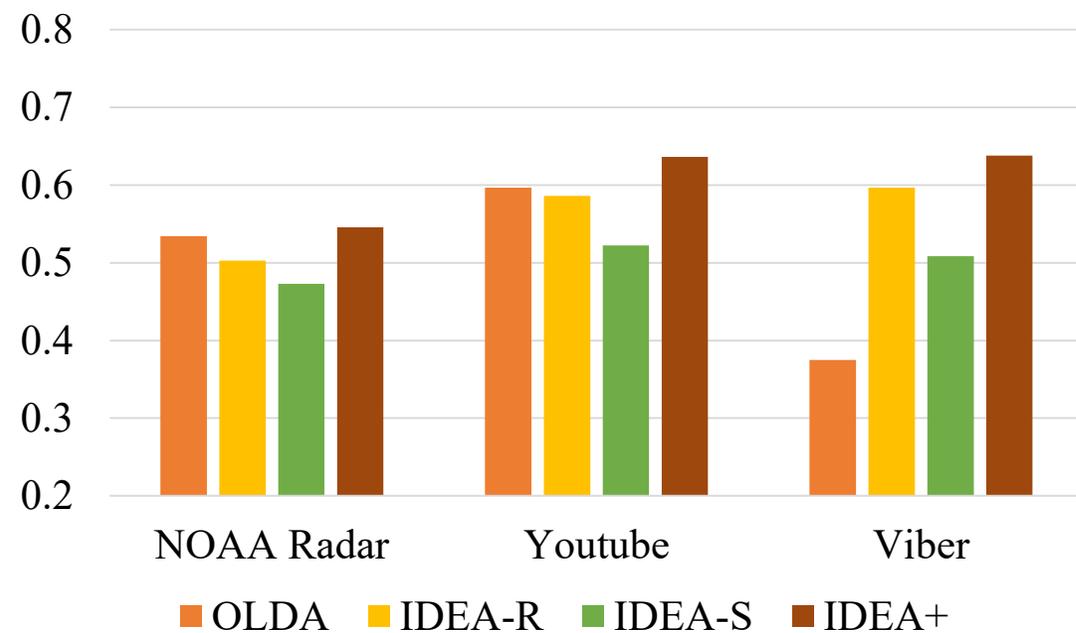
$$Recall_{Label} = \frac{Label \cap Changelog}{Changelog}$$

$$F_{hybrid} = 2 \times \frac{Precision_{Emerge} * Recall_{Label}}{Precision_{Emerge} + Recall_{Label}}$$

Validation



$Precision_{Emerge}$



F_{hybrid}

Precision
60.4%

29.6%

Recall
60.3%

6.3%

F-score
58.5%

23.1%

IDEA's Results



	v11.07	v11.10	Versions
Topic 1	<u>description box: 2.03</u>	say playback error: 1.12	
Topic 2	user interface: 1.25	<u>split screen: 1.23</u>	Emerging Issue

Phrase-level Issues

Sentence-level Issues

	v11.07	v11.10
Topic 1	I mean it work ... you would <u>click a link in the description</u> and it doesn't even let me go through the next video: -0.050843	I can't watch YouTube on my phone because the app won't open and ... doesn't work: -1.03
Topic 2	But right now the lack of multitasking have actually : -0.79	Add <u>split view and slide over</u> but no picture in picture: -1.366723

Industry Practice



4/20+ Tencent apps



Tencent
Bugly Group

□ 500~5,000 reviews per day

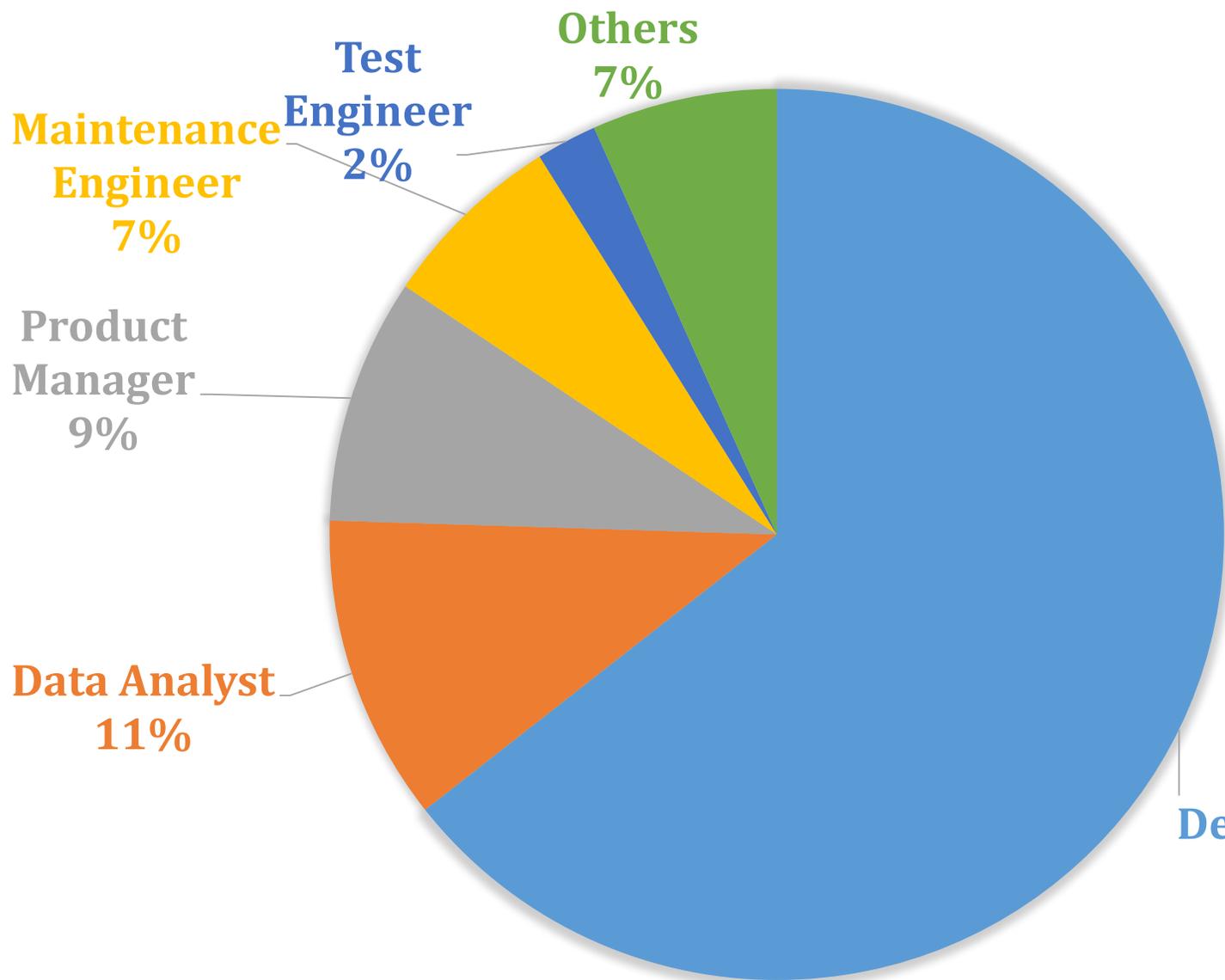


One PC with Intel(R) Xeon
E5-2620v2 CPU (2.10 GHz,
6 cores) and 16GB RAM

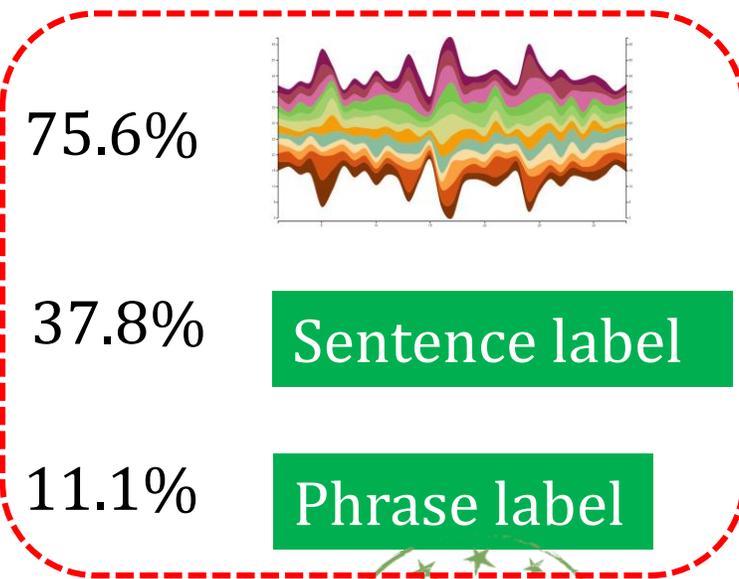
For 36,000 product reviews per
version:

- ✓ ~160 reviews per second
- ✓ 1.02 GB on average

Industry Survey



68.9% Reflect modified issues



75.6%

37.8%

11.1%

Sentence label

Phrase label

88.9%

Performance



Outline

- Topic 1: Online app review analysis for **detecting emerging app issues**
 - ❑ Motivation
 - ❑ Framework of detecting emerging app issues
 - ❑ Evaluation
 - ❑ Summary

Conclusion of Topic 1

- We first study **emerging issues** in app reviews.
- We propose 'IDEA' for **effectively** identifying emerging issues in both **research** and **industry** scenario in an **end-to-end mode**.
- We proposed to use **changelogs** for verification, which can also be used in other verification scenarios.
- We release our code for publicly available.



Outline

- Topic 1: Online app review analysis for detecting emerging app issues
- Topic 2: Issue Prioritization **across different app platforms**
- Topic 3: Exploration on the effects of in-app ads on user experience
- Conclusion and future work

Outline

- Topic 2: Issue Prioritization across different app platforms
 - Motivation
 - Framework of CrossMiner
 - Evaluation
 - Summary

Outline

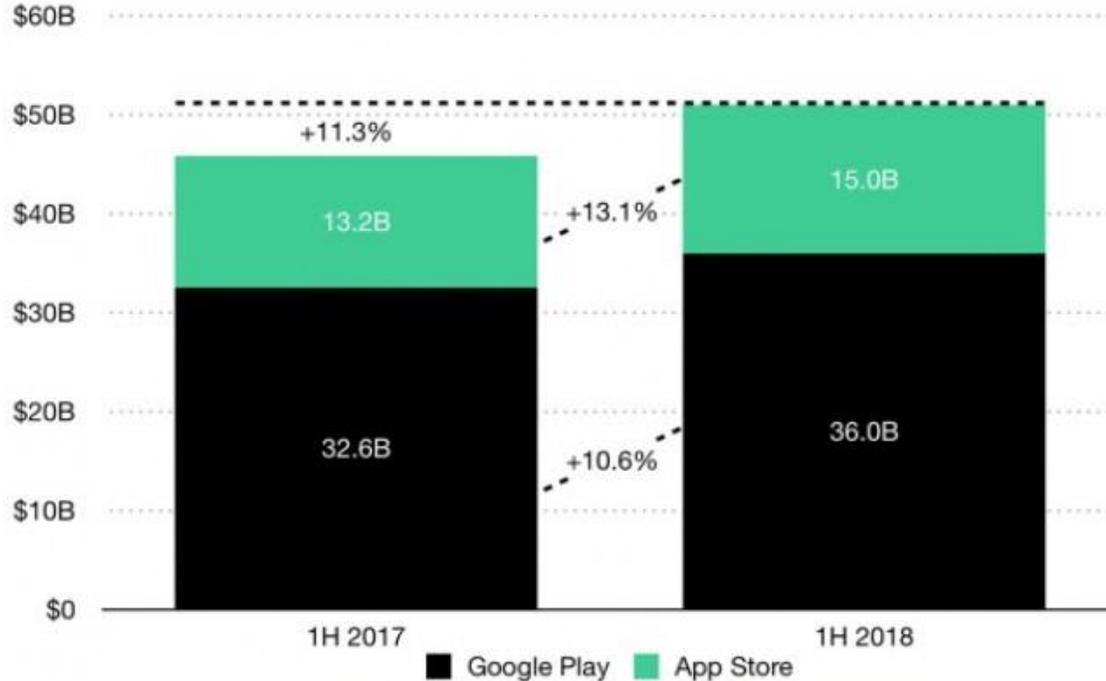
- Topic 2: Issue Prioritization across different app platforms
 - Motivation
 - Framework of CrossMiner
 - Evaluation
 - Summary

Apps Distributed in Different Platforms

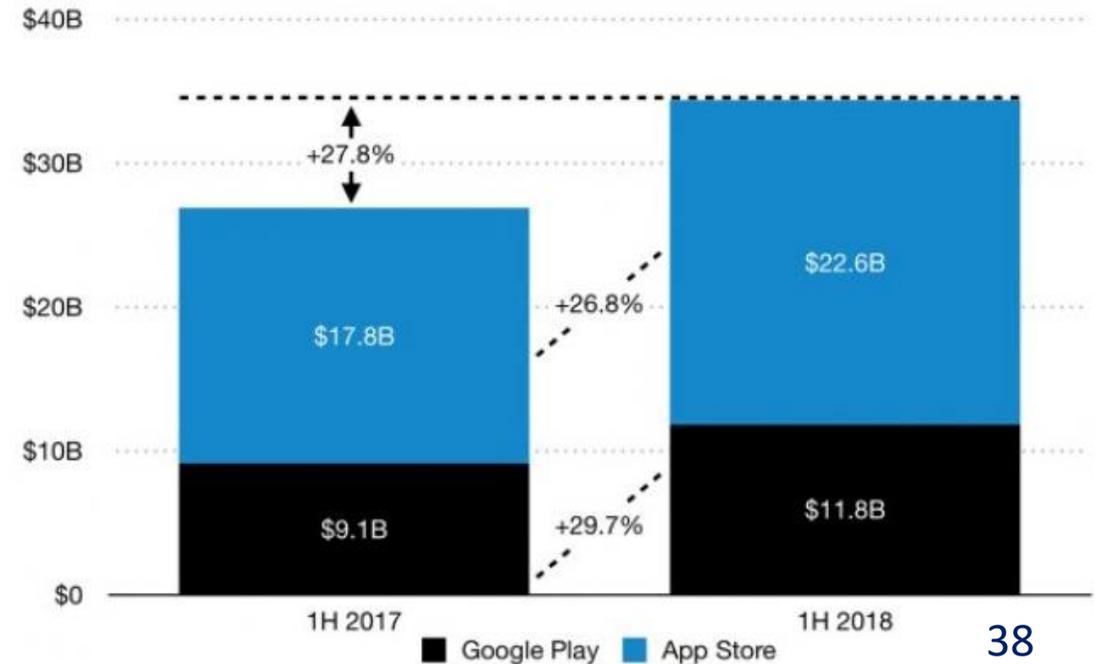


User  Revenue

Worldwide App Downloads - First Half 2018



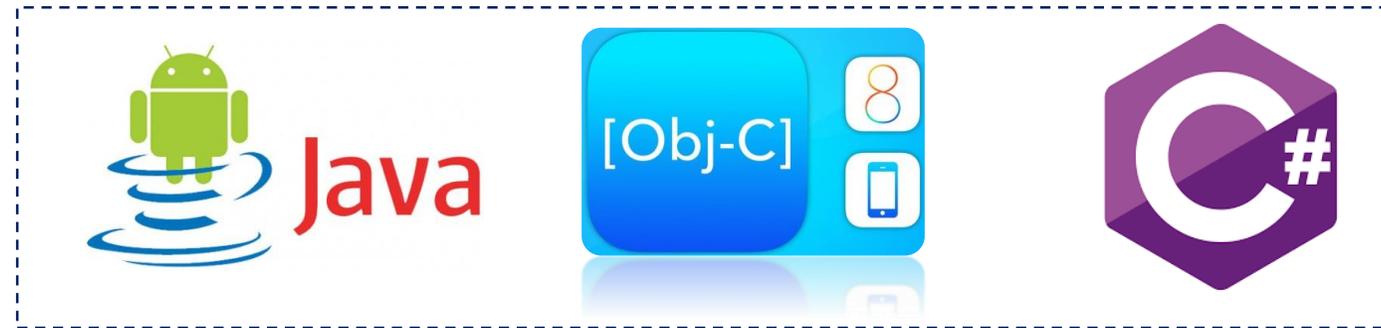
Worldwide Gross App Revenue - First Half 2018



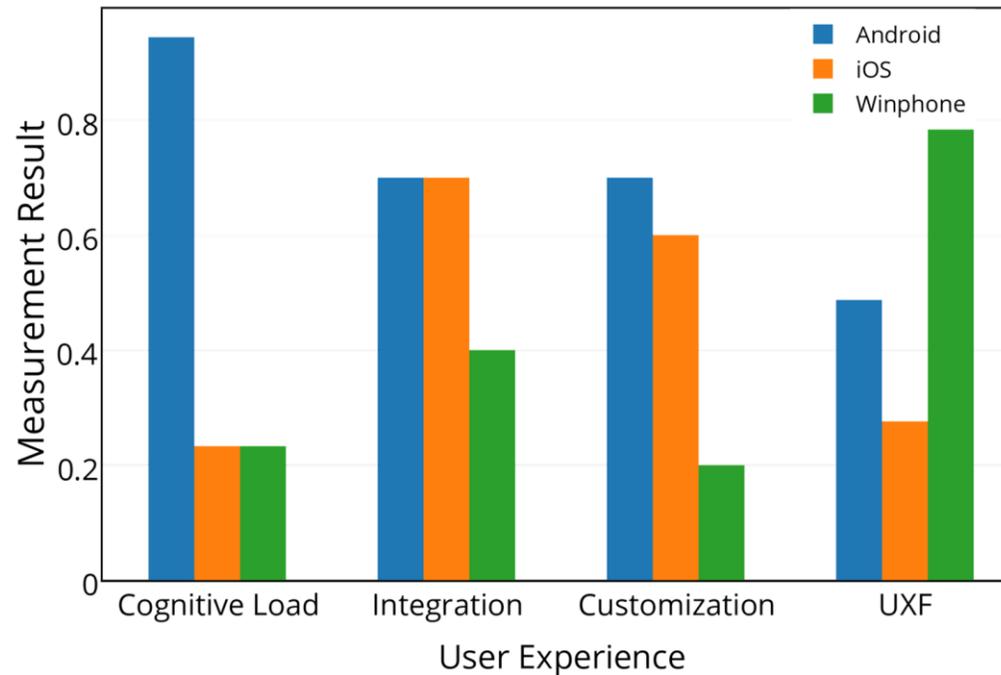


Differences of Platforms

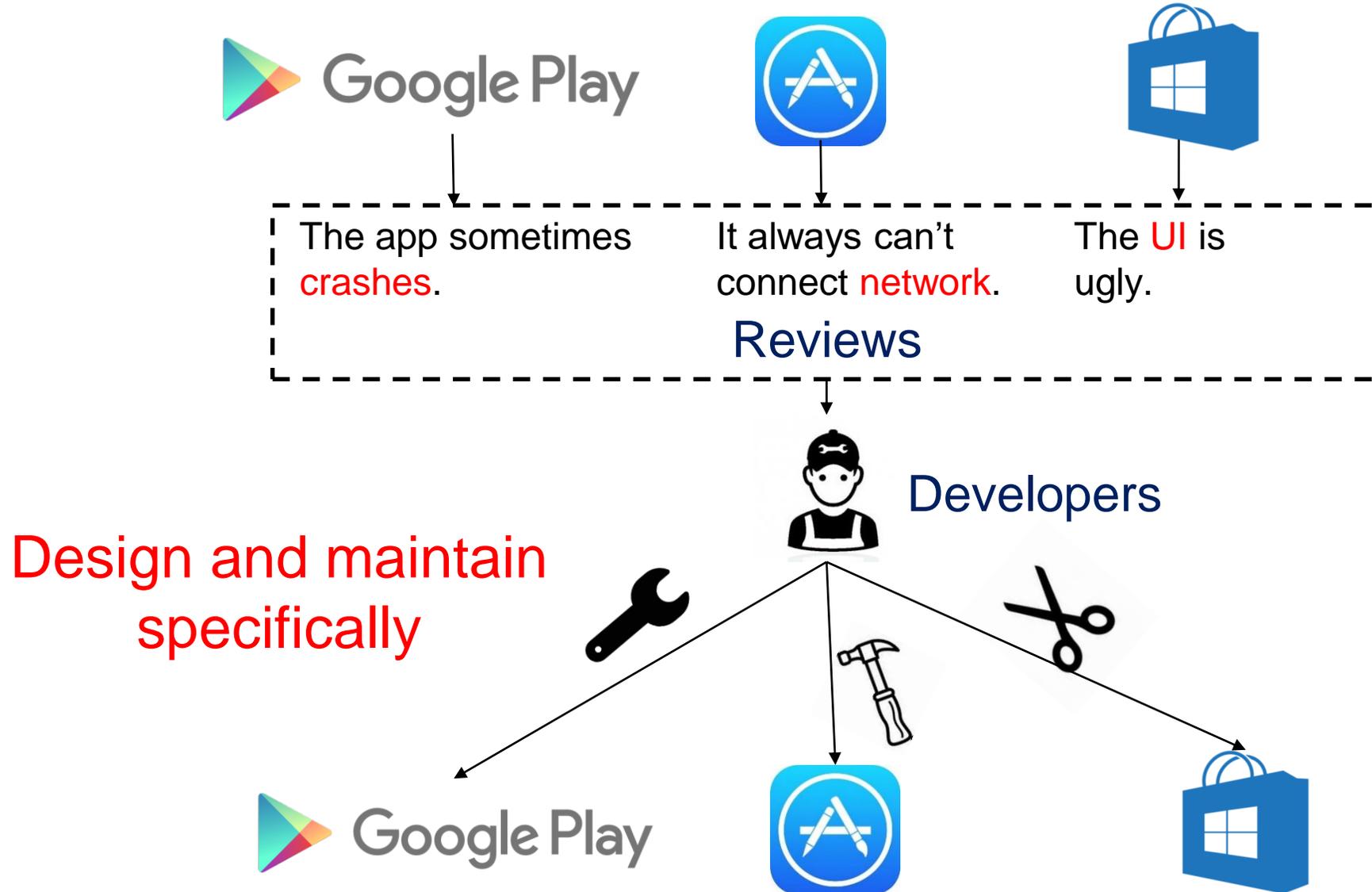
Programming Languages:



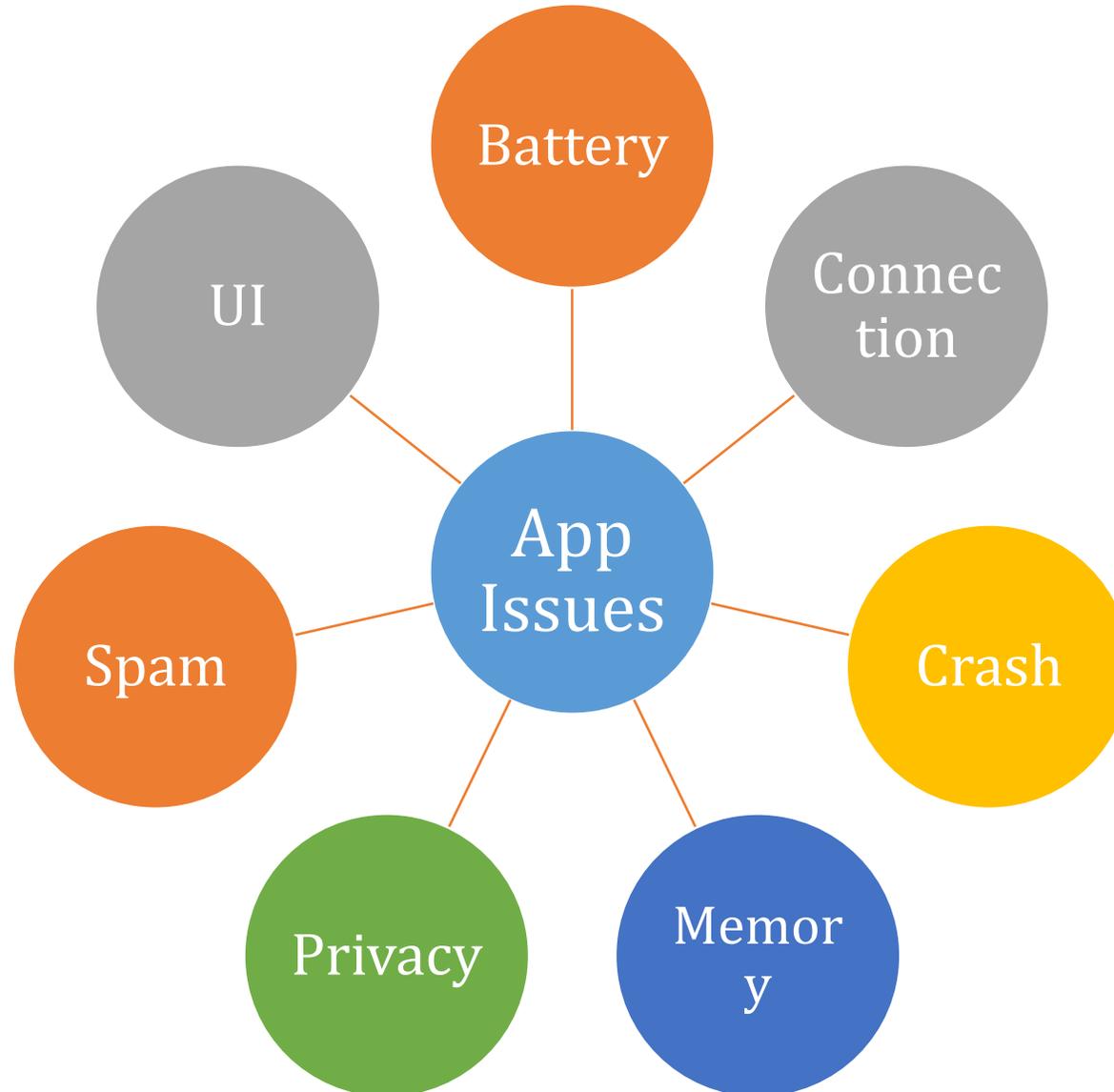
User Experience:



Reviews in Different Platforms



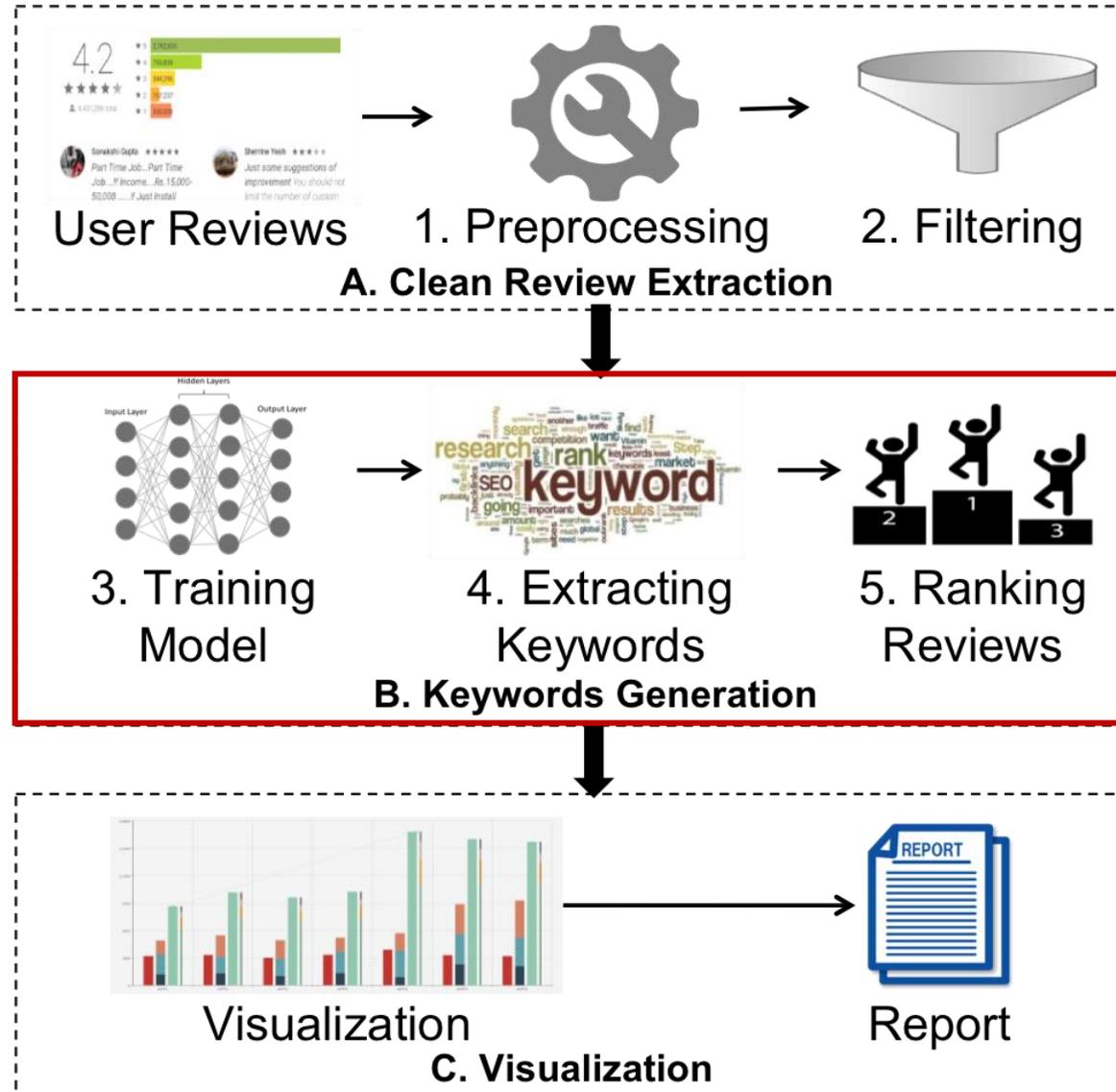
Seven Basic App Issues



Outline

- Topic 2: Issue Prioritization across different app platforms
 - Motivation
 - Framework of CrossMiner
 - Evaluation
 - Summary

CrossMiner Framework



Extracting Keywords

😊 Similar words for each issue based on *word2vec*:

$$similarity = \frac{\sum_{i=1}^n I_i W_i}{\sqrt{\sum_{i=1}^n I_i^2} \sqrt{\sum_{i=1}^n W_i^2}},$$

→ App Issue

→ Examined Word

😊 Combined with clustering method:

Issue	Similar Words
Battery	battery, drain, usage, consumption, overheat, drainer, consume, cpu , power, ram , hog, electricity, drainage, charger, batter, standby, discharge, energy, deplete , memory , foreground

Keyword List

Issue	Keywords
Battery	battery, drain, usage, consumption, overheat, drainer, consume, power, hog, electricity, drainage, charger, batter, standby, discharge, energy
Crash	crash, freeze, foreclose, lag, crush, stall, close, shut, laggy, glitch, hang, load, stuck, startup, buffer, open, laggs, freez, glitchy, buggy
Memory	memory, storage, space, gb, internal, gigabyte, ram, 6gb, occupy, 4gb, mb, 300mb, 8gb, 500mb, 16gb, byte, 5gb, gig, 2gb, 1gb, 1g
Network	network, connectivity, internet, consumption, wifi, connection, reception, conection, connect, signal, 4g, wi, 3g, broadband, fibre, lte, reconnecting, fi, wireless, reconnect, disconnect
Privacy	privacy, security, invade, safety, personal, policy, invasion, breach, protection, protect, private, disclosure, secure, unsafe, insecure, permission, fingerprint, encryption, violation, encrypt
Spam	spam, spammer, scammer, unsolicited, harassment, unwanted, bot, bombard, junk, scam, advertisement, pop-ups, scraper, hacker
UI	ui, interface, design, layout, gui, ux, clunky, redesign, aesthetic, navigation, usability, desing, sleek, appearance, aesthetically, intuitive, minimalistic, ugly, slick, graphic, unintuitive

Ranking Reviews

Ranking score defined as:

$$score(t) = e^{-r \left(\frac{1}{\ln(n)+1} + \frac{1}{\ln(n_t)+1} \right)}$$

Rating
#Keyword

Length



Rank	User Review	Score
1	Seriously bad user experience and interface. Once you've liked or unliked a song, there's no way to go back even if you've made a mistake. I don't know why Spotify is so popular with suck poor graphic design.	0.943
2	Clunky unintuitive interface missing basic and obvious music player features . You must get the basics right first before trying to push rubbish the user doesn't want.	0.914
3	Don't like the new design, in the now playing view the artwork is smaller to fit in the artwork on either side . I don't care what's on either end of my current playing track, or at least show it in a way that doesn't take up artwork space. The album art is always an awesome part of the music's personality so it shouldn't be minimised like this . Also the now playing bar at the bottom of the screen isn't flat looking, looks like design from windows XP. Not happy. An awesome service needs an awesome interface.	0.890
...

The top three reviews related to

Outline

- Topic 2: Issue Prioritization across different app platforms
 - Motivation
 - Framework of CrossMiner
 - Evaluation
 - Summary

Subject Apps

20 Apps



Review # 2,637,438

Review # 1,687,003

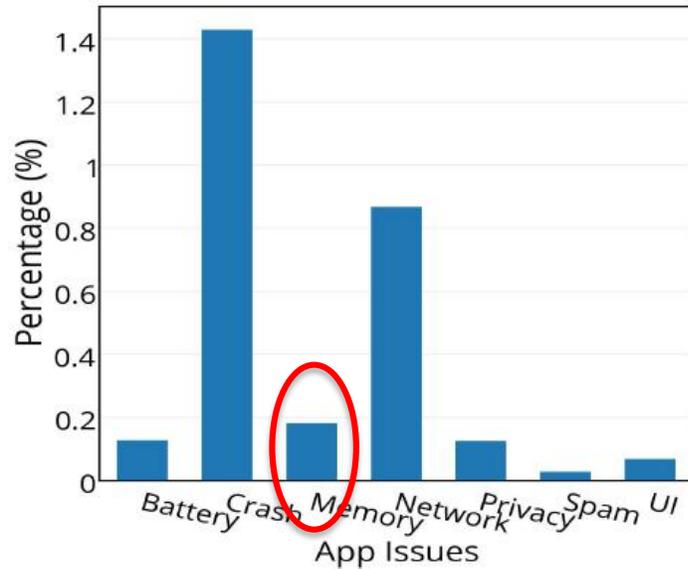
Review # 338,875

App	Platform	Title	Review	Date	Star
Facebook	Google Play	Crashes	Crashes non stop	2015-10-22	1

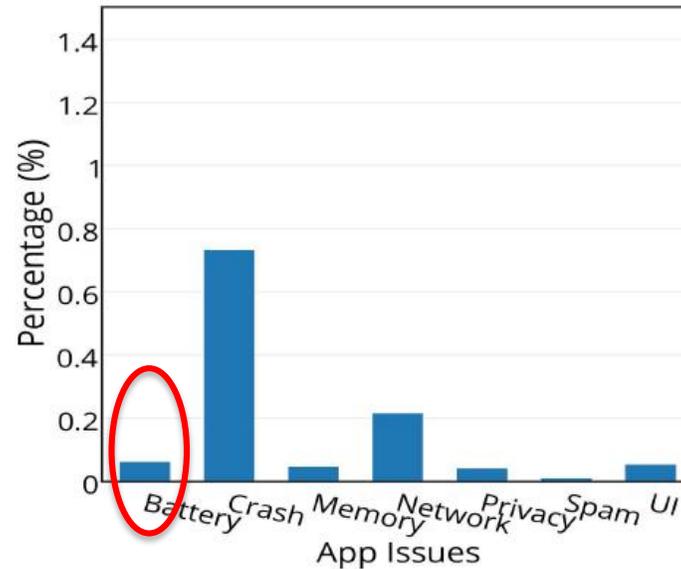
Experimental Result



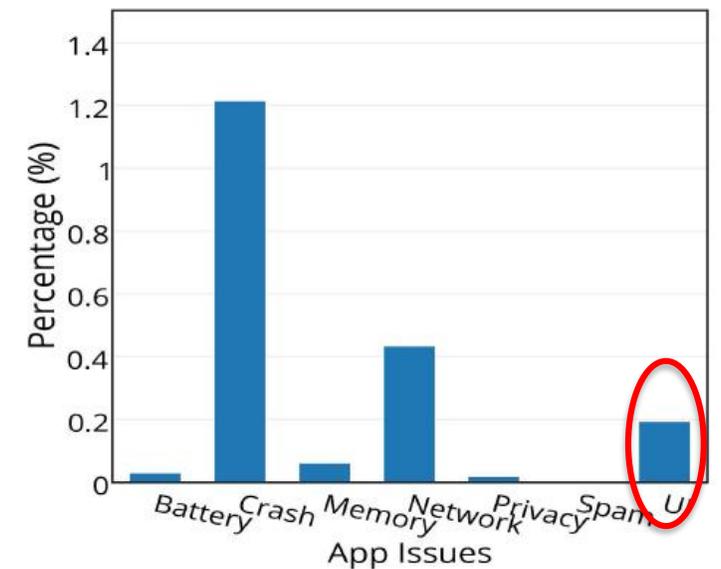
Percentage Distribution on Issues of Spotify Music



(a) Google Play



(b) App Store



(c) Windows Store

Performance Evaluation



Ranked Issues from **Android Community** of Spotify Music

Rank	Views	User Feedback	Issue
1	56416	No internet connection available	Network
2	32495	No SD Card storage !!	Memory
3	24797	Spotify for Android causing massive battery drain and heating of phone	Battery
4	11315	Spotify crashes on Android	Crash
5	1796	Issues with Android UI context menu touch area	UI
6	197	Intrusive or what!!!!!!	Privacy
7	80	Tired of the push notification spam!	Spam

We use the well-known Normalized Discounted Cumulative Gain (NDCG):

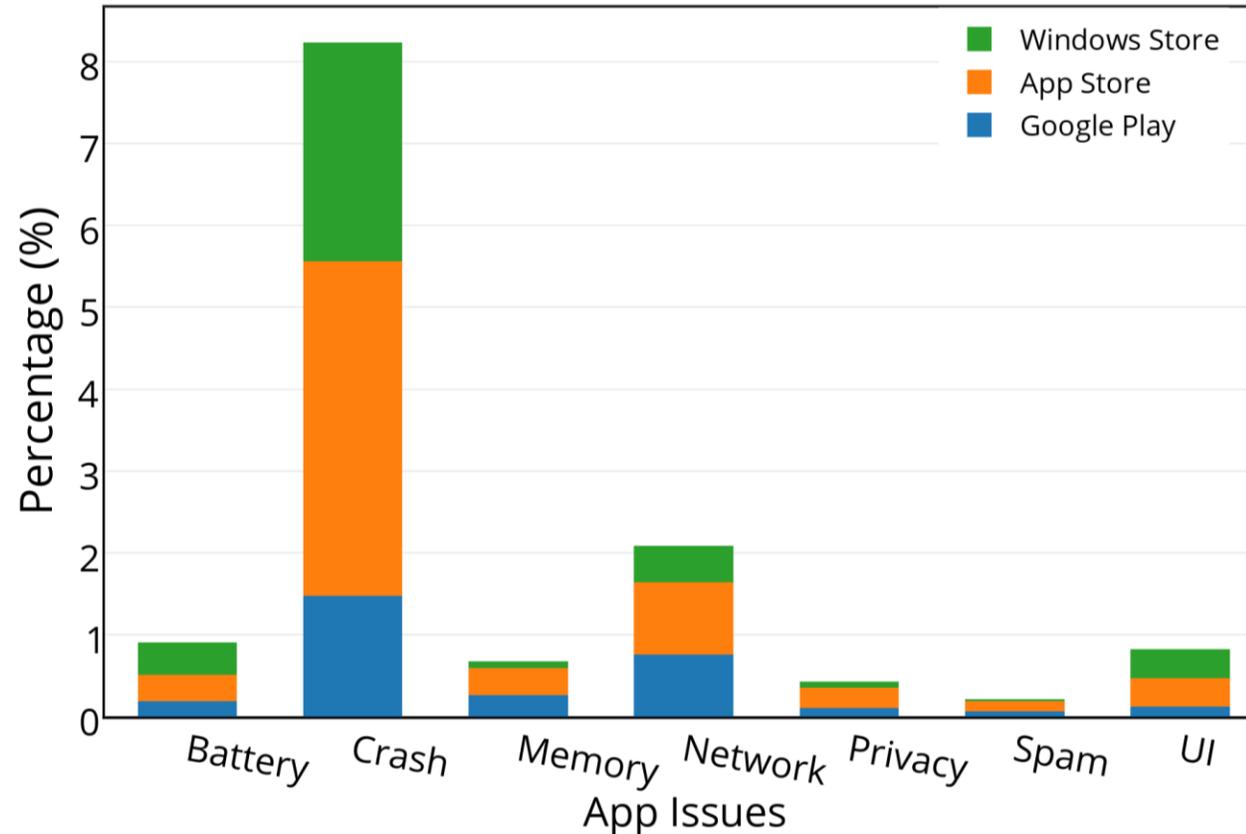
$$NDCG@k = \frac{DCG@k}{IDCG@k}$$

NDCG evaluation results:

	Android	iOS	Windows Phone
NDGC@7	0.943	0.911	0.982

Platform Comparison

Average Percentage Distributions on Issues for Different Platforms



p -value of ANOVA: $0.012 < 0.05$

Cross-platform issue Distributions are Significantly Different.

Outline

- Topic 2: Issue Prioritization across different app platforms
 - Motivation
 - Framework of CrossMiner
 - Evaluation
 - Summary

Conclusion of Topic 2

- We propose a framework **CrossMiner** to extract issue-related keywords comprehensively from real user reviews.
- We discover the **differences** on different platforms from users' perspective.
- We demonstrate that our framework reflects the **importance of user concerns accurately**.

Outline

- Topic 1: Online app review analysis for detecting emerging app issues
- Topic 2: Issue Prioritization across different app platforms
- Topic 3: Exploration on the effects of **in-app ads** on user experience
- Conclusion and future work

Outline

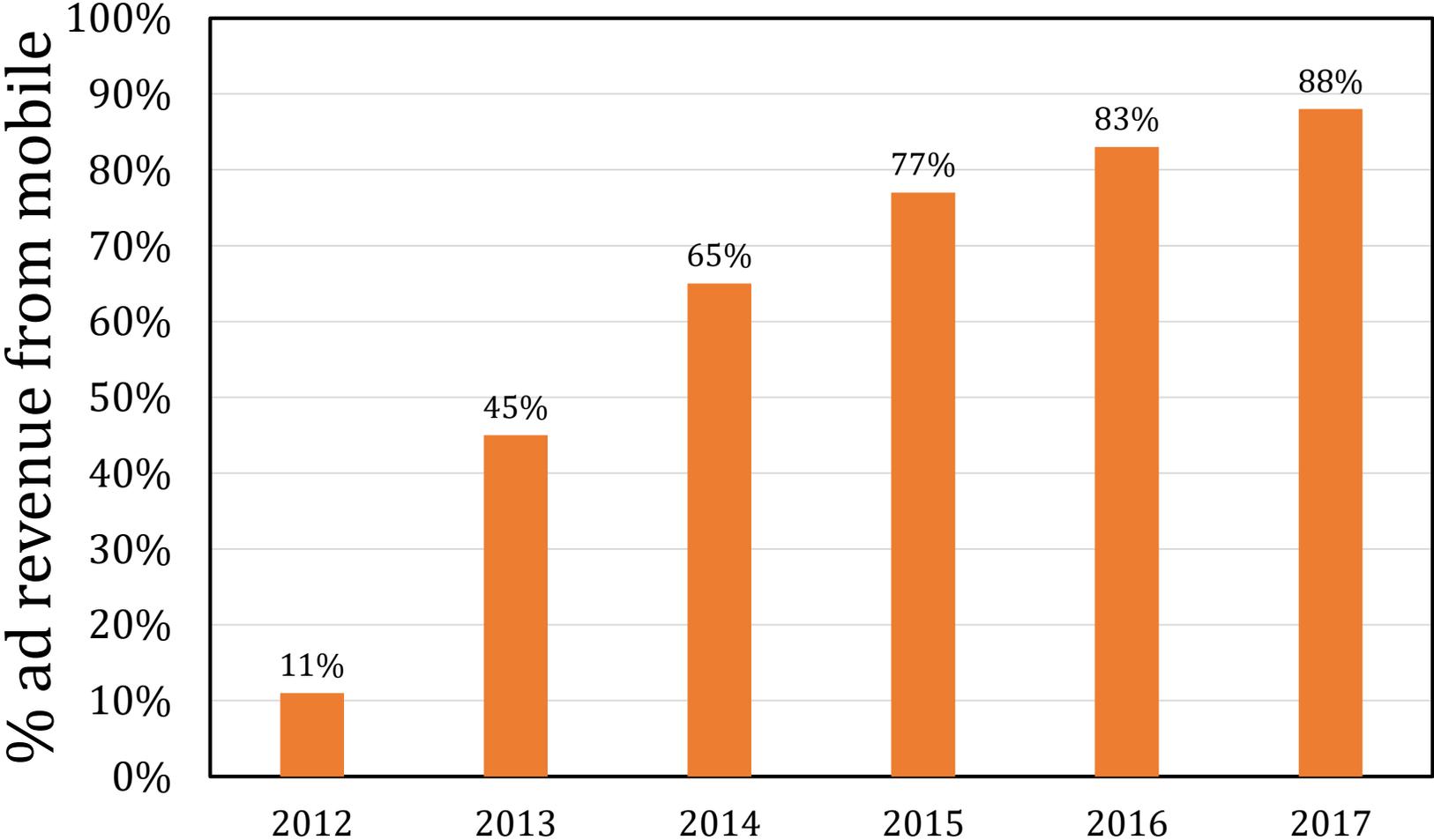
- Topic 3: Exploration on the effects of in-app ads on user experience
 - Motivation
 - Workflow of in-app ads exploration
 - Evaluation
 - Summary

Outline

- Topic 3: Exploration on the effects of **in-app ads** on user experience
 - Motivation
 - Workflow of in-app ads exploration
 - Evaluation
 - Summary

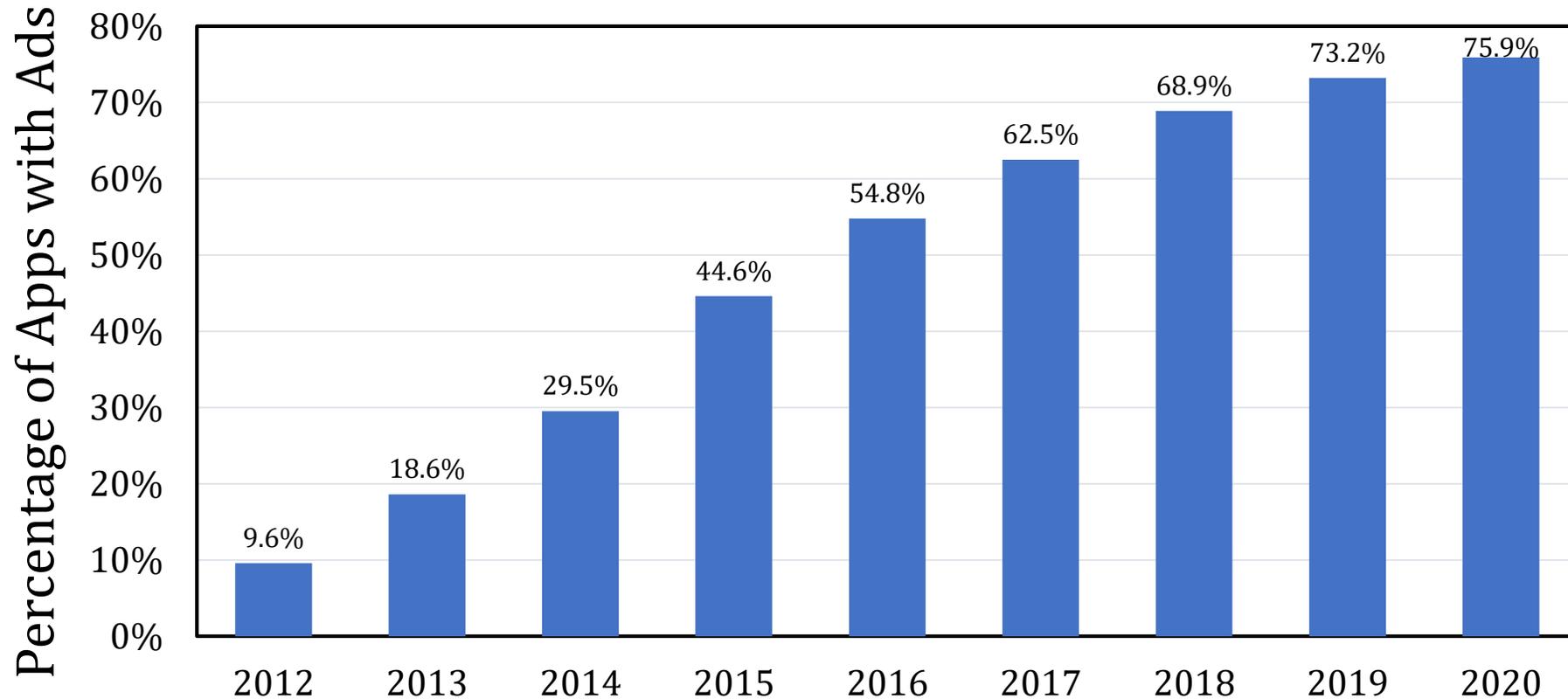


In-App Ads are Important



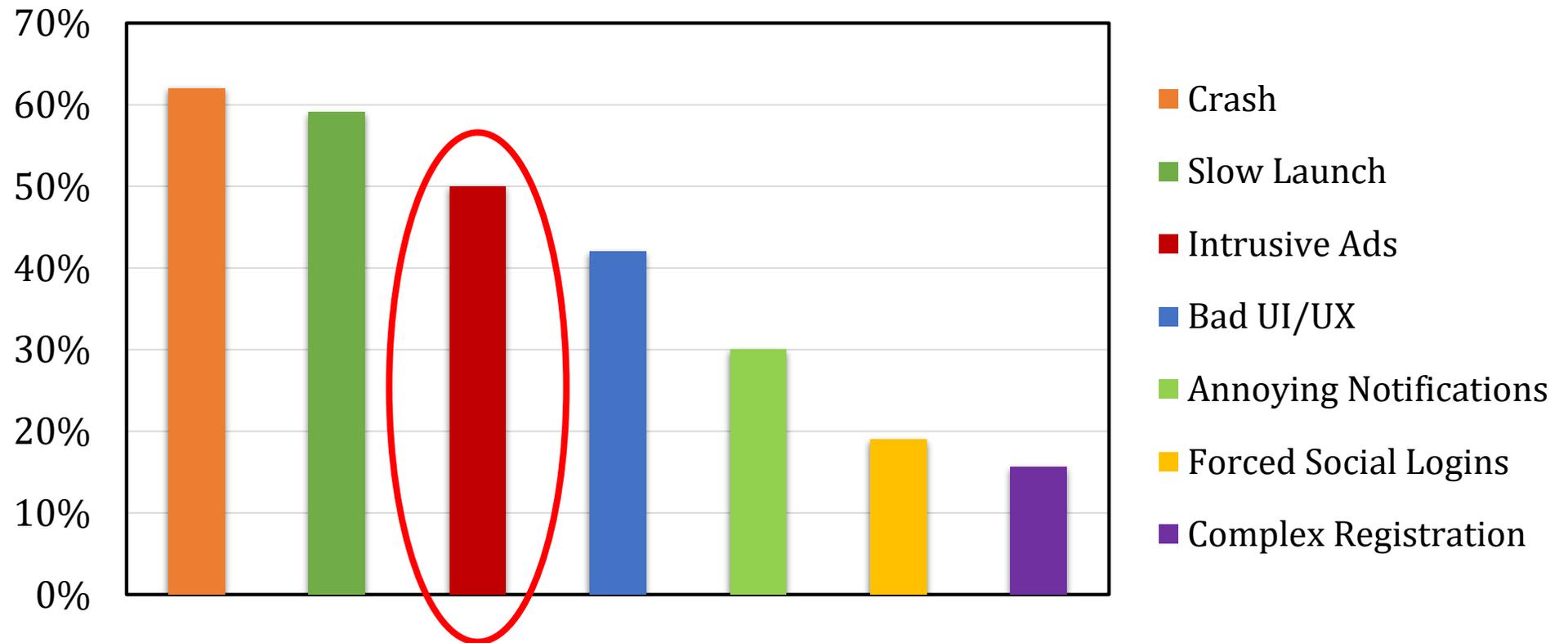
Apps with Ads are Growing

Display ads on mobile growth trajectory

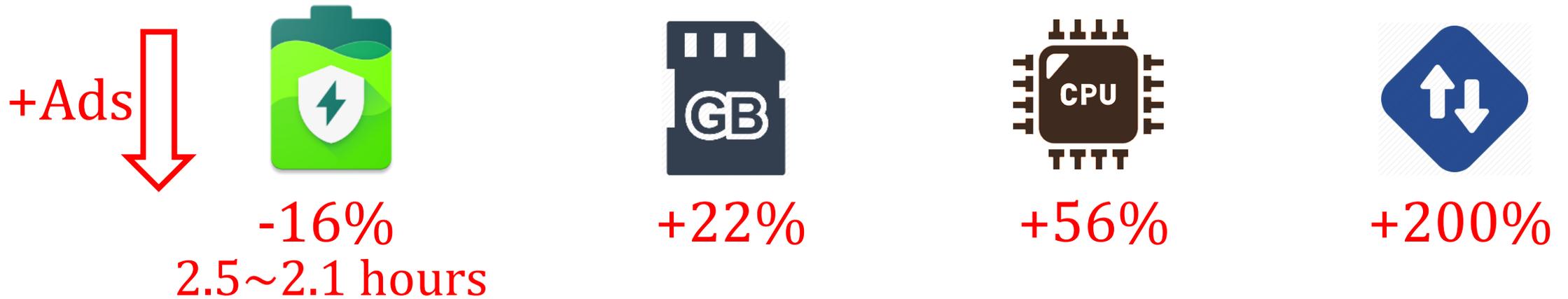


User Experience Matters

Top 7 Reasons Why People Uninstall Apps



Ads Cause Much Performance Cost



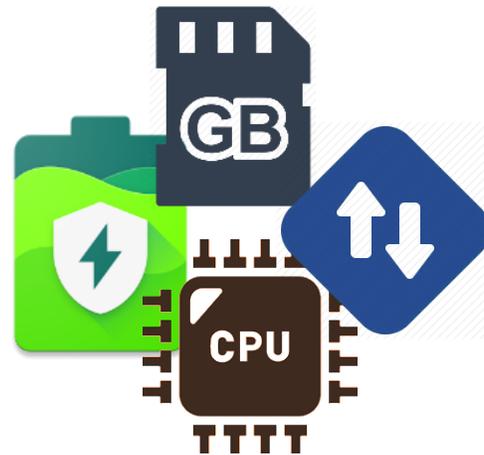
Our task



Mobile Ads



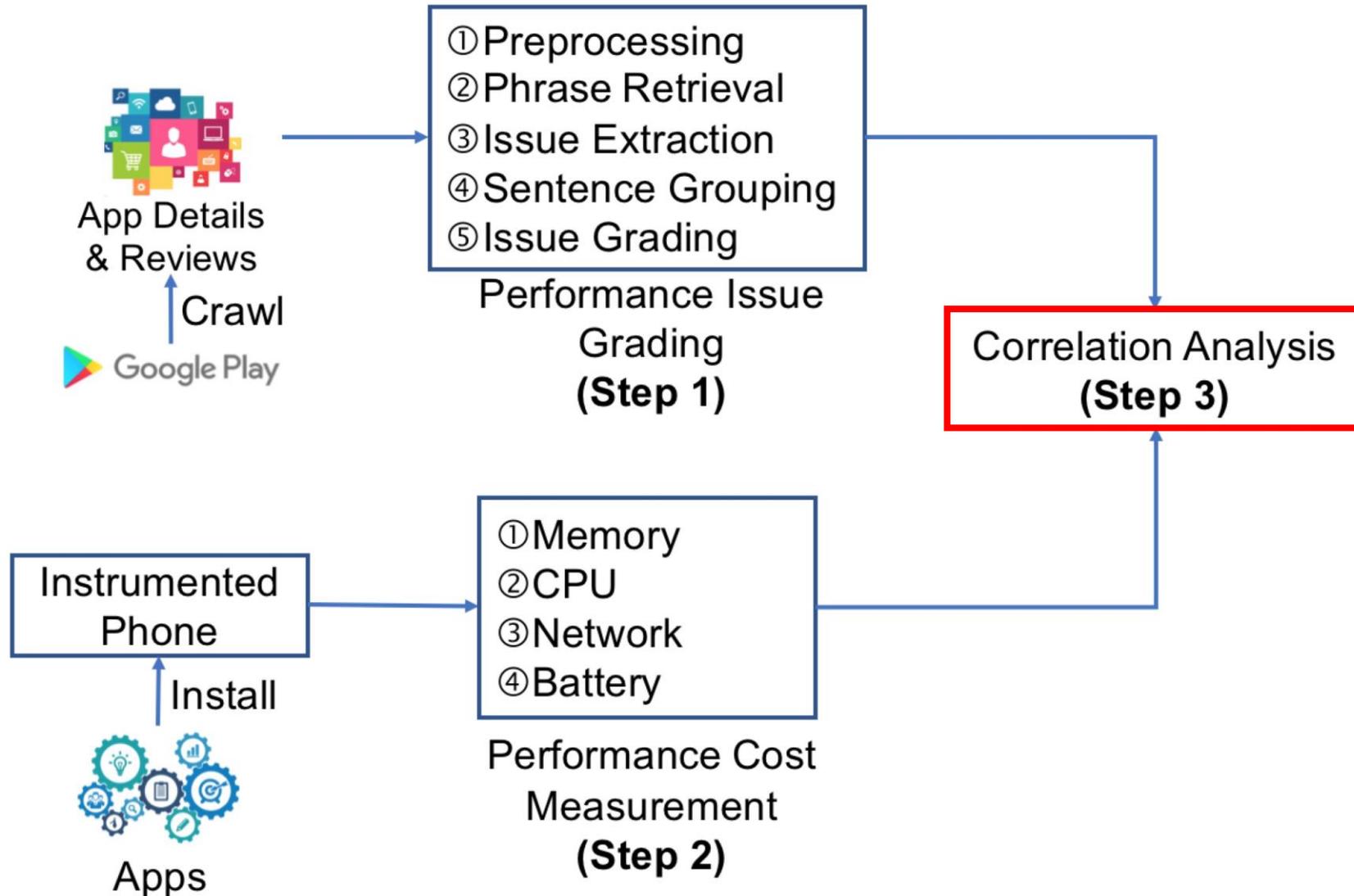
Users



Outline

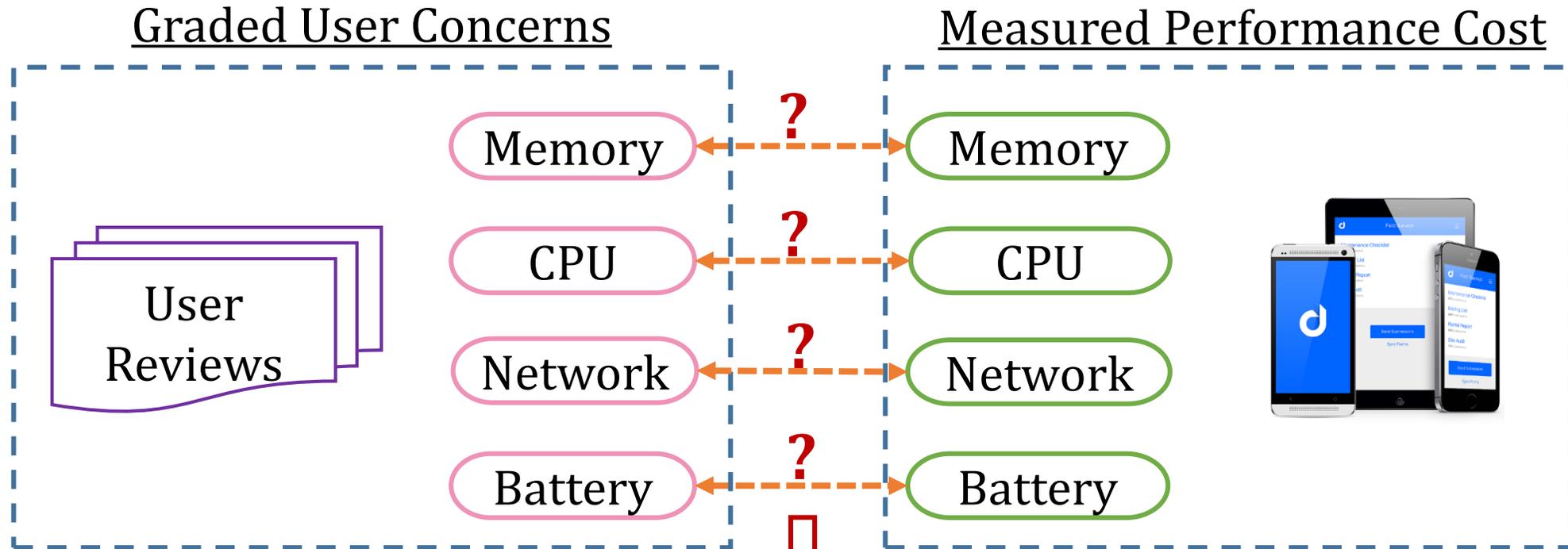
- Topic 3: Exploration on the effects of in-app ads on user experience
 - ❑ Motivation
 - ❑ Workflow of in-app ads exploration
 - ❑ Evaluation
 - ❑ Summary

Workflow of In-App Ads Exploration



Correlation Analysis

n subject apps



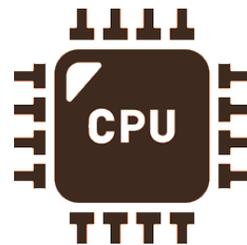
- ❑ Pearson Correlation Computation (PCC)
- ❑ Spearman's Rank Correlation (SRC)

Outline

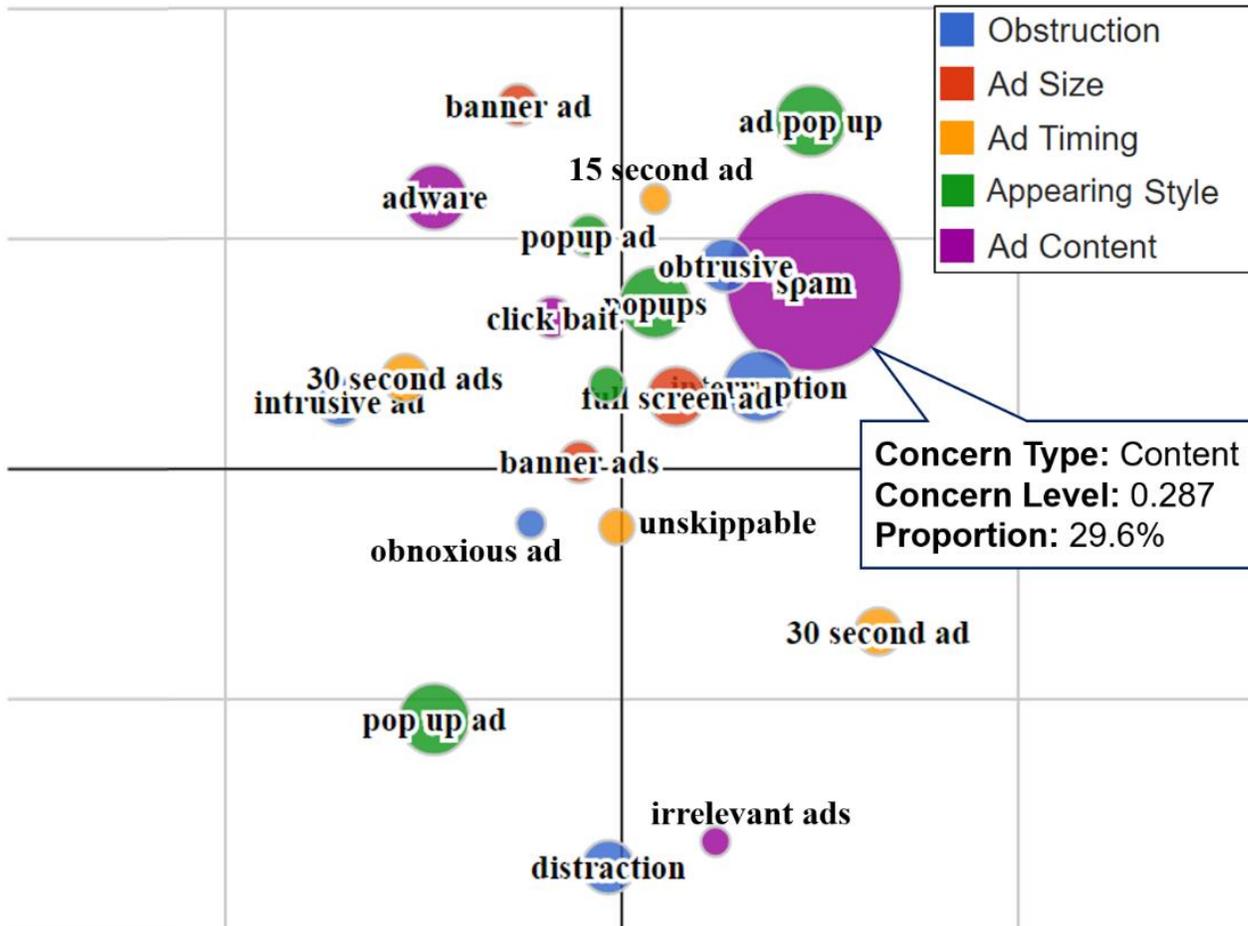
- Topic 3: Exploration on the effects of in-app ads on user experience
 - Motivation
 - Workflow of in-app ads exploration
 - Evaluation
 - Summary

Correlation Analysis Result

Cost Type	Memory		CPU		Network		Battery	
	<i>r</i> -score	<i>p</i> -value						
PCC	0.132	0.578	0.166	0.482	-0.281	0.229	0.534	0.015
SRC	0.372	0.105	0.213	0.366	-0.127	0.591	0.679	0.0009



User-Concerned Ad Issues



Insight:

- Choose ad SDKs that can **recommend relevant ads**;
- Avoid **pop-up ads**;
- Avoid **full-screen ads**;
- **Shorten** the compulsory **video ads**.

Outline

- Topic 3: Exploration on the effects of in-app ads on user experience
 - Motivation
 - Workflow of in-app ads exploration
 - Evaluation
 - Summary

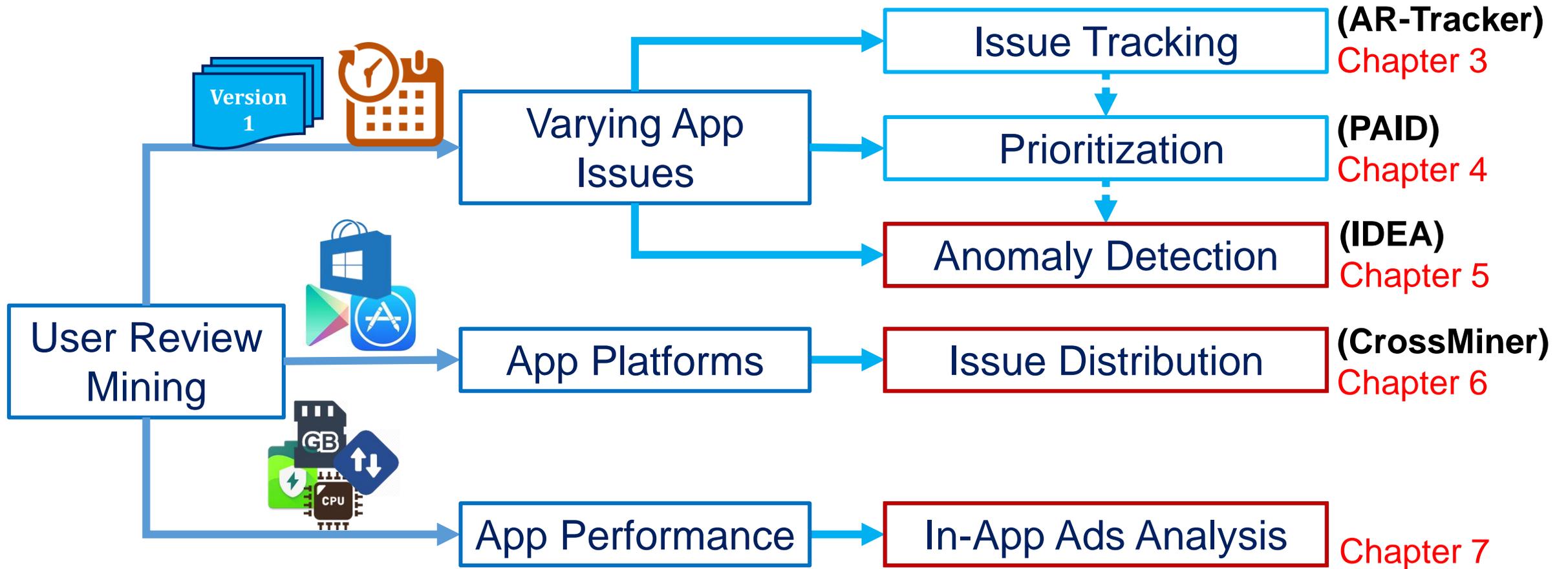
Conclusion of Topic 3

- We provide **detailed findings and insights** captured from an extensive analysis on user feedback.
- We carry out the **first empirical study** to explore correlations between **user concerns and performance costs** of ads in practice.
- We make **publicly available the source code** for cost measurement and user review analysis.

Outline

- Topic 1: Online app review analysis for detecting emerging app issues
- Topic 2: Issue Prioritization across different app platforms
- Topic 3: Exploration on the effects of in-app ads on user experience
- **Conclusion and future work**

Thesis Contributions

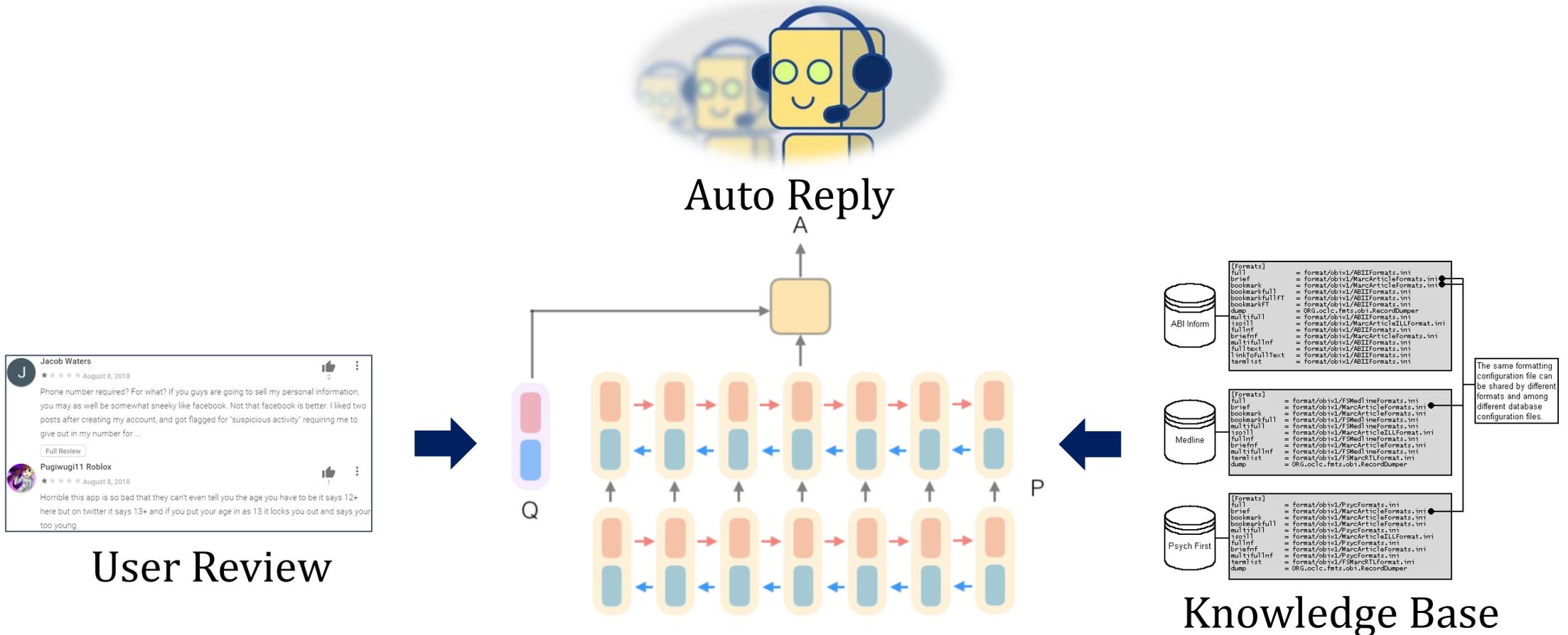


Conclusion

- We **design various frameworks and algorithms** for observing issues along with versions, across platforms, and for specific app component.
- We conduct **large-scale experiments** on verifying the effectiveness of our frameworks in various applications.
- We develop **several tools** for developers to use **in practice**.

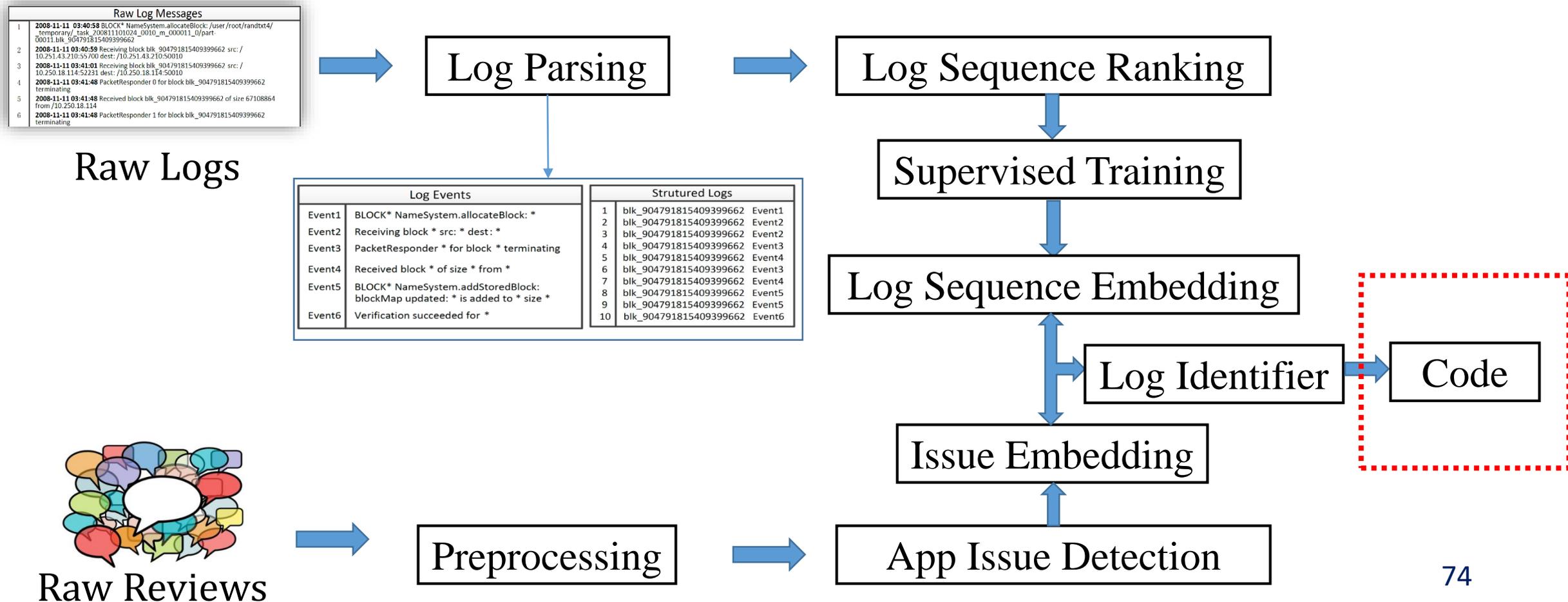
Future Work (1)

- Automatic user review reply.



Future Work (2)

➤ Review-based code localization.



Publications (1)

1. **Cuiyun Gao**, Jichuan Zeng, Federica Sarro, Michael R. Lyu, Irwin King. Exploring the Effects of Ad Schemes on the Performance Cost of Mobile Phones. The 1st International Workshop on Advances in Mobile App Analysis (A-Mobile), 2018.
2. **Cuiyun Gao**, Jichuan Zeng, David Lo, Chin-Yew Lin, Michael R. Lyu, Irwin King. INFAR: Insight extraction from app reviews. The 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), demo track, 2018.
3. Jichuan Zeng, Jing Li, Yan Song, **Cuiyun Gao**, Michael R. Lyu, Irwin King. Topic memory networks for short text classification. The 28th International Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018.
4. **Cuiyun Gao**, Jichuan Zeng, Michael R. Lyu, Irwin King. Online app review analysis for identifying emerging issues. The 40th International Conference on Software Engineering (ICSE), 2018.
5. **Cuiyun Gao**, Yichuan Man, Hui Xu, Jieming Zhu, Yangfan Zhou, Michael R. Lyu. IntelliAd: Assisting mobile app developers in measuring ad costs automatically. The 39th International Conference on Software Engineering (ICSE), poster track, 2017.

Publications (2)

6. Yichuan Man, **Cuiyun Gao**, Michael R. Lyu, Jiuchun Jiang. Experience report: Understanding cross-platform app issues from user reviews. The 27th IEEE International Symposium on Software Reliability Engineering (ISSRE), 2016.
7. **Cuiyun Gao**, Baoxiang Wang, Pinjia He, Jieming Zhu, Yangfan Zhou, Michael R. Lyu. PAID: Prioritizing app issues for developers by tracking user reviews over versions. The 26th International Symposium on Software Reliability Engineering (ISSRE), 2015.
8. **Cuiyun Gao**, Hui Xu, Junjie Hu, Michael R. Lyu. AR-TRacker: Track the dynamics of mobile apps via user review mining. The IEEE Symposium on Service-Oriented System Engineering (SOSE), 2015.
9. Hui Xu, Yangfan Zhou, **Cuiyun Gao**, Yu Kang, Michael R. Lyu. SpyAware: Investigating the privacy leakage signatures in app execution traces. The 26th International Symposium on Software Reliability Engineering (ISSRE), 2015.

Thanks!
Q&A