AWARE: Aspect-Based Sentiment Analysis Dataset of Apps Reviews for Requirements Elicitation

Nouf Alturaief | Hamoud Aljamaan | Malak Baslyman

Presented by
A.D.S.R.Gunathilaka

Presentation Structure

- Introduction
- AWARE Dataset
- Baseline Model Results

Introduction

- Motivation
- Background
- Contributions

Introduction:Motivation

- Requirement Elicitation is the practice of understanding and capturing the business domain knowledge, stakeholder goals, and user needs.
- It is a critical activity in the Requirement Engineering (RE) process, and it plays a significant role in the overall quality of the RE outcome [1]
- Crowd-generated content (e.g. apps reviews) is an essential source of knowledge that can be utilized to create a customer-centric experience.
- Utilizing apps reviews instead of traditional approaches (e.g. surveys or interviews) brings big benefits and enhancement to the requirement elicitation activity.
- For example, it enables reaching a much larger number of users with diverse socio-demographic backgrounds, which is important because most modern apps have a global target audience.

[1] D. Zowghi and C. Coulin, “Requirements elicitation: A survey of techniques, approaches, and tools,” in Engineering and managing software requirements. Springer, 2005, pp. 19–46.

Introduction : Background

- Identifying cases such as: is there a privacy concern? Are users satisfied with the new update? Which apps features matter most to users? Can help to gain a deeper understanding of specific apps features.
- To support such analysis, we can utilize Aspect-Based Sentiment Analysis (ABSA) [2], which identifies the sentiment with respect to a specific aspect.
- For example, in the sentence "**Screen sharing is superb, but sometimes the sound does not work.**", screen is an aspect
  with a positive sentiment and sound is another aspect with a negative sentiment.
- Having the aspect information leads to a fine-grained analysis that adds more accurate understanding of opinions [3].
- ABSA consists of three sub-tasks:
  (i)aspect category classification
  (ii) aspect term extraction
  (iii) aspect sentiment analysis.
- Each sub-task can be used to extract information with different granularity levels.

[2] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004, pp. 168–177.
[3] Y. Li, B. Jia, Y. Guo, and X. Chen, "Mining user reviews for mobile app comparisons," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 1, no. 3, pp. 1–15, 017.

Introduction : Contributions

- ABSA task has not yet been investigated in the context of smartphone apps reviews and requirements elicitation.
- This paper introduce AWARE: ABSA Warehouse of Apps REviews, which is an ABSA dataset of smartphone apps reviews.
- AWARE is annotated with aspect terms, aspect categories, and sentiments. As the diversity of mobile apps implies different needs and expectations [4], The paper provide annotated reviews of three distinct domains: (i) Games, (ii) Productivity, and (iii) Social Networking.
- The aspect categories are derived by analyzing the reviews and validated with stakeholders.
- To show the effectiveness of the dataset and to encourage further research, researchers have built machine learning baselines for the three tasks.

[4] C. Alves, G. Ramalho, and A. Damasceno, “Challenges in requirements engineering for mobile games development: The meantime case study,” in 15th IEEE International Requirements Engineering Conference. IEEE, 2007, pp. 275–280.

- Introduction
- AWARE Dataset
- Baseline Model Results

AWARE Dataset

AWARE Dataset

- AWARE is benchmark dataset of 11323 apps reviews that are annotated with aspect terms, categories, and sentiment.
- It contains reviews that were collected from three domains: productivity, social networking, and games.
- The data set contains two aspect definitions
    - **Aspect Term**: A term describing an aspect of an app that is expressed by the sentiment and that exists in the sentence.
    - **Aspect Category**: A predefined set of domain-specific categories.
- The aspect categories for each domain were derived using content analysis and they validated them with domain experts in terms of importance, comprehensiveness, overlapping, and granularity level.

AWARE Dataset

TABLE III
DEFINITIONS OF THE DERIVED ASPECT CATEGORIES.

Category	Definition
Aesthetics	Beauty of the appearance including graphics, audio, and video.
Cost	Price or value for money of the app or a feature
Compatibility	How the app can be launched on another environment or co-exist with another system.
Effectiveness	Usefulness and practicality of the app or a feature, including how the app offers its users the ability to achieve various objectives and reach the final goal.
Efficiency	Consumption of resources such speed, memory consumption, and battery consumption.
Enjoyability	Degree of pleasure derived from using the app or a feature such as fun, disappointment and engagement.
General	General positive or negative sentiment that does not fit any of the pre-defined categories.
Learnability	Ability to understand and master the app, such as difficulty in overcoming a particular challenge or understanding certain concepts.
Reliability	Crashing issues, failures, data loss, and connection errors.
Safety	Safety of users such as concerns about harassment, stalking and bullying.
Security	Personal information protection, such as encryption and authentication mechanisms, and how it is used and controlled.
Usability	Ease of use or convenience in using the app, including properties such as accessibility, look & feel, placement or existence of components, etc.

AWARE Dataset

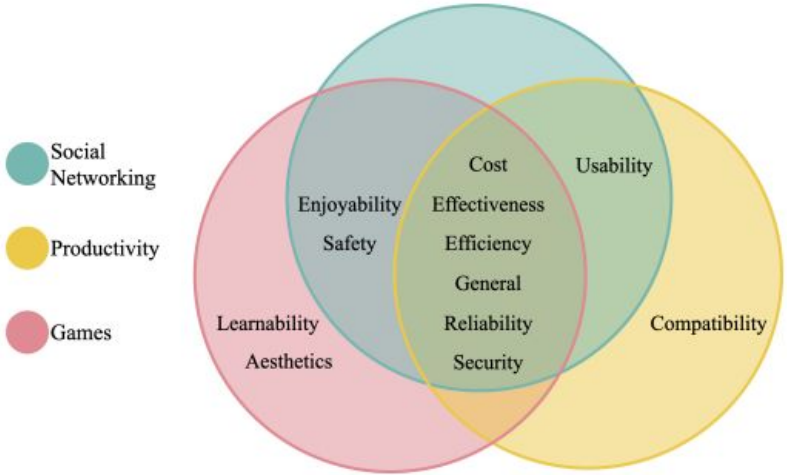


Fig. 2. Logical Relationships of the Aspect Categories.

AWARE Dataset

Fig. 3. AWARE Sentiment and Not-Opinion Distribution.

AWARE Dataset

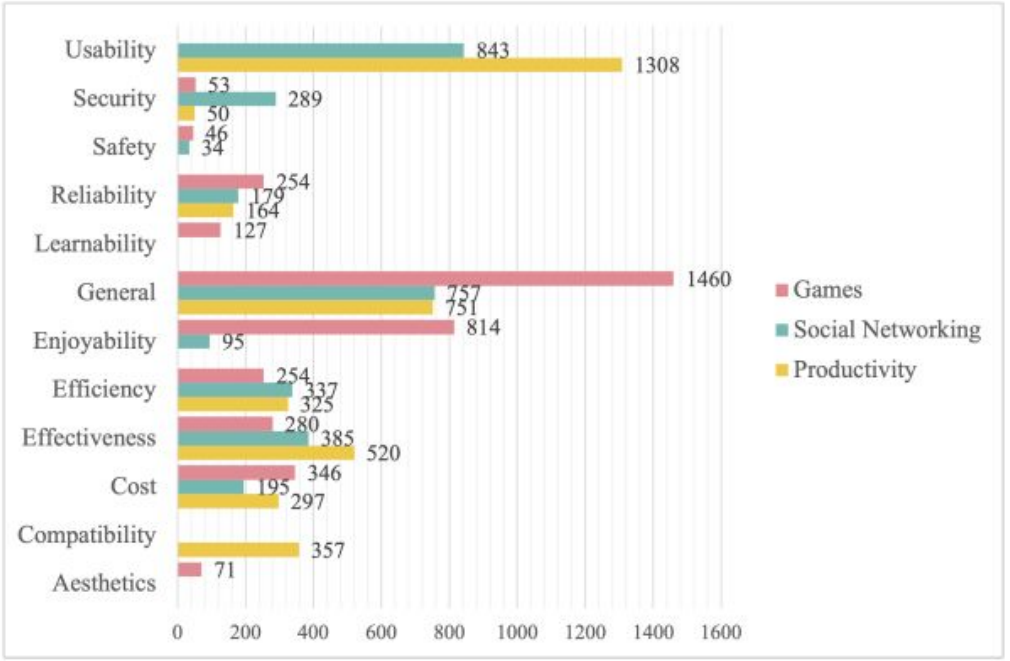


Fig. 4. AWARE Categories Distribution for Each Domain.

- Introduction
- AWARE Dataset
- Baseline Model Results

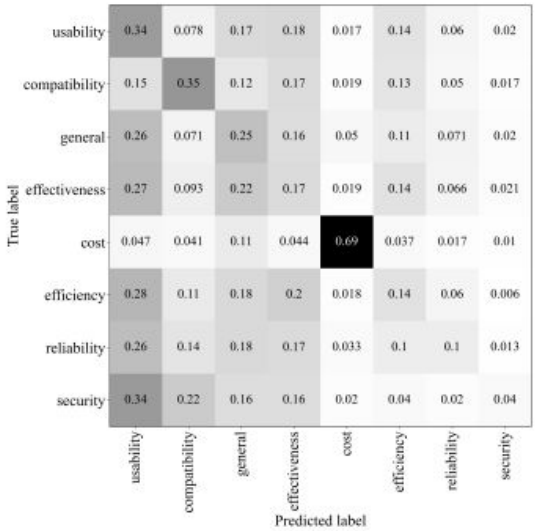
Baseline Model Results

Baseline Model Results

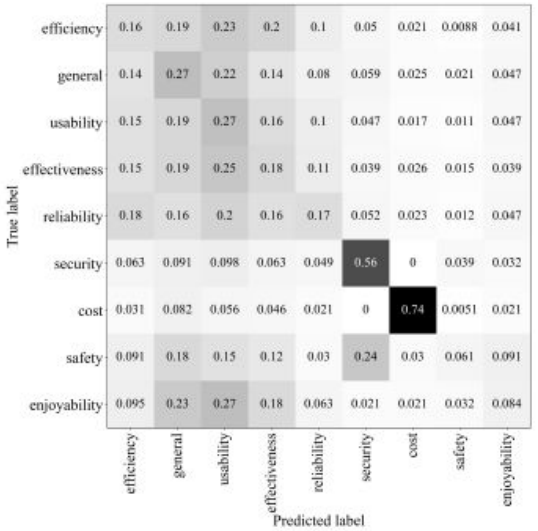
TABLE IV
BASELINE MODELS RESULTS.

Task		Model	Metric	Result
Aspect Category Classification	Productivity	SVM	F1	0.33
		MLP		0.32
	Social Networking	SVM		0.32
		MLP		0.31
	Games	SVM		0.32
		MLP		0.29
Aspect Sentiment Classification	Productivity	SVM	Acc.	68.71%
		MLP		66.11%
	Social Networking	SVM		69.72%
		MLP		67.32%
	Games	SVM		67.49%
		MLP		64.79%
Aspect Term Extraction		POS	F1	0.82

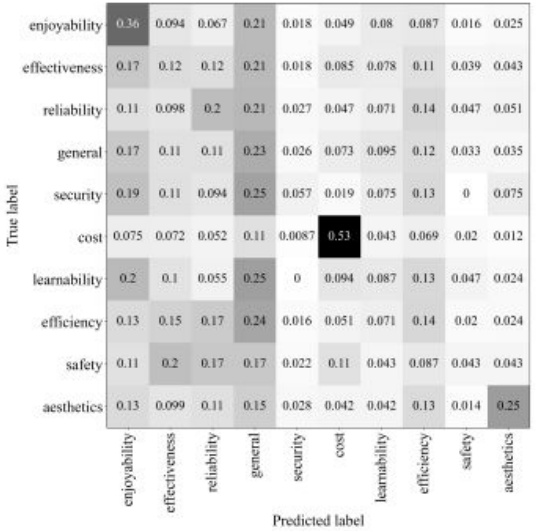
Baseline Model Results



(a) Productivity



(b) Social Networking



(c) Games

Fig. 6. A Confusion Matrix Representation of the Aspect Category SVM Classifier for Each Domain.

Thank You

Q & A