Aspect-based Sentiment Analysis on Mobile Application Reviews

A.D.S.R.Gunathilaka

Presentation Structure

- Introduction
- Methodology
- Experiments & Results

- Introduction
- Methodology
- Experiments & Results

Introduction

[30] J. Verma and A. Patel, “Evaluation of unsupervised learning based extractive text summarization technique for large scale review and feedback data,” Indian Journal of Science and Technology, vol. 10, pp. 1–6, 05 2017.

Introduction : What is Requirement Elicitation

- 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].

[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:Importance of user feedback in the context of mobile app development

- User involvement is a major contributor to success of software projects [1].
- Feedback typically contains multiple topics related to the application such as user experience issues, bug reports, and feature requests[2][3].
- W. Maalej et al [3] showed that majority of low star rating feedback usually contains shortcomings and bug reports of the application where four to five star ratings mainly consist of praise. It was noted that the feature requests are mostly coming from three to five star rating feedback.
- User comments can be used to improve user satisfaction of software products[4].

[1] M. Bano and D. Zowghi, “A systematic review on the relationship between user involvement and system success,” Information and Software Technology, vol. 58, 06 2014.
[2] D. Pagano and B. Bruegge, “User involvement in software evolution practice: A case study,” 05 2013.
[3] D. Pagano and W. Maalej, “User feedback in the appstore: An empirical study,” 07 2013.
[4] H. Li, L. Zhang, L. Zhang, and J. Shen, “A user satisfaction analysis approach for software evolution,” 2010 IEEE International Conference on Progress in Informatics and Computing, vol. 2, pp. 1093–1097, 2010.

Introduction : Why ABSA?

- Consider following user review: “UI is awesome and easy to use but application drains the battery faster”. Here “UI” is an aspect with a positive sentiment and “battery” is another aspect with a negative sentiment.
- Having the aspect information along with their respective sentiment leads to a fine-grained analysis [2].
- To support such analysis, we can utilize Aspect-Based Sentiment Analysis (ABSA) [3], which identifies the sentiment with respect to a specific aspect.
- ABSA consists of three sub-tasks:

  (i)aspect category classification

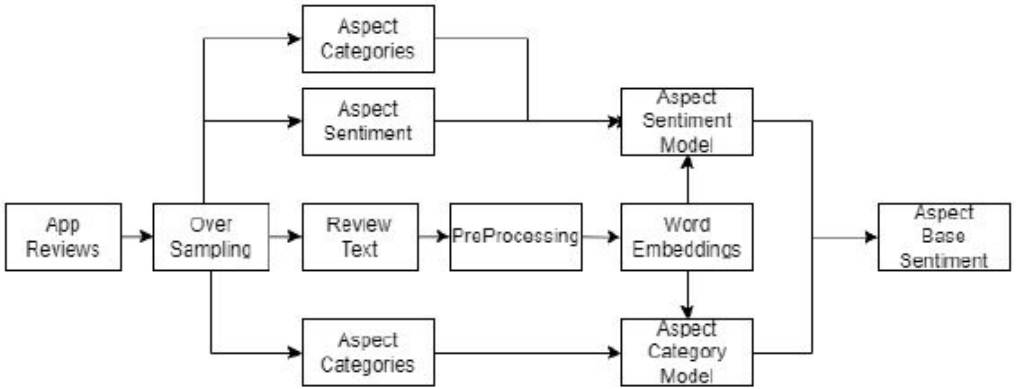
  (ii) aspect term extraction

  (iii) aspect sentiment analysis.

[2] 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.
[3] 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.

Methodology

Methodology : Proposed Approach Overview



Methodology : Dataset[4]

- 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.

[4] N. Alturaief, H. Aljamaan and M. Baslyman, "AWARE: Aspect-Based Sentiment Analysis Dataset of Apps Reviews for Requirements Elicitation," 2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW), 2021, pp. 211-218, doi: 10.1109/ASEW52652.2021.00049.

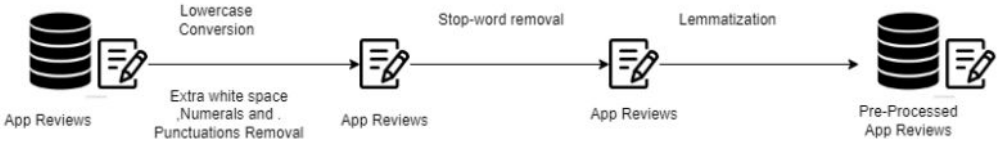
Methodology : Over Sampling the Data

- Contextual augmentation by Google Bert [5].

- Data Augmentation by Round-trip translation(RTT).

[5] Kobayashi, Sosuke. (2018). Contextual Augmentation: Data Augmentation by Words with Paradigmatic Relations. 452-457. 10.18653/v1/N18-2072.

Methodology : Preprocessing

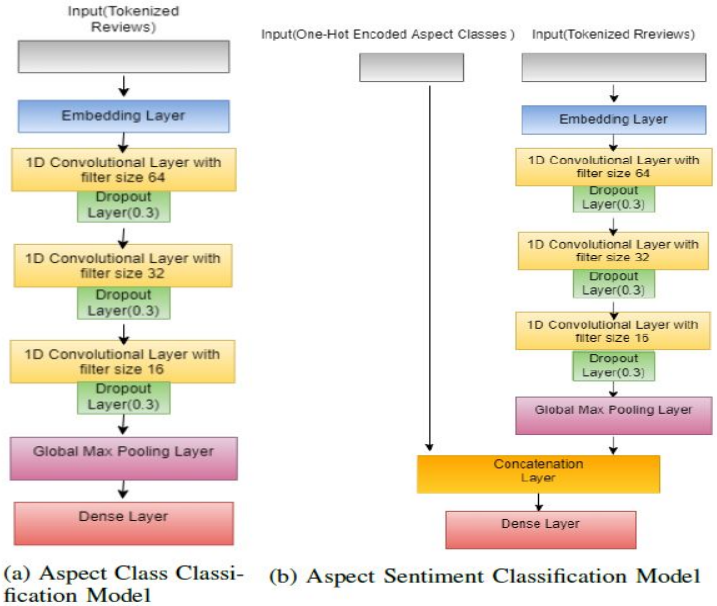


Methodology : Embeddings

Pre-trained Models

- FastText
- Glove
- Word2Vec

Methodology : Feature extraction and classification



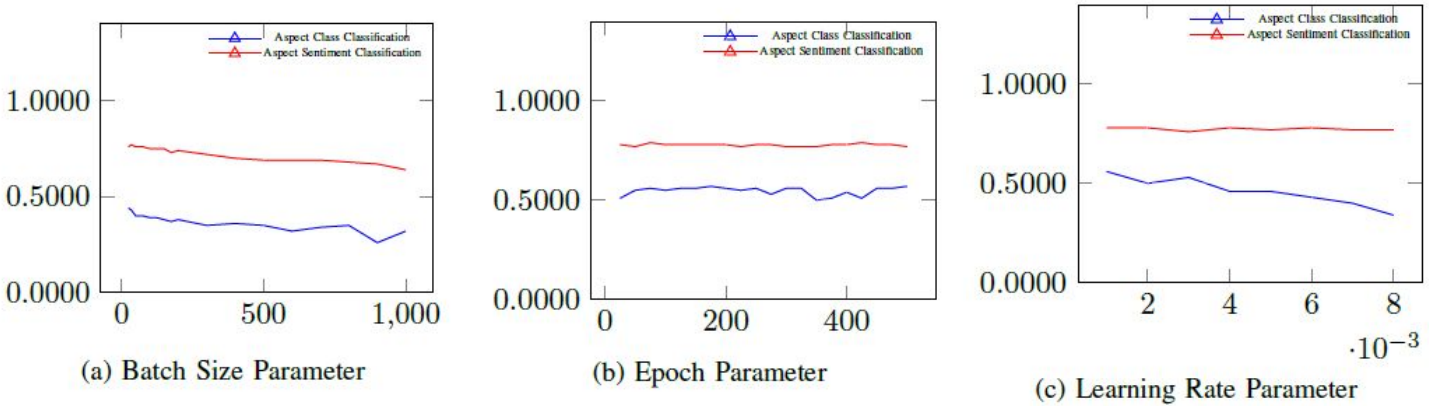
(a) Aspect Class Classification Model

(b) Aspect Sentiment Classification Model

- Introduction
- Methodology
- Experiments & Results

Experiments & Results

Hyper Parameter Tuning



(a) Batch Size Parameter

(b) Epoch Parameter

(c) Learning Rate Parameter

Aspect Category Classification

Dataset	Word Embedding	Preprocessing	BERT	RTT(DE)	RTT(CN)	RTT(TR)	RTT(JP)
Productivity	Fasttext	Disabled	0.60	0.59	0.25	0.61	0.60
		Enabled	0.63	0.61	0.23	0.62	0.59
	Word2Vec	Disabled	0.61	0.62	0.24	0.61	0.61
		Enabled	0.62	0.62	0.26	0.61	0.60
	Glove	Disabled	0.54	0.53	0.24	0.52	0.55
		Enabled	0.56	0.57	0.25	0.58	0.58
Gaming	Fasttext	Disabled	0.42	0.45	0.19	0.35	0.43
		Enabled	0.40	0.39	0.22	0.28	0.45
	Word2Vec	Disabled	0.42	0.41	0.23	0.37	0.44
		Enabled	0.39	0.42	0.21	0.37	0.44
	Glove	Disabled	0.42	0.44	0.20	0.34	0.42
		Enabled	0.30	0.30	0.21	0.24	0.31
Social	Fasttext	Disabled	0.62	0.62	0.58	0.25	0.60
		Enabled	0.60	0.61	0.58	0.27	0.60
	Word2Vec	Disabled	0.60	0.62	0.61	0.29	0.61
		Enabled	0.58	0.62	0.61	0.28	0.61
	Glove	Disabled	0.54	0.56	0.54	0.27	0.55
		Enabled	0.54	0.55	0.55	0.26	0.57
Average	Fasttext	Disabled	0.55	0.56	0.34	0.41	0.55
		Enabled	0.55	0.54	0.35	0.39	0.55
	Word2Vec	Disabled	0.55	0.55	0.36	0.43	0.56
		Enabled	0.53	0.56	0.36	0.42	0.55
	Glove	Disabled	0.50	0.51	0.33	0.38	0.51
		Enabled	0.47	0.48	0.34	0.36	0.49

TABLE I: Aspect Class Classification

Aspect sentiment Classification

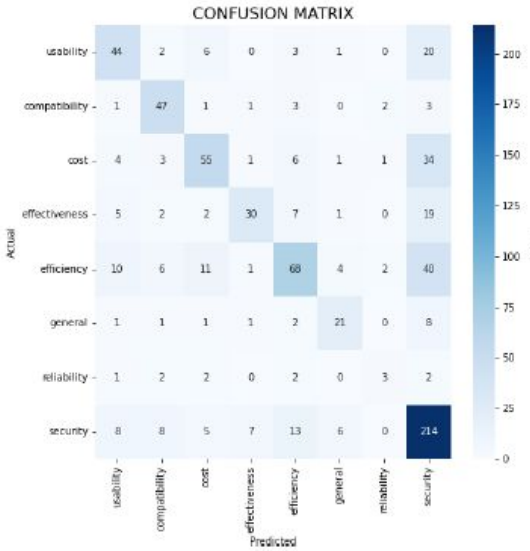
Dataset	Word Embedding	Preprocessing	BERT	RTT(DE)	RTT(CN)	RTT(TR)	RTT(JP)
Productivity	Fasttext	Disabled	0.81	0.81	0.62	0.80	0.78
		Enabled	0.79	0.79	0.63	0.81	0.81
	Word2Vec	Disabled	0.80	0.80	0.62	0.82	0.80
		Enabled	0.79	0.79	0.64	0.82	0.81
	Glove	Disabled	0.80	0.79	0.61	0.79	0.81
		Enabled	0.79	0.80	0.62	0.80	0.77
Gaming	Fasttext	Disabled	0.70	0.71	0.68	0.65	0.70
		Enabled	0.71	0.70	0.68	0.65	0.71
	Word2Vec	Disabled	0.70	0.70	0.67	0.64	0.70
		Enabled	0.72	0.69	0.68	0.66	0.70
	Glove	Disabled	0.71	0.72	0.70	0.65	0.70
		Enabled	0.70	0.69	0.69	0.65	0.70
Social	Fasttext	Disabled	0.83	0.80	0.81	0.63	0.83
		Enabled	0.82	0.83	0.81	0.62	0.82
	Word2Vec	Disabled	0.81	0.86	0.81	0.64	0.80
		Enabled	0.82	0.86	0.82	0.64	0.83
	Glove	Disabled	0.82	0.84	0.82	0.64	0.81
		Enabled	0.79	0.85	0.82	0.63	0.81
Average	Fasttext	Disabled	0.78	0.78	0.71	0.70	0.77
		Enabled	0.78	0.78	0.71	0.70	0.78
	Word2Vec	Disabled	0.77	0.79	0.70	0.70	0.77
		Enabled	0.78	0.78	0.72	0.71	0.78
	Glove	Disabled	0.78	0.79	0.71	0.70	0.78
		Enabled	0.76	0.78	0.71	0.70	0.76

TABLE II: Aspect Sentiment Classification

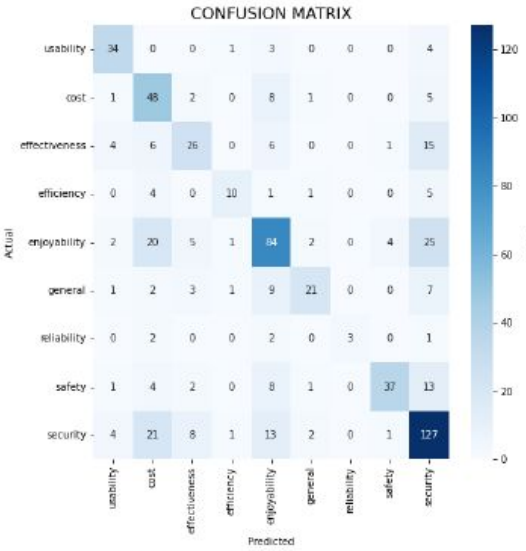
Final Results and Error Analysis

Task		Baseline	Results	Metric
Aspect Category Classification	Productivity	0.33	0.62	F1
	Social Networking	0.32	0.62	F1
	Games	0.32	0.42	F1
Aspect Sentiment Classification	Productivity	68.71%	80%	Acc.
	Social Networking	69.72%	86%	Acc.
	Games	67.49%	70%	Acc.

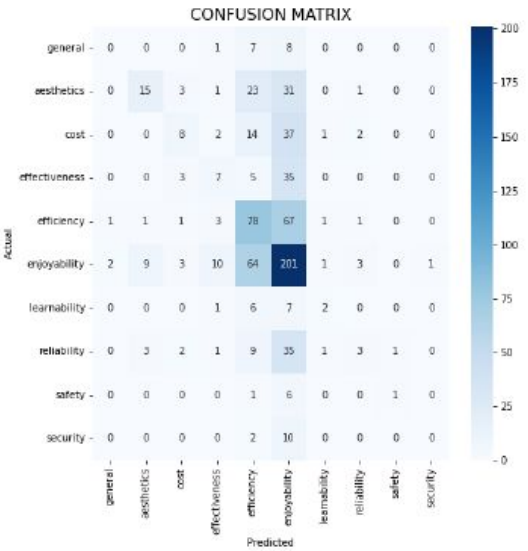
Final Results and Error Analysis



(a) Productivity



(b) Social Networking



(c) Game

Conclusion and Future Works

Thank You

Q & A