Comparison Study between the challenges of AI and Android Development through an empirical study of StackOverflow Posts

Md Rezwan Hassan Khan
Carleton Univeristy
Ottawa, ON, Canada
mdrezwankhan@cmail.carleton.ca

Nawab Haider Ghani
University of Ottawa
Ottawa, ON, Canada
nghan041@uottawa.ca

ABSTRACT

Developers are regularly facing difficulties in their respective domains and more solutions to their problems are also provided every time a problem comes up. It is very important to learn about the difficulties that developers in each domain face regularly and how popular the domains are in terms of other developers being interested in the topics of the respective domain. Artificial Intelligent is such a domain that has gained interest from developers during the last five years and developers are facing new problems day after day and solutions are quite fast to get since other developers are also interested. On the other hand, Android is a very popular domain that has been worked on for many years by developers and remains an important domain for mobile application developers mainly but still the developers are facing new difficulties working in android each day. Thus a comparison of these two domains is necessary to understand how developers in these domains face difficulties and the topics that they are interested in. Thus, in this paper, we perform an empirical study of the Stack Overflow posts related to each of these domains and find insights on the popular topics, the difficulties that developers are facing each year and the type of questions that goes unanswered in each domain. We have performed topic modelling on domain specific posts to obtain the topics that developers are interested in and then examined the difficulties and popularity of these topics for each domain. Finally we compared both these domains to find trends about the unanswered posts in these domains. Our results show that Android developers mostly face difficulties in solving UI or database problems while AI developers are facing most difficulties to solve compilation errors. Also insights show that AI developers are more interested in Neural Network models while Android developers are interested in UI related works. Overall comparison shows that AI development is more difficult than android and the trends of unanswered posts show that even though many android related posts are still unanswered, the number of AI related unanswered posts are rising quite drastically each year.

KEYWORDS

android, artificial intelligence, topic modelling, Stack Overflow

1 INTRODUCTION

Mining software repositories is a field where it uses data that has been created during development process by the developers to analyse and come up with insights that is useful for the developers and software industry[7].With recent vast improvement in technology, a lot of new software projects are being created everyday, new programming language, tools and framework has been introduced in the market and recently a lot of work is going on. Beside this new implementations, work in Artificial Intelligence and android mobile are increasing day by day. Previously research has been conducted on extracted data produced during software development but not many analysis was properly done on specific software domain especially in AI field. Beside this, only few work in the past was conducted where a detailed comparison analysis of the issues faced by developer in different software development field is performed. This motivated us, the authors, to conduct an empirical study on the developer difficulties in the AI and Android development domains.

In this paper, at first we did a separate analysis on the challenges Android and AI developers faces from Stack Overflow's posts. Secondly we did a comparison analysis between this two development domain to examine the difference issue developers face in each domain. Then we looked in to the evolution of the number of posts for each domain along with the topic wise evolution of number of posts each year. Lastly we constructed a graph for each domain for the number of unanswered posts and identified the trends for each domain that supports the difficulty levels for the respective domains. Our implementation include topic modelling to categorise the topic in each domain and answer 3 research questions to find out valuable insights[3].

After we find out the topics in AI and Android development that define different issues in each development we try to find out 3 research questions[4–6, 11] to come up with our final analysis and these questions are :

- RQ1: What are the topics developers are asking about ?
- RQ2: What are the type of questions developers are asking ?
- RQ3: Which is the most difficult question to answer development?

Based on these questions, we found out insights for each domain and compared these insights between these two domains to identify the more popular and more difficult, etc domain amongst these two.

The remaining part of the paper are designed as follows. Section 2 talk about the related work like our paper. Section 3 discuss about the methodology and Section 4 explained our analysis. Section 5 go through the evolution of AI and Android development and the trends for unanswered posts. And section 6 deals with threats validity and paper conclude with section 7 with the conclusion and future work.

2 RELATED WORK

In this section,we have discussed the similar work like ours that work with the similar data to analyse the challenges in development industry.

2.1 Challenges in chat bot Development

This research tried to find out the problems developers faced during chat bot development and try to give some suggestions that can improve the efficiency and productivity of chat bot development by analysing Stack Overflow's posts.[3]. This paper motivate us to work with different domain of development industry to find our research questions for the comparison analysis between 2 different fields of software development.

2.2 Use of Stack Overflow Data for analysis

There are numerous studies that has been done to analyse Stack Overflow data. Researchers tried to work with different part of data to come up analysis that define some specific issues.[14]

2.3 Topic Modelling on Stack Overflow Data

In some previous work they used LDA to define cell phone related questions in different categories.[13] . Beside this work, Bagherzadeh et al[5] worked with LDA to topic model the posts related to Big Data.

So from all these researches, we find out one thing that was not dealt before and that was the comparison analysis of the challenges between different domain in development industry . With this knowledge we try to come up with an analysis that can show us the different challenges developers face in different domains of development industry. In our case, we have selected Android and AI developments to analyse different challenges developers in these two industry and find out the different characteristics of the developers of each domain.

3 METHODOLOGY

For our research we have first extracted the domain specific data using relevant tags from Stack Overflow through some pre-processing. Then we performed Topic Modelling to find the appropriate topics for each domain and then performed the analysis on the topic-labelled data based on three research questions. All these steps will be discussed in detail in the following subsections.

3.1 Data Extraction and Pre-processing

The main objective of our work is to study and analyse the similarities, trends and any differences between the difficulties that developers of different domains (Artificial Intelligence(AI) and Android) face in their work. A good approach to perform such a study is to analyse the Stack Overflow discussions of developers working in the respective domains. We need topic-refined information/dataset to work on this project, but StackOverflow does not contain any such data dumps for specific domains. Therefore we had to follow a few steps to extract the dataset as much refined as possible for the specific topics. The steps were:

- Step1: We initially found a set of data from stack exchange [2] that were presented as filtered by domains. Since we are working with AI and Android domain, we downloaded the data dump of these two domains from stack exchange.
- Step2: For each domain, we performed a few general Natural Language Processing(NLP) techniques to filter out some common tags from the data. For that we first removed any punctuation marks or hashtags,etc from the post titles and then tokenized them to create arrays of words from the title. Then we removed any stopwords such as the "is", "has",etc from the tokenized arrays using the NLTK library and created a new dataset of tokenized arrays. Lastly in this step we created a dictionary of all the words extracted in the tokenized arrays to come up with some related tags to the respective domains.
- Step3: Once we had the dictionary created, we retrieved a few related tags for each of the domains and finally used Biq Query UI in the Google Cloud console [1] to extract all the posts from StackOverflow that included the tag words we retrieved earlier. For example, for AI, we searched for posts with words "artificial Intelligence", "Tensorflow", "Classification", "data", "keras", "loss-function" and "neural-network", etc. This gave us the refined dataset that we needed to proceed with the study. It should be noted that we received in total about 150,000 posts for Artificial Intelligence, and even though we collected almost 3 times the data for Android, we only took about 150,000 posts randomly.

3.2 Topic Modelling

After we have created the dictionary of words from the title of the "question_posts" as mentioned above , we have implemented Latent Dirichlet Allocation technique as this work good with NLP problems and can accurately define categories of the topics[10]. It is an unsupervised learning and LDA assigns a random topic to each word in the dictionary we have created. So each word from the dictionary was assigned with a probabilities basis on per topic to find the how likely that word is related to that specific topic. Topic with highest probability based on the probability of the words are considered as the dominant topic.We use the step that was taken by the paper written by Ahmad Abdel Latif and his co-authors[3] and followed the implementation steps by Susan Li with our own little modification.[9]

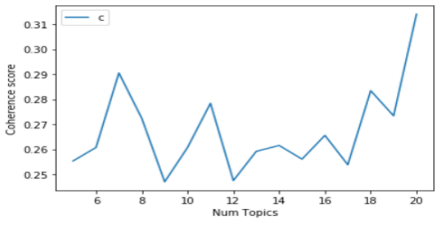

Figure 1: Android Coherent Score vs Number of Topic

One of the hardest things we faced was finding the accurate number of topics for our data set. High number of topic will make our topic modeling too specific and low number of topic will make it too generic . So to overcome this issue we did our topic modelling with various number of topic ranging from 6 to 20 and find the

coherent score.[3, 12]. From Figure 1 and Figure 2, we can see that coherent score increases with increasing number of topics. But there is a sudden big decrease in coherent score for both Android and AI's bag of words after Topic number is equal to 7. It make more sense to go for topic value before the first major decrease in coherent score as suggested by Shashank Kapadia's work[8].So for both AI and Android we have found out the optimal number of topic in each corpus is 7 as shown in Figure 1 and Figure 2.

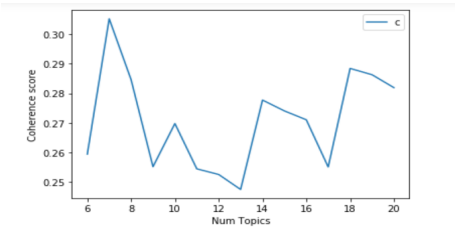

Figure 2: AI Coherent Score vs Number of Topic

Table 1: Android development topics,category and their search keyword.

Topic	Main Category	Search Keywords
Android studio's issues	Development	Android, studio
Graphical User Interface	User Interaction	Android, change, layout
Fragment UI	User Interaction	Android, fragment
Google API	Integration	Android, google, error
Firebase	Development	Android, database
Image	Development	Android, image
Active function	Development	Android, active

Table 2: AI development topics,category and their search keyword.

Topic	Main Category	Search Keyword
Tensorflow	Integration	Tensorflow
Classification	Model	classif
Data	Preprocessing	data
Keras	Model	keras
Neural Network	Model	neural network
Compilation Error	Error	error,object
Loss function	Evaluation	loss function

After finding out the number of topics, we manually go through each topic's 30 posts to decide our final topic name. After that we also divided our 7 topics into few main categories to explain our findings in more details. For this we implement the same strategy that Ahmed AbdelLatif[3] implemented in their work.And then we find the three research questions as mentioned in section 2.

4 CASE STUDY ANALYSIS

In this section, we tried to find our 3 research questions through the analysis of Android and Artificial Intelligence's posts. We mainly extracted data from question's title,body and the topics we have selected to come up with our final analysis.

4.1 Reasons to select these questions

- RQ1: What are the topics developers are asking about ?
  The 3 reasons to select these question are :
  - To find out topics that are common in both AI and Android Developments.
  - To find out topics that are not common in both AI and Android Development.
  - What are the most popular topics in each domain?

  How we find out its answer ?
  We use 3 metric measurements to find the popularity of topics :
  - Average view of a post: More a topic is viewed, the more likely it is to be popular in the development community.
  - Average Favourite: These are the posts that are interesting for developers and likely to recurrent more in the development process.
  - Average score of a post : It tells us how the developer's community sees this post as interesting or not.
- RQ2: What are the type of questions developers are asking ?
  From the previous work[5] we find out that the most types of the questions developers ask are 'why', 'how' and 'what'. And consider the remaining type as others.
  First ,we have randomly selected 30 posts from each topic. For android ,Graphical User Interface and Active function's were exceptions. To select a random post from each topic we have shuffled the post first with dataframe.sample() function of pandas library. After that we have selected the first 30 posts from each topic's randomized dataframe.
  In our next step, we manually go through each 30 posts from each topic and manually label each post as why,what,how or others. A short description of each label is giving below to give an idea how we perceive each label[5]:
  - How :These are all the posts that ask about the ways to implement something.
  - What : These are all the posts that ask about the meaning of something
  - Why :These are all the posts that ask about the purpose of something.
  - Others :These are all the posts that don't belong to the three labels mentioned above .

  After that we find out the overall percentage of each label for the entire data set. Along with this we also find out the individual percentage of label in each topic.

- RQ3: Which is the most difficult question to answer in development? We want to find out the posts that are most difficult to solve/answer in Stack Overflow. This analysis can help development industry to figure out important issues developers are facing in their development process and come up with suggestions such as improving the tools,framework and improved version of documentation. We measure a metric used in a previous work[5,6] to find the difficulties of the posts:
  - The percentage of a post without an accepted answer . So the post having less accepted answers tends to be the most difficult post we assume.

4.2 Android

In this section we will analyse the answer of 3 research questions based on Android Development's posts.

- RQ1: What are the topics in the Android domain, developers are asking about ?

Table 1 shows that there are 7 topics which we then categories manually into 3 main categories which are Development,User Interaction and Integration.

Table 3: Android development topics,category and their popularity.

Topic	Main Category	Posts	Avg View	Avg fav	Avg Score
Android studio's issues	Development	2338	1526	0.34	1.23
Graphical User Interface	User Interaction	118	1153	0.36	1.35
Fragment UI	User Interaction	1965	1111	0.27	0.93
Google API	Integration	151	1667	0.25	1.06
Firebase	Development	4383	961	0.26	0.64
Image	Development	8412	1624	0.38	1.1
Active function	Development	1194	1002	0.32	1.2

Development: As shown in Table 3, this category mainly deals with posts related to building android apps with different frameworks and questions about implementation and configurations. Most of the issues are related to the topic 'image'(8412 posts) .Topic 'image' deals with questions related to use of images in android app development like how to edit,load,convert etc. After "image' , "Firebase/Database" has 4383 posts which deals with the issues related to Database in android app development. For example , 'About saving View objects in firebase database android'. On 3rd position, it is " android studio's issues" . An example of its post is "How to detect system background touch in android studio programmatically". So most of the posts in this topic are development issues regarding android studio app development.

Next, Active function's has 1194 posts that deals with issues related to functions/features that developers want to make it active in the app. For an example : " How to keep active the service that listens to notifications on android? " These 4 topics we categorise as Development .

User Interaction : This category deals with issues related with UI and design .Topic 'Graphical User Interface' with 118 posts and 'Fragment UI' with 1965 posts fall into this category. 'Graphical User Interface' deals with issues related to change/implement layout of GUI. For an example 'Can anyone put a light on this (txtView and button layout on android studio not working properly)". And Fragment UI deals with issues related to Fragment interactions and errors. For an example : "Using volley on android, where to declare the queue using fragments?"

'Fragment UI' has the least average score which shows that developers did not find this post very interesting. Graphical User Interface' has the least post among all other 6 topics so it shows that these are not that popular among developers . It's quite surprising to see a very less number of posts related to this topic as it's a very important part in android development . So we went through each post and saw that the posts from this topic have very good favorite scores and average scores. So we came up with a conclusion that though it has the lowest post, developers don't have many issues related to this topic.

Integration : Topic is 'Google API' in this category with 151 posts which are related to Questions related to integration errors or problems for Google Api. Though it has few posts , this topic is very popular among the developers as it has the most number of average views but it has the lowest average favourite score. So from this we can say that developers are searching for answers for these issues frequently but developers don't find this post very helpful and unlikely to take any help in their development process.

Summary of the analysis :

1. Most posts are related to the topic **"image ".**
2. Post from **"image"** seemed to be most helpful for the developers and it tends to recurren more during android development.
3. The most popular topic in android developments is **Google API** but these posts are not very helpful for the developers**.**
4. Most interesting topic in android development is **Graphical User Interface** with developers facing the least issues in this topic.
5. The least interesting topic in android development is **Fragment UI.**
6. In general the popularity of all the topics are very similar as the average view of each topic is very close to each other **.**

- RQ2: What are the type of questions Android developers are asking ?

From the Table 4 we can see that around 45% of the questions are of 'how ' type followed by "why"(28.2 %) and "others"(16.8%) and "what"(9.9%). So from this , we can say that android developers are looking for posts that ask about the ways to implement something. From the table we can also infer that , "image" has the most post of "how type". It shows that developers are asking questions like how to integrate images in the app and how to load,edit and convert images in android frameworks. The "android studio's issues" has a category with most "what " type which means developers wanted to know about the posts in this topic that ask about the meaning of

Table 4: Percentage of Android development posts type in StackOverflow

Topic	Main Category	How	What	Why	Others
Android studio's issues	Development	50	16.7	20	13.3
Graphical User Interface	User Interaction	59.3	14.8	14.8	11.1
Fragment UI	User Interaction	46.7	13.33	16.7	23.3
Google API	Integration	10	3.3	83.3	3.3
Firebase	Development	46.7	6.7	23.3	23.3
Image	Development	60	10	10	20
Active function	Development	44	4	28	24
All Android Development		**45.04**	**9.9**	**28.2**	**16.8**

something . For example, developers ask more about the android studio, basic info about all the frameworks ;basically they want to know about something related to the topic "android studio's issues". From the table we can see that Google API has the most "Why" type post which shows that developers face most difficulties when they implement anything related to Google API and get maximum issues when implementing it;thus developers ask question why they are getting this issue for an example, " why getting this error: package com.google.android.gms.location does not exist". There are around 16.2 % posts that are classified as "others" which means that there are a huge number of posts that are not directly cannot be classified directly as other 3 type as it's very confusing to tag them what the posts are actually asking about. So further investigation should be done on this post to classify it .

Summary of the analysis :

1. Most of the posts are of '**How**' type and least are '**What**' type .
2. Developers in "image" (60%) looks for posts in android development by posting "**How**" type questions.
3. Developers in "android studio's issues" (16.7%) looks for posts in android development by posting "**What**" type questions.
4. Developers in "Google API" (83.3%) looks for posts in android development by posting "**Why**" type questions.
5. There are around 16.2 % posts that are classified as "**others**" .

- RQ3: Which is the most difficult question to answer in Android development?

From table 5 we can see that overall every topic has a huge number of posts with unaccepted answers. "Android studio's issues" being the highest with 67.8% of the posts without an accepted

Table 5: Percentage of Android Posts without Accepted Answers

Topic	Percentage of Posts without Accepted Answers
Android studio's issues	67.8
Graphical User Interface	50.8
Fragment UI	57.6
Google API	61.0
Firebase	56.9
Image	57.5
Active function	52.7

answer. After analyzing each post of "Android studio's issues" we found out that the posts without accepted answers have very low scores by developers which may mean that the post information is not very structured and difficult to understand and more likely not get an accepted answer. Followed by "Android studio's issues",the most difficult posts are related to Google API(61%) . Among all posts related to Graphical Interface(50.8%) has the least unaccepted answer and after analysing the post we say that the average scores and average favourite for Graphical interfaces are higher compared to all other topics. That gives us an insight that these posts are easily understandable, interesting and tend to give answers that have high probability to be accepted .

Summary of the analysis :

1. Posts related to "**Graphical Interface**" tend to be answered more frequently .
2. Posts related to "**Android studio's issues**" have the most difficult posts to answer followed by "**Google API**"
3. Posts with accepted answers have high scores and favorite scores in general.
4. Posts without accepted answers have low score and low favorite score in general.

4.3 Artificial Intelligence

Table 2 shows that there are 7 topics which we then categorize manually into 5 main categories which are Integration, Model, Preprocessing, Error and Evaluation. The table also includes the search words we used as tags to search for the related posts of that particular topic.

- RQ1: What are the topics in the AI domain, developers are asking about ?

Integration: From the table 6 we can see that ,Topic is 'Tensorflow' in this category with the highest post among all other topics which are related to Questions related to integration errors or issues with Tensorflow. For an example : "How to fix an EnvironmentError occuring in a virtual env while installing tensorflow". These are basically the questions related to tensorflow, its module and other

frameworks when trying to integrate with tensorflow is a kind of issue developers ask about. Beside this topic is the most popular topic among developers having an average score of 1381.So from this we can say that posts from topic tensorflow visualize the most .

Model: In this category we mainly put topics that post related to AI models such as binary classifier , CNN model to classify images and so on. We have put 3 topic under this classifications and they are :

Classification: This topic mainly deals with posts related to AI's classifications models such as binary classification , image classification model and in general classifications post that is related to AI. For an example, post from this post look like this "How to check image classification model used by teachablemachine used for image classification?"

Neural Network : This topic mainly deals with posts related to neural networks for example posts regarding convolutional layers,cnn,weight and biases. For example, this post looks like this "Plotting decision boundary for a neural network with two layers?". It has the highest average view among the model category having 1366 average views. Beside this it has the highest average score (0.96) and average score(2.07). So this score states that post from this topic, developers find it the most helpful and are more likely to use this knowledge during their development process in near future.Beside this post from this topic is also the most interesting topic among the developers .

Keras :This topic mainly deals with posts related to a library that is used to solve machine learning problems known as keras. For example, posts from this topic look like this " how to predict multiple steps for multiple targets in keras?" Post from this topic is also very popular among developers.

On average Model has a similar average view for each topic like Tensorflow having more than 1100 views on average which shows that these topics in Model are very popular and posts from these topics are often visualised .

Table 6: AI development topics,category and their popularity.

Topic	Main Category	Posts	Avg View	Avg fav	Avg Score
Tensorflow	Integration	12663	1381	0.49	1.52
Classification	Model	6002	1156	0.59	1.27
Data	Preprocessing	12565	964	0.48	1.93
Keras	Model	6450	1127	0.40	1.46
Neural Network	Model	3706	1366	0.96	2.07
Compilation Error	Error	297	856	0.16	0.70
Loss function	Evaluation	1286	909	0.46	1.51

Preprocessing : This category deals with the topic named "Data". Basically this topic has all the questions related to data augmentation , dataset,pre process steps that were done on the sample data and so on. For example : "Will learning curves be good to use if the data is skewed?". Data has the 2nd highest number of posts among other 6 topics which shows that this topic is hot among the developers.

Error : This category deals with the topic named "error". Basically this topic has all the questions related to different types of error messages developers get when running an AI model. For example "'TensorBoard' object has no attribute 'writer' error when using Callback.on_epoch_end()". It has the lowest number of posts which shows that it is the least popular among the developers having just 297 posts. Beside this it has the lowest average favourite and it has the least average which shows that its posts are not very helpful as well the least interesting for the developers.

Evaluation : This category deals with the topic named "Loss function" and this topic has all the questions related to the loss function of an AI model to evaluate the result and accuracy of a model. For example : "how to use weighted_categorical_crossentropy loss function".It has the 2nd lowest number of posts which shows that it is also that popular among the developers having just 1286 posts.

Summary of the analysis :

1. Most posts are related to **"Tensorflow "** and **"Data".**
2. Post from **"Neural Network"** seemed to be most helpful for the developers and it tends to recurren more during android development.
3. The most popular topic in android developments is **"Tensorflow".**
4. Most interesting topics in android development is **"Neural Network"**
5. The least interesting topic in android development is **Compilation Error.**
6. The least popular topic is **"Compilation Error"** with developers facing the most issues in this topic and not very helpful to the developers.
7. In general the popularity of all the topics are very similar as the average view of each topic is very close to each other .

- RQ2: What are the type of questions AI developers are asking ?

From table 7 ,we can say that around 43.3 percent of the questions are of 'how ' type followed by "why"(28.1 %) and "what"(16.7%) and "others"(11.9%). So from this , we can say that AI developers are looking for posts that ask about the ways to implement something. From the table we can also infer that , "keras" has the most posts of "how type"(53.3%). It shows that developers are asking questions like how to use this library and its module. For example, how to import a data module from the library or how to save a keras trained AI model.

The "Loss function" has a category with most "what " types(30%) which means developers wanted to know about the posts in this topic that ask about the meaning of something . For example, developers asking about the loss function they will use in the last layer and which will be more suitable and so on

From the table 7 , we can see that "error " has the most "Why" types(60%). These are the posts developers ask mainly when they run an AI model and it gives a compilation or any kind or error in general. So it is obvious that this topic will have the maximum post because they will want to know why this error happened followed by what type to know what these errors are.

There are around 11.9 % posts that are classified as "others" which means that there are a huge number of posts that are not directly cannot be classified directly as other 4 type as it's very confusing

to tag them what the posts are actually asking about. So further investigation should be done on this post to classify it .

Table 7: Percentage of AI development posts type in StackOverfloww

Topic	Main Category	How	What	Why	Others
Tensorflow	Integration	46.7	3.33	36.7	13.3
Classification	Model	40	13.33	26.7	20
Data	Preprocessing	60	6.7	23.3	10
Keras	Model	53.3	20	13.3	13.3
Neural Network	Model	50	20	20	10
Compilation Error	Error	10	23.3	60	6.7
Loss function	Evaluation	43.3	30	16.7	10
All AI Development		**43.3**	**16.7**	**28.1**	**11.9**

- RQ3: Which is the most difficult question to answer in AI development?

From the table 8 we can see , overall every topic has a huge number of posts with unaccepted answers. Compilation error being the highest with 71.4% of the posts without an accepted answer. After analyzing each post of Compilation error we found out that the posts without accepted answers have very low scores by developers which may mean that the post information is not very structured and difficult to understand and more likely not get an accepted answer. Followed by Compilation error ,the most difficult posts are related to Tensorflow(67.5%). Neural Network has the least unaccepted answer(59.5%) and after analysing the post we say that the average scores for Neural Network are higher compared to all other topics. That gives us an insight that these posts are easily understandable and tend to give answers that have high probability to be accepted .

Table 8: Percentage of AI Posts without Accepted Answers

Topic	Percentage of Posts without Accepted Answers
Tensorflow	67.2
Classification	63.0
Data	63.8
Keras	64.1
Neural Network	59.5
Compilation Error	71.4
Loss function	63.7

Summary of the analysis :

1. Posts related to the **Neural Network** tend to be answered more frequently .
2. Posts related to **Compilation error** have the most difficult posts to answer.

4.4 Comparison Analysis Between Android and AI development:

- Comparison Analysis Based on Question 1 : Table 9 shows the differences and similarities between the difficulties faced by developers of the Android and AI development domains and also how the topics differ in each domain in terms of popularity and interests from the developers.

Table 9: Android vs AI - RQ1

Android Development	AI Development
Most Topic are related to Development and that is most posts come from android studio issues	Most posts are related to AI Models ;these are the issues related with data preprocessing and tensorflow libraries
Issues from android studio are the most useful posts for the developers and more like to take help from these posts in their development	Issues from the AI model "Neural Network" seem to be the most helpful for the developers more like to take help from these posts in their development
Developers face most difficulties understanding topics from UI and Database	Developers face most difficulties understanding posts related to errors occurred during AI development
The most popular topic(Google API) in android developments is the most helpful for the developers having high favourite scores. This gave an analysis that popular posts seemed to be understandable and well informed	The most popular topic (Tensorflow) in android developments is the least helpful for the developers having low favourite scores. This gave an analysis that popular posts seemed to be hard to understand and ill informed
Most interesting topic in android development is from UI	Most interesting topic in android development is from the AI model "Neural Network"
The least interesting topic in android development is related Database.	The least interesting topic in android development is Compilation Error.
Popularity of all the topics are evenly spread	Popularity of all the topics are evenly spread.

- Comparison Analysis Based on Question 2 : Table 10 describes the differences and similarities to answers based on research question 2. It identifies how each domain contained different types of questions and the distribution of question types for each topic, compared amongst the two domains.
- Comparison Analysis Based on Question 3 : Table 11 shows the level of difficulty that each domain have in terms of each topic labelled for the respective domains. The difficulty was based on the percentage of posts without accepted answers. It contains the similarities and differences for our answer to research question 3 compared between the two domains.

Table 10: Android vs AI - RQ2

Android Development	AI Development
Most of the posts are of 'How' type	Most of the posts are of 'How' type
Least of the posts are 'What' type	Least of the posts are 'Other' types
Overall there are around 16.2% posts that are classified as "others"	Overall, there are around 11.9% posts that are classified as "others".
Questions in android developments are not well structured and ill informed as there are more % of posts that we were unable to label as Why, What or How compared to AI.	Questions in AI developments are well structured as there are a less % os post that we were unable to label as Why, What or How.

Table 11: Android vs AI - RQ3

Android Development	AI Development
% post without accepted answer is 58.4	% post without accepted answer is 0.64
Posts related to"Graphical Interface" tend to be answered more frequently	Posts related to the Neural Network tend to be answered more frequently
Posts related to "Android studio's issues" have the most difficult posts to answer followed by "Google API"	Posts related to Compilation error have the most difficult posts to answer.
Posts with accepted answers have high scores and favorite scores in general.	Posts with accepted answers have high scores and favorite scores in general.
Posts without accepted answers have low score and low favorite score in general.	Posts without accepted answers have low score and low favorite score in general.

5 EVOLUTION ANALYSIS

5.1 Evolution of AI

Artificial Intelligence has been a very interesting topic in recent years and more developers are working on this domain everyday. To analyse this evolution of AI throughout the years we will be examining two things: a) the number of posts related to Artificial Intelligence that were posted each year in Stack Overflow, and b) a topic wise analysis of the number of posts posted each year respective to each topic in the AI domain. Figure 3. Shows a graph plot for the number of posts that were posted each year related to AI. We can visualize the number of posts from 2013 to 2020 (this range was taken to keep consistency with android).

As we can see from the graph, the number of posts are increasing each year. 2020 shows a dip in numbers but that is only because we have data from 9months, and after the last 3 months we believe the numbers will be higher than 2019. We can notice a hike in numbers from 2016 to 2017. The reasons we could find after researching that in 2017 a lot of AI developments took place such as Google creating

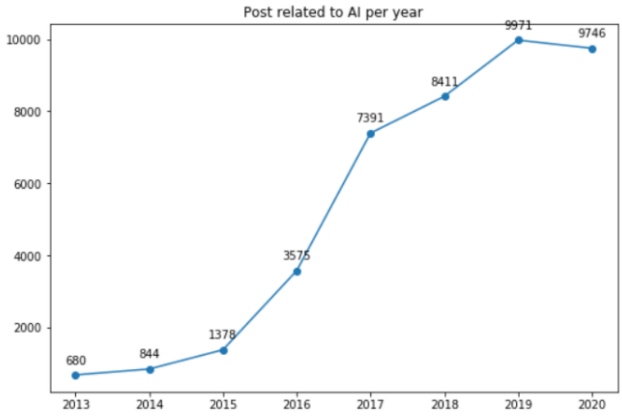

Figure 3: Year wise evolution of AI posts

AutoML which is an Ai creating AI, Nvidia creating fake faces of humans using AI that looks real, and the world recognizing the first AI citizen. Now to look further into the evolution, we will look into the topic wise posts that were posted each year for the AI domain, presented in Figure 2.

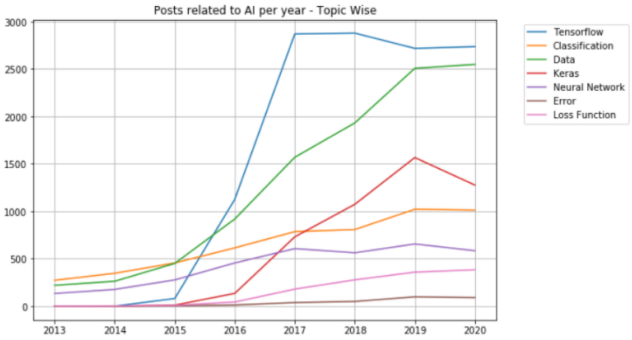

Figure 4: Topic wise evolution of AI posts

Looking into figure 4 we can see that almost all the topics have a positive gradient, suggesting increasing posts each year. Here again, 2020 only has data for 9months, so a full year's data is expected to be higher than the year 2019. It can be noticed that since 2015 more and more questions related to data have been posted, which is when developers started to take advantage of Big Data to establish different systems. Also Keras and Tensorflow had a hike in the number of posts since 2016 as they were discovered in November and March of 2015 respectively and developers were learning the libraries since 2016. Overall, we can say that AI has been a hot topic since the year 2016 and it will continue to be one since a number of functions and utilities are being discovered for AI each day.

5.2 Evolution of Android

Similar to AI domain, we will also be looking into a) the number of posts related to android posted each year in Stack Overflow and b) the number of posts in each topic that were posted each year.

Figure 3. Shows the graph of the number of posts related to android posted each year.

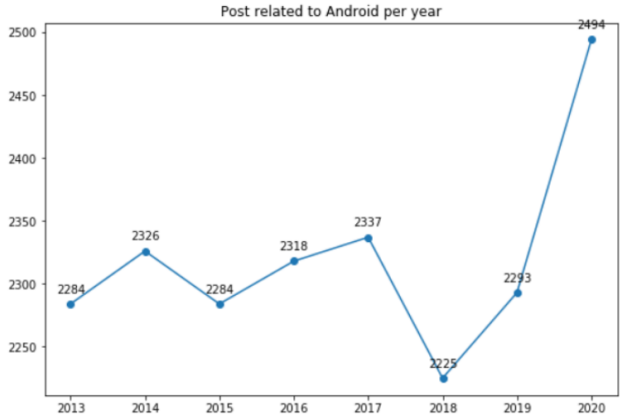

Figure 5: Year wise evolution of Android posts

From Figure 5, we can see that the number of posts related to android amongst the topics we selected have a general positive increase in number. However we can notice a drop in 2015 and 2018 but this could be because the initial data we extracted for android was randomly chosen for about 150 000 posts and this may have led to a number of instances from these two years to be dropped out.

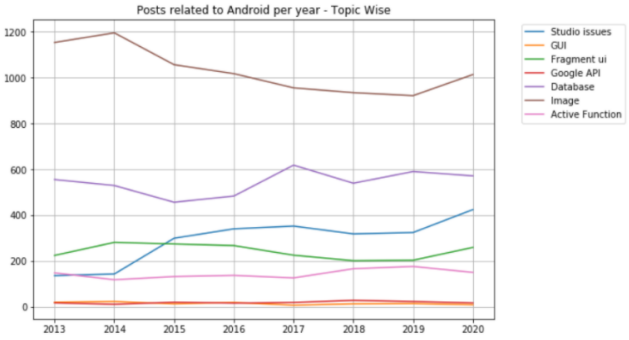

Figure 6: Topic wise evolution of Android posts

Now let us examine the year wise number of posts for each topic related to android, given in Figure 6. From the graph, we can see that in general all the topics have a constant gradient and this is because android is not a new domain and developers have been working for years in this domain thus the posts have remained similar over the years. However, it must also be noticed that compared to other topics, posts for android related to image or queries about image editing in android are the highest each year. On the other hand, posts related to Graphical User Interface and Google API are the least each year. As we can recall from our first research question summary, developers face the least issue in GUI and posts related to Google API are quite repetitive thus not many new posts appear each year.

In general we can summarize that the number of posts in android are still rising each year as developers continue to work with android but most topics have similar number of posts each year and the numbers are not that high.

5.3 Trends in Unanswered Posts

To determine the trends in unanswered posts for each domain we have plotted two separate graphs of the ratio of unanswered posts to total posts for that year against the year for each domain. The graphs can be seen in figure 7 and 8.

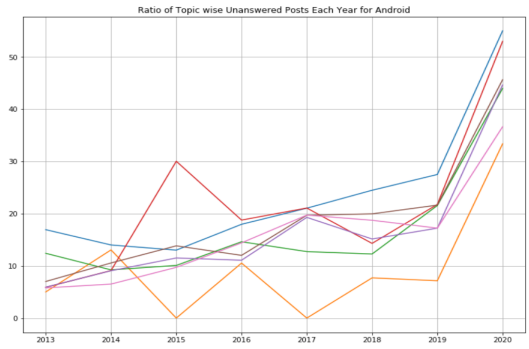

Figure 7: Unanswered posts ratio per year for android

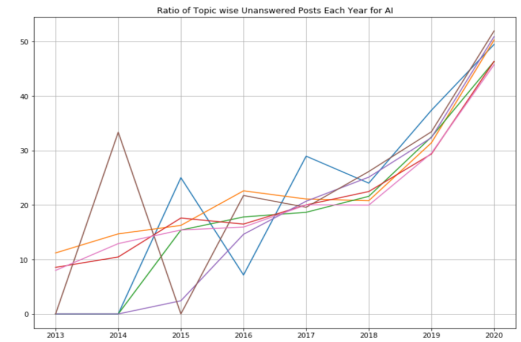

Figure 8: Unanswered posts ratio per year for AI

From the graphs we can identify a general increasing trend of number of unanswered posts each year for both the domains but the increase for AI unanswered posts is more steep than that of android suggesting that AI developers are facing new problems every year and other developers of the same domain are finding it difficult to solve these problems.

6 THREATS TO VALIDITY

6.1 Internal Threats

Internal validity concerns factors that could have influenced our results. We selected tags related to both android and AI from the repository of Stack Overflow. This could result in the fact that some posts are mislabelled and we did not retrieve them from the repository of StackOverflow. However we tried to mitigate this

threat by applying NLP techniques on both the title and the tags of the posts to search for the related tags. Therefore even if a post is mislabelled, the related topic could be extracted from the post title. One other potential threat is the selection of the optimal value of K = 7 for the LDA. This directly relates to the quality of the topic modelling of the LDA and is known to be difficult to select in general. However we tried to select the best value for K using approaches from previous works and examining the coherence of the topics as well. Lastly, the labelling of post types is also a threat to validity since we have manually labelled them ourselves. Nevertheless we have tried our best to label them properly using the process where both the authors have manually labelled the post types and then matched the labels. For the ones that did not match we both voted to select the best type to our knowledge with proper justification.

6.2 Construct Threats

Construct validity considers the relationship between theory and observation. Our labelling of the topics from the LDA might not be consistent with the posts associated with the topics. Here also, we tried to avoid this threat by using the same process of both the authors deciding on the topics and then justifying on the topics that did not match between ourselves to decide the topic. We have also taken about 30 random posts from each topic and manually examined them to determine that the topics were chosen properly.

6.3 External Threats

Threats to external validity concern the generalization of our findings. In this study we extracted data from posts on Stack Overflow, and there are still other forums that have similar discussions of developers regarding the chosen domains. However we believe that our study allows the generalizability since Stack Overflow is one of most popular forums for developers' discussions and a huge number of questions can be found from its repository. Lastly, another threat to validity of this paper could be the fact that we extracted only about 150 000 data from the repository instead of the whole data because of computation resources and also manual analysis of such a large dataset will not be possible in the time we performed this study. Nevertheless, we did try to extract proper related posts from the repository and our study findings support the general ideas we have about the domains, therefore the generalizability should not be a problem.

7 CONCLUSIONS

In this paper we analyze the challenges Android developers and AI developers faced. This paper also did a comparison analysis between these 2 development industry and try to find some insights. Beside this we also try to understand the the characteristic of developers of each domain and try to come up with some insights. We find out 7 topics for each development's domain that has been categorized into few main categories. Android developers are highly interested in posts that are related to UI whereas it is related to AI model like Neural Network and classification AI development. Questions in android developments are not well structured and ill informed compare to AI development. AI posts are more popular than Android posts and AI developers faced more challenges compare to Android developers.Developers face most difficulties understanding posts related to errors occurred during AI development whereas it is UI and Database in AI development. So from this we can say that AI development is more hard compare to Android development.Therefore more documentations, tutorials and improved tools should be introduced to aid the AI Developers. The evolution of AI and Android posts show that AI has been a more popular domain since 2016 and the number of posts are rising every year, whereas the android related posts have somewhat maintained a similar number of posts each topic over the period of years 2013 to 2020. However the analysis of unanswered posts also show that AI domain have significant increase of unanswered posts each year which means that developers are facing new problems each year that other developers are not able to solve that easily. The Trend for Android's unanswered posts also show an increase each year but that may be because android developers already have similar problems posted in other questions and thus these questions remain unanswered.

Our study has opened a lot of opportunities for further research. We would like to extend this research in the future for a greater set of data and from other forums as well. Also we would like to perform similar analysis for more domains and find insights in a large scale about each domain. This would also help us to understand the issues of each domain better and then suggest the implications of the findings we obtain. One other analysis we aim to perform is the median time that the posts in each topic take to get accepted answers and this would allow us to better understand the difficulty of the topics. An important addition from this research would be to conduct interviews with developers of the respective domains to know about their struggles and use these interviews to find insights that support our empirical study. Finally, we would like to investigate the repositories of each of these domains by examining the commits and bug reports to find the issues that developers are facing in the respective domains and compare with our findings.

REFERENCES

[1] 2020. bigquery: cloud data warehouse google cloud_2020. =https://cloud.google.com/bigquery.

[2] 2020. Stack exchange data dump : stack exchange, inc. =https://archive.org/details/stackexchange.

[3] Ahmad Abdellatif, Diego Costa, Khaled Badran, Rabe Abdalkareem, and Emad Shihab. 2020. Challenges in chatbot development: A study of stack overflow posts. In *Proceedings of the 17th International Conference on Mining Software Repositories*. 174–185.

[4] Syed Ahmed and Mehdi Bagherzadeh. 2018. What do concurrency developers ask about? a large-scale study using stack overflow. In *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. 1–10.

[5] Mehdi Bagherzadeh and Raffi Khatchadourian. 2019. Going big: a large-scale study on what big data developers ask. In *Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 432–442.

[6] Kartik Bajaj, Karthik Pattabiraman, and Ali Mesbah. 2014. Mining questions asked by web developers. In *Proceedings of the 11th Working Conference on Mining Software Repositories*. 112–121.

[7] Vangie Beal. 2020. mining software repositories - MSR. https:https://www.webopedia.com/TERM/M/mining-software-repositories-msr.html.

[8] Shashank Kapadia. 2019. Evaluate Topic Modeling and Latent Dirichlet Allocation (LDA). https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24?fbclid=IwAR2l_dgT2_gW1BD6eC8S288mH1tufJNVY5kc_oqkeCQz51Dw6zZF96zLiCU.

[9] Susan Li. 2018. Topic Modeling and Latent Dirichlet Allocation (LDA) in Python. https://towardsdatascience.com/topic-modeling-and-

latent-dirichlet-allocation-in-python-9bf156893c24?fbclid=IwAR2l_dgT2_gW1BD6eC8S288mH1tufJNVY5kc_oqkeCQz51Dw6zZF96zLiCU.

[10] Edward Ma. 2018. 2 latent methods for dimension reduction and topic modeling. https://towardsdatascience.com/2-latent-methods-for-dimension-reduction-and-topic-modeling-20ff6d7d547.

[11] Sarah Nadi, Stefan Krüger, Mira Mezini, and Eric Bodden. 2016. Jumping through hoops: Why do Java developers struggle with cryptography APIs?. In *Proceedings of the 38th International Conference on Software Engineering*. 935–946.

[12] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*. 399–408.

[13] Christoffer Rosen and Emad Shihab. 2016. What are mobile developers asking about? a large scale study using stack overflow. *Empirical Software Engineering* 21, 3 (2016), 1192–1223.

[14] Haoxiang Zhang, Shaowei Wang, Tse-Hsun Chen, and Ahmed E Hassan. 2019. Reading answers on stack overflow: Not enough! *IEEE Transactions on Software Engineering* (2019).