Sharing Open Deep Learning Models

Ayse Dalgali
Syracuse University School of Information Studies
ayocal@syr.edu

Kevin Crowston
Syracuse University School of Information Studies
crowston@syr.edu

Abstract

We examine how and why trained deep learning (DL) models are shared, and by whom, and why some developers share their models while others do not. Prior research has examined sharing of data and software code, but DL models are a hybrid of the two. The results from a Qualtrics survey administered to GitHub users and academics who publish on DL show that a diverse population shares DL models, from students to computer/data scientists. We find that motivations for sharing include: increasing citation rates; contributing to the collaboration of developing new DL models; encouraging to reuse; establishing a good reputation; receiving feedback to improve the model; and personal enjoyment. Reasons for not sharing include: lack of time; thinking that their models would not be interesting for others; and not having permission for sharing. The study contributes to our understanding of motivations for participating in a novel form of peer-production.

1. Introduction

Deep learning (DL) models (e.g., AlexNet or GoogLeNet) [1] refer to a more advanced type of machine learning (ML) that uses neural networks to learn a complex mapping of inputs to outputs (e.g., from an image to a label for the image). There are many DL applications, from image recognition to machine translation. Although neural networks were first discussed in 1943 [11], DL models have yielded more satisfactory performance in the last ten years, as indicated by the increase in DL applications.

Neural networks are structures patterned on the function of a human brain, more specifically, to mimic how neurons in the human brain work. In a human brain, neurons receive inputs and apply a non-linear interaction to compute an output. Similarly, in a neural network, artificial neurons act as computational nodes between inputs and outputs (as shown in Figure 1). When inputs enter the neuron, they are multiplied by an associated weight. The sum of the multiplication (inputs and associated weights) is then translated to an output signal via an activation function. The term deep learning refers to neural networks with complex architectures that have many layers of neurons between the inputs and outputs.

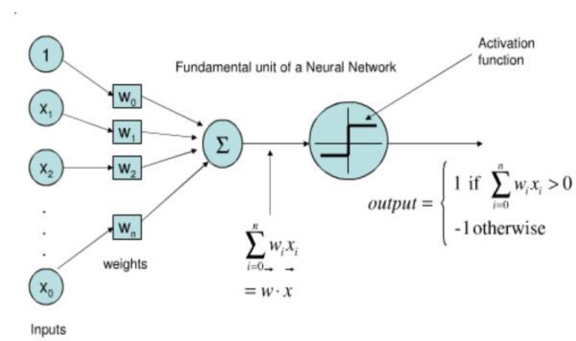

Figure 1. Mathematical model of an Artificial Neuron [3]

Training a neural network from scratch means determining the network configuration [9] and then adjusting the weights throughout the network. Weights are initialized randomly and updated during the model training as the network is given a large set of input images, text or sound. Thus, designing a neural network and training it from scratch requires a lot of effort, time, and training data.

Pre-trained DL models are models for which their weights can be downloaded and used without training from scratch. These pre-trained models may be used by others for a new application as is (model reuse), rather than building and training a new model. Another approach is to retrain only part of the network, e.g., training only the final layer while using the early stages of the model as they are to extract the learned features. This approach is called **transfer learning**, referring to a ML/DL method in which a DL model created to perform a task is reused as the starting point of another DL model for a second task. Reusing or fine-tuning a pre-trained network via transfer learning is usually much faster and easier than training from scratch.

Utilizing transfer learning is very common in many DL applications, such as computer vision and natural language processing tasks, because having the

URI: https://hdl.handle.net/10125/59650
ISBN: 978-0-9981331-2-6

amount of data and time necessary for training a DL model in these domains is very difficult [9]. The utility of pre-trained DL models has led people to share, modify, reuse and redistribute them. Thus, we offer a new concept, **Open Deep Learning Model (ODLM)** which we define as 'a DL model that the public can freely reuse, modify and redistribute'.

For actually executing ML/DL models, various open source libraries are used, such as TensorFlow, Caffe, Torch, Keras and Theano. Many of these libraries are themselves open source software. However, in this study, we focus on sharing and reuse of DL models, rather than sharing open source libraries or general open source usage/contribution behavior for ML/DL, as will be discussed below.

1.1. Research Questions

The goal of this study is to investigate the reasons for sharing ODLMs, as well as reasons for not sharing DL models. The research questions in this study are:

Research question 1: Where and by whom are ODLMs shared?

Sub question 1.1: What are the differences or similarities between sharing ODLMs and sharing data or software?

Research question 2: Why do some DL developers share their DL models, while others do not?

1.2. Problem Statement & Significance of the Study

Platforms where ODLMs are shared provide the public the opportunity to use, to distribute, and to contribute to the development of new models. Moreover, these platforms enable DL developers to collaborate, which facilitates development of more complicated, useful, and advanced DL models in a shorter time through "accelerating scientific progress and faster adoption of machine learning methods in other disciplines" [2].

Sharing, reusing, and discussing ODLMs in these platforms can lead to the development of models for future applications of DL. Hence, it is important to understand how developers share and use ODLMs. Furthermore, in the information age, accessing scientific products (e.g., scientific data and software code) is quite easy and quick because sharing them via various tools (e.g., websites, databases, social media, blogs, online libraries, etc.) is quite common and practical. This study aims to explore whether the situation is similar for ODLM sharing. Since ODLMs are a product of code and training data, understanding how code and data are shared might provide insight into how ODLMs are shared. Hence, this study compares the similarities and differences between sharing ODLMs and sharing code and data.

This study contributes to understanding the reasons for sharing pre-trained DL models, as well as reasons for not sharing such models. Thus, it contributes to the collaboration of developing new ODLMs for promising applications and sheds light on a novel form of peer-production. To the best of our knowledge there is no work in the literature that focuses on directly sharing ODLMs. This study can fill this gap and contribute to the future work more concerned with peoples' experiences sharing ODLMs.

2. Literature Review

As mentioned before, because ODLMs can be seen as a mix of code to perform some kind of data processing and training data, understanding code and data sharing properly is crucial to understand how ODLMs are shared. Thus, we start with a review of research on sharing data and code.

2.1. Data Sharing

Data are the basis of sound scientific conclusions in research and are dependent on logical use and processing of data. Nowadays, "science is [more] data intensive and collaborative" [10] because the technological tools allow scientists to work quickly and build more connections with fellow scientists. Scientists and researchers share many types of data in their research, such as experimental data, observational data, survey data, and interview data. However, sharing outputs of research projects is more common than sharing input data [14]. Researchers usually share the outputs obtained from input data via publications such as research articles in journals, conferences or workshops. In addition, they also share at least some of their data results either on their organization's website, a national network, a global network, or a personal website [14]. Furthermore, Synthesis Centers [4] and Open Science Framework (OSF) are other platforms where data may be shared.

2.1.1. Data Sharing: Reasons for Sharing. Data sharing is seen as important for "improving data integrity and for enhancing transparency and reproducibility of the scientific enterprise" [4]. Data sharing can provide some personal advantages to scientists and researchers, including increasing citation rates [4]. Furthermore, if they share their raw data, other scientists can re-analyze and verify results in addition to potentially applying different insights or methods. All of this can improve the quality and efficiency of research processes and findings [14].

In a study [14], the authors conducted a survey with 1329 scientists in order to explore their current data sharing experiences. The study showed that scientists whose research field is atmospheric science (related to earth, planets etc.) shared their data more than the people from different disciplines [14]. On the other hand, data in medicine and social sciences are shared less than the data in other disciplines [14]. People over 50-years-old tend to share their data more than those who are 20–39-years-old [14]. Respondents who are 20–39-years-old share their data if they have legal permission to share [14]. For those over 50-years-old, permissions are not so important [14]. Besides, respondents between 40 and 50-years-old were less likely to believe that creating new datasets from previous datasets is possible than both the respondents who are 20–39-years-old and over 50-years-old [14].

In that study [14], the work focus is separated as "research" and "teaching". People who tend to do research more than teach are defined as "research-intensive," and those who tend to teach more than research are defined as "teaching-intensive". The difference in data sharing between research-intensive and teaching-intensive respondents are less than the difference that stems from discipline or age. Nevertheless, research-intensive respondents tend to share their data more than teaching-intensive respondents [14].

Another study [18] suggested that data sharing is increased by personal motivations such as career benefit (e.g., credits and reputation). The same study found that normative pressure positively affects social scientists' data sharing behavior: the scientists share data because it is a valuable norm in their research communities [18]. Finally, another study showed that organizational support for improving data quality is a key factor that motives scientists to share data [12].

2.1.2. Reasons for not Sharing Data. Scientists report that main reasons behind not sharing data are "insufficient time" and "lack of funding" [14]. In addition, [14] also reports that the risks of shared data rewarding other scientists than themselves, and ethical concerns are other reasons for not sharing data. [18] also suggested that the primary barriers to data sharing are effort (time and seeking funding) and lack of institutional support to reduce the effort. This study also found that human subjects' privacy and confidentiality constraints are other reasons for not sharing data. Finally, [12] pointed out concerns about data quality, time constraints, organizational constraints (e.g., permission to share), and legal and policy requirements.

2.2. Code Sharing

Since ODLMs perform some kind of data processing, they can be seen as a kind of software code. Technological developments have inspired people to share code for developing new software programs. With sharing, modifying and redistributing code, the concept of "open source software (OSS) [13]" has emerged. Having access to the source code means that users can also modify the program, thus facilitating collaboration.

Code can be shared in many venues. Today, one of the most popular places where code is shared is GitHub. GitHub users share files that include code so that other users can download and use it. The Open Science Framework (OSF) provides a facility specifically for sharing scientific code [5]. The author of [5] stated that "OSF is debuted in 2012 with an aim to increase sharing, collaboration, and transparency in research" (p. 76). He adds that OSF provides a free online platform so that researchers can share their data and code, thus making OSF suitable for project collaboration. Other sites where code may be shared include Bitbucket, Banyan, SciGit, figshare, and Zenodo [5].

2.2.1. Code Sharing: **Reasons for Sharing.** Researchers have studied developers' motivations for sharing code [e.g., 6]. One reason mentioned in [6] is that software developers share code in order to contribute to the community where new software is developed. In another study [7], helping people improve their programming skills thanks to community feedback and support was an important reason for sharing code. Other very strong motivations for sharing are enjoyment and being creative while contributing to software development. Another study [17] showed that software developers have intrinsic and extrinsic motivations. According to their study the main intrinsic motivation is altruism; the main extrinsic motivations are personal needs (e.g., efficient learning tools, communication with the community) and peer recognition. Writing higher-quality code, being "part of a community and benefitting from [also] the code shared by others, thus reducing software development time for ourselves and others [reusers]" [8] are other reasons for sharing code.

2.2.2. Code Sharing: **Reasons for not Sharing.** Despite the mentioned motivations and advantages of sharing code or data, some people hesitate to share code due to their concerns for not having sufficient legal rights, destroying the software industry, not finding code that is compatible with standards, and destroying intellectual property [15]. Intellectual property (IP) is defined as "creations of the mind, such as inventions; literary and artistic works; designs; and symbols, names and images used in commerce" [16]. IP is protected by laws regarding patents, copyright and trademarks. This allows people to obtain recognition or financial benefit from what they invent or create" [16]. Thus, IP protection is important for people sharing software as well. Furthermore, there is software industry that contain a community making money from software sales. Therefore, harming software industry is sometimes given as a reason for not sharing code.

2.3. Summary of Literature Review

Based on the literature review, we present a table (Table 1) that combines the people, locations, reasons for sharing and not sharing data and code, which informs our planned study of sharing and reuse of pre-trained deep learning models.

Table1.Summary of the Literature Review

CATEGORY	
People	Scientists
	Researchers
	Faculty (academics)
	Software developers
	Post-doctoral research associates
	Graduate students
	Undergraduate students
	Organizations
Locations	with research articles in journals
	in an organization's website
	on the author(s)' own website(s)
	in a national network
	in a global network
	Synthesis Center
	GitHub
	Open Science Framework (OSF)
	Bitbucket
	Banyan
	SciGit
	figshare
	Zenodo
Reasons for sharing	increasing citation rates
	improving the quality and efficiency of scientific progress and findings via applying different insights or methods to existing data by different people
	contribute the community
	career benefits (credits and reputation)
	improving knowledge thanks to community feedback and support
	personal enjoyment
	using creativity while contributing software development
	altruism
	personal needs (e.g., efficient learning tools, communication with the community)
	peer recognition
	writing higher-quality code thanks to showing other developers
	being part of the community
	solving problems which may be encountered during the development of the projects
	sharing is a valuable norm in their research communities
Reasons for not sharing	insufficient time
	lack of funding
	organizational constraints (permission to share)
	legal and policy requirements
	rewards others than themselves
	risks of bad reputation (concerns about the quality of data)
	ethical concerns (violation of privacy and confidentiality)
	not having sufficient legal rights
	destroying of intellectual property
	destroying software industry

3. Methodology

3.1. Research Design

First, in the research design section, we provide a brief overview of the research questions and the overall process to answer them**.**

3.2. Data Collection and Analysis

Since we want to explore experiences related to sharing ODLMs, our target population is DL developers. Hence, we designed a survey using an online survey tool, Qualtrics. We aimed to recruit respondents from different countries to have a diverse sample so that we can increase the transferability of the study.

Table 2. Research questions, methods, and expected findings

Research Questions	Method	Expected Findings
RESEARCH QUESTION #1		
Where and by whom are ODLMs shared?	Survey	Sites where ODLMs are shared and groups of people sharing ODLMs
Sub question 1.1: What are the differences or similarities between sharing DL models and sharing data or source code?	Comparing the results of the survey with the literature review	Discussion on the comparison of sharing DL models and sharing data or source code
RESEARCH QUESTION #2		
Why do some DL developers share their DL models, but others do not?	Survey	Reasons for sharing DL models Reasons for not sharing DL models

3.2.1. Survey Design. In the first page of the survey, we added an information letter to introduce our project and ourselves in order to gain respondents' trust and consent. Table 1, at the end of the literature review, includes factors regarding people, locations, reasons for sharing and not sharing. Since we aim to explore whether these factors are similar to ODLM sharing, we created survey questions based on those factors. After the preliminary information and a consent question, we provided questions from five different question blocks: experience sharing DL models, reasons for not sharing DL models, experience reusing DL models, reasons for not reusing DL models, and demographic information. The survey questions were designed based on the literature review.

3.2.2. Data Collection. The survey link was sent to DL researchers by email. We collected the email addresses of DL project contributors from GitHub profiles in addition to papers that provided DL model descriptions and authors' email addresses. We sent 118 invitation emails but received only 4 responses. Thus, to recruit more participants, we implemented a version of snowball sampling: in the invitation emails, we also requested that contributors forward the survey link to other contributors they were familiar with. By doing so, we aimed to increase the reliability, validity and generalizability of the study. In addition to emails, we shared the survey link in different DL project repositories on GitHub. To find the appropriate repositories we queried the GitHub search on 24 April 2018 using the keyword "Deep Learning (project)" to search repositories that involve DL projects. With emails and link sharing, we received and recorded a total of 117 responses from the survey.

3.2.3. Data Analysis Procedure. Based on the results of the survey, we made descriptive statistics and correlations to explore factors that affect DL models' sharing. For exploring the correlations, we used the R Chi-square test because we looked at the correlations between nominal data.

4. Results

In this section the descriptive statistics based on the survey are provided. Because the percentages are rounded, their total sometimes is different from 100%. The discussion section includes the relationships.

4.1. Locations for DL Model Sharing

Based on the survey, the most common sites where ODLMs are shared are GitHub (61%), research institute websites (15%), university websites (6%), personal websites (6%) and Caffe Model Zoo (6%) respectively. ODLMs are also shared on commercial organization websites although it is not as common (3%). We note though that the majority of the survey responses (114) were obtained from GitHub users, explaining the high fraction of that response.

Table 3. Sites where ODLMs are shared

Answer	%
Caffe Model Zoo	6%
GitHub	61%
GitXiv	0%
Personal website	6%
Research Institute website	15%
University website	6%
Commercial organization website	3%

4.2. People Sharing ODLMs

The majority of people reporting sharing ODLMs are PhD students (25%), researchers (22%), and computer/data scientists (19%). Then, researchers in industry also share ODLMs (10%), but not as many as computer/ data scientists or academics (13%) or researchers do (Table 4).

Table 4. Status of people sharing ODLMs

Answer	%
Academics	13%
Researcher	22%
Master student	3%
PhD student	25%
Undergraduate student	10%
Computer/Data scientist	19%
Employee in industry	10%

The majority of people sharing ODLMs are from computer science and engineering departments (68%). Those from mathematics departments are the second highest (9%). It is interesting that the people from medicine departments (4%) have a higher percentage than people from social science departments (3%), economics departments (3%), and atmospheric science departments (3%) (Table 5). In the survey, nobody chose mathematics alone as their focus area. They chose mathematics with another discipline such as: economics + mathematics, medicine + mathematics, computer science and engineering + mathematics. Furthermore, there is a diversity among the people sharing ODLMs. These results indicate that there are many DL applications in various domains.

Table 5. Departments of people sharing ODLMs

Answer	%
Computer science and engineering	68%
Medicine	4%
Social sciences (e.g., education, psychology, sociology)	3%
Economics	3%
Atmospheric science (e.g., fields are related to earth, planets)	3%
Mathematics	9%

4.3. Reasons for Sharing ODLMs

The most common reason for sharing ODLMs is contributing to the collaboration of new ODLM development (24%) (Table 6). The second most common reason is desiring to receive feedback to improve the model (18%). The third most common reason is providing a base for a new ODLM development (15%). Namely, these three reasons indicate that most people share their models to support the creation of new models, applications, and methods; all of which further research in DL. Other reasons recorded are "It is the norm in my area of work to share models" (6%) and "others expect me to share my models" (1%), but these are not so common reasons.

Internal motivations such as "increasing the citation rates of my papers" (13%), "getting a good reputation (13%)" and "having personal enjoyment" (10%) are also seen by participants as important reasons for sharing ODLMs.

Table 6. Reasons for sharing ODLMs

Answer	%
in order to increase the citation rate of my papers	13%
in order to get good reputation	13%
in order to contribute to the collaboration of new Deep Learning models' development	24%
in order to get feedback to improve the model	18%
in order to have personal enjoyment	10%
in order to provide a base for new Deep Learning models' development	15%
It is the norm in my area of work to share models	6%
Others expect me to share my models	1%

4.4. Reasons for Not Sharing DL Models

Participants indicated that the main reason for not sharing DL models is not having trained a DL model of their own (Table 7). Some participants also indicated that they have trained DL models but still do not share their models because they do not think their models would be of use or interest to others, or that they do not have permission to share their models.

Other commonly reported reasons for not sharing are: not having enough time (13%), concerns about losing the advantage from the models (6%), concerns about ownership of training data (6%) or thinking that sharing models is not the norm in their work setting (8%). Another reason was concerned with trust: 5% of respondents indicated they would only share models with reliable and experienced people. This, however, appears to not be a very significant reason for not sharing. Similarly, ethical concerns (such as risks of violations of ethical rules by people with malicious purposes) are not very common reasons for not sharing DL models according to this survey.

Table 7. Reasons for not sharing DL models

Answer	%
Because I haven't trained a Deep Learning model of my own	21%
I have ethical concerns (risks of violations of ethical rules by people with malicious purposes)	3%
I don't have enough time	13%
I don't find a safe place for sharing	2%
I want to share my pre-trained models with only reliable and experienced people whom I know (such as colleagues, professors, scientists etc.), not with everyone	5%
I have concerns about ownership of training data	6%
I don't think my models would be of use or interest to others	17%
I don't have permission to share the models or the data used to train them	13%
Sharing models is not the norm in my work setting	8%
I am concerned about losing my advantage from the models	6%

4.5. Correlations

We explored correlations between the factors that affect DL model sharing. While deciding which correlations are tested, first we looked at the relationships in data sharing and the relationships in code sharing mentioned in the literature review. We then tested other relationships, since we predict potential relationships based on the descriptive statistics and previous relationships mentioned in this study.

ODLM/Training Data Sharing: A correlation between DL model sharing and training data sharing is found as significant. 71% of the people sharing ODLMs also share training data. Only 29% of them do not share training data with the models.

ODLM/Source Code Sharing: DL model sharing and source code for ODLM sharing are also found as related to each other. The pie chart in Figure 2 shows the relationship between source code sharing and DL model sharing. 83% of the respondents of the survey reported that if they share ODLMs, they also share source code. Only 17% of them do not share source code although they share DL models.

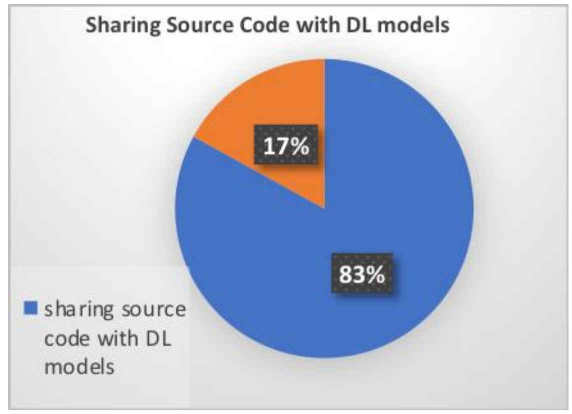

Figure 2. Source Code/DL model sharing

ODLM Sharing/Discipline: Discipline and ODLM sharing are also found linked to each other. Participants from the disciplines of computer science and engineering tend to share their models more. It is interesting that while people who focus on the medicine + mathematics combination do not share their DL models, people who focus on medicine+ social sciences do tend to share them.

50% of the people who do not have permission to share their models are students from computer science and engineering, mathematics and medicine; 17% of them are computer/data scientists. The remaining 33% are employees in the industry. Moreover, 80% of the respondents who express as a reason for not sharing that "sharing models is not the norm in my work setting" are employees in industry.

ODLM Sharing/Age: There is a relationship between sharing ODLMs and age. 64% of the respondents between 18-25 years old reported that they do not share any DL models (perhaps because they are still students). Similarly, 80% of the respondents who are between 25 and 32 reported that they do not share any DL models. 50% of the respondents who are between 32-39 years old share their models. People who are between 46-50-years-old tend to share their models more than other age groups: only 6% of the respondents between 46-60-years-old reported that they do not share DL models.

ODLM Sharing/Person: There is a relationship between ODLM sharing and participant-affiliation; whether a participant identifies himself/herself as an academic, computer/data scientist, master's student, etc. This study finds that researchers, graduate students or computer/data scientists tend to share ODLMs more than other groups of people. On the other hand, people who defined themselves as an industry employee were the least likely to share their models.

5. Discussion

The results of the survey are compared with the information in the literature review to answer the research question of similarities and differences between data, code and ODLM sharing.

The location for sharing: Both ODLMs and data are shared in a variety of locations. For example, ODLMs are shared on sites such as GitHub, Caffe Model Zoo, personal websites, research institute websites and university websites. Data can be also shared in different locations, such as university websites, journals, conferences, personal websites, research institute websites and organization websites. Some of the locations that data and ODLMs are shared are the same, such as, personal websites, research institute websites, and university websites. Despite this similarity, there is a difference in terms of sharing the same scientific product in two places at the same time: while data that are shared via journals (e.g., a paper) usually can be published in only a single journal, a DL model can be shared in more than one place, such as sharing the same model in GitHub and GitXiv at the same time.

As for code sharing, code may be shared in various locations as mentioned in section 2.2. Moreover, it can be shared in more than one place at the same time. GitHub is the most popular place for code sharing. Namely, GitHub is a place for both source code sharing and ODLM sharing.

The people sharing them: Data are usually shared by scientists, academics and researchers, based on the literature. There may be overlap: a person can be both an academic and a scientist working in a university and doing research; hence, one person can be a researcher, scientist and academic at the same time.

According to the survey, the majority of people sharing ODLMs are PhD students, researchers and computer/data scientists. But, the survey participants were DL papers' authors or GitHub users; therefore, it may exist a sample bias. This is because the majority of the sample may be PhD students. Then, researchers in industry also share ODLMs, but not as many as computer/ data scientists or academics or researchers do. Similarly, based on the literature, code is also usually shared by computer/data scientists, academics, and researchers. Sometimes researchers from giant software companies such as Facebook, Google, Microsoft also share ODLMs and source code.

Reasons for sharing: The main reason for sharing data (generally to support scientific papers) is increasing the citation rates. For ODLM sharing this reason is also common. In the survey, 12% of the respondents reported "increasing the citation rates" as a reason for sharing their ODLMs. Many ODLMs are shared with papers that describe these models. For example, it is written in Caffe Model Zoo that ODLM users should refer to the scientific papers that describe the relevant models. Namely, if someone uses a pre-trained model for his/her own DL application, he/she needs to cite the scientific paper where the DL model is explained (the ODLM developers are usually the authors of those papers). It looks similar to reusing a scientific paper. If we are using information from someone else's paper to write our own paper, we need to cite it.

Another reason that motivates people to share data and code is the potential to contribute to the research community and to further research. This reason also inspires ODLM developers: in the survey, this reason is seen as one of the most common reasons for sharing ODLMs (24%).

Literature demonstrates that age and discipline are factors that affect data sharing. In the literature review, it was mentioned that scientists whose research discipline is atmospheric sciences (it contains objective data) share their data more than people from other disciplines. On the other hand, people from the fields of medicine and social sciences are less likely to share data than the people from other disciplines as their data concern people.

As for ODLMs, the participant's discipline is also an important factor that affects sharing these models, but the situation is different from that happens in data sharing. Based on the survey, although the majority of respondents sharing ODLMs from computer science, people from medicine and social sciences surprisingly also tend to share many ODLMs. The reason for this might be that while sharing data in their field usually includes personal data, and thus sharing them may harm the people whose personal data is shared, when it comes to ODLM sharing may not harm anyone, it may actually help people. For example, sharing ODLMs in medicine specifically for healthcare purposes may help many people, thus the desire to share may be higher.

In addition, some of the common internal motivations for sharing ODLMs and code are to obtain a good reputation and for personal enjoyment. Feedback to improve the models is another reason for DL model sharing. Moreover, there are similar reasons for code sharing: improving their programming skills thanks to community feedback and support and personal needs (e.g., efficient learning tools, communication with the community).

Reasons for not sharing: Based on the literature, the main reasons for not sharing data are insufficient time, lack of funding, risks to rewards others, challenges in finding a safe place and ethical concerns such as confidentiality and privacy. The reasons for not sharing code are different: concerns about not having sufficient legal rights, destroying of intellectual property and destroying software industry. Based on the survey, the main reason behind not sharing DL models is lack of experience in doing so. If they have trained a DL model and they do not share it, however, it is likely because they do not think their models would be of use or interest to others; that or they do not have permission to share their models.

Age is also a factor that affects sharing ODLMs and sharing data. In the literature review, it was mentioned older people tend to share more data. This situation similar in ODLM sharing as well. The older researchers tend to share more. Based on the survey, younger researchers have less experience, more concerns about the quality of their DL models and legal permissions to share. For example, 40% of the respondents between 18-25-years-old who do not share their models reported that they have not trained a DL model; 40% of them do not think their models would be of use or interest to others. Similarly, respondents who are between 25-32 years-old and not sharing DL models mentioned same reasons for not sharing. Moreover, they added other reasons: "ethical concerns", "I want to share my models with only reliable and experienced people", "concerns about ownership of training data", and "losing my advantage of the models". It may be that industry or academia is more competitive for young researchers. Thus, they may do not want to share their models with others who may obtain rewards with their model. Because older scientists tend to be more established, often having better positions, they do not have the same concerns and share ODLMs more.

Other common reported reasons for not sharing DL models are not having enough time, concerns about losing the advantage from the models, ownership of training data and private intellectual property. The first three reasons here are similar to the reasons for not sharing data and more common in younger researchers: insufficient time and rewards to other scientists. Besides, although the literature indicates that ethical concerns is an important reason behind not sharing data, it is not as critical a reason for not sharing DL models.

In order to eliminate existing reasons for not sharing DL models, we present here four policy recommendations. First, policies are needed to protect owners of models from having their work expropriated. For example, the owners of the DL can allow others to use a model as long as others cite relevant papers in which the DL models that will be used are presented. Thus, the owners of DL models and training data do not lose the advantage from the models, ownership of training data and private intellectual property. Reviewers and editors need to be vigilant to be sure that such credit is given.

Second, governments encourage researchers to share DL models by providing grants and funding that require data sharing (including models). However, we found that 80% of the respondents who do not share DL models and express that "sharing models is not the norm in my work setting" were employees in industry, who likely are not supported by government grants. Therefore, other incentives will need to be created. Third, research is need on what to create DL models and to share training data in a way that does not violate the privacy and security of the data subjects. As a final recommendation, governments that invest in DL research should provide additional financial support to enable DL researchers sharing their DL models.

6. Conclusion

In this study, we investigated how ODLMs are shared, finding some similarities to sharing open source libraries (such as the libraries that execute ODLMs) or research data, but also some differences. This study contributes to understanding the reasons for sharing pre-trained ODLMs, as well as reasons for not sharing such models. Thus, it contributes to the collaboration of developing new ODLMs for promising applications.

Developing important DL models in a shorter time and with less data can be done via accessing and reusing existing models that have already been trained. More people tend to share their ODLMs and contribute to the collaborative work that goes into creating new ODLMs. Thus, we believe, in the future, ODLMs will be more popular and DL applications will continue to expand into various other domains.

References

[1] P. Ballester, and R.M. Araújo, "On the Performance of GoogLeNet and AlexNet Applied to Sketches" , In *AAAI*, 2016, pp. 1124-1128.

[2] A. Ben-Hur, C. S. Ong, S. Sonnenburg, B. Schölkopf, and G. Rätsch, “Support vector machines and kernels for computational biology”, *PLoS Computational Biology*, 4(10), 2008, e1000173.

[3] R. Cadène, N. Thome, and M. Cord, “Deep learning for visual recognition”, (Master's thesis). Retrieved from arXiv:1610.05567v1 [cs.CV] 18 Oct 2016, 2016.

[4] R.G. Curty, K. Crowston, A. Specht, B. W. Grant, and E. D. Dalton, “Attitudes and norms affecting scientists’ data reuse”, *PLoS ONE,* 12 *(12),* 2017.

[5] J. Kubilius, “Sharing code”, *I-Perception*, *5*(1), 2014, pp. 75-78.

[6] K. R. Lakhani, and E. Von Hippel, “How open source software works: ‘free’ user-to-user assistance”, *Research Policy, 32*(6), 2003, pp. 923-943.

[7] K. R. Lakhani, and R. G. Wolf, “Why hackers do what they do: Understanding motivation and effort in Free/Open source software projects”, In *Perspectives on Free and Open Source Software,* edited by Joe Feller, Brian Fitzgerald, Scott Hissam, and Karim R. Lakhani. Cambridge: MIT Press, 2005, pp. 3-21.

[8] R. LeVeque, “Top Ten Reasons to Not Share Your Code (and why you should anyway)”, https://sinews.siam.org/Details-Page/top-ten-reasons-to-not-share-your-code-and-why-you-should-anyway, 2016.

[9] MathWorks, “Transfer Learning Using AlexNet.”, https://www.mathworks.com/help/nnet/ug/pretrained-convolutional-neural-networks.html?s_tid=srchtitle#bvm8bu_, 2018.

[10] National Science Foundation, “Scientists seeking NSF funding will soon be required to submit data management plans”, http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928, 2010.

[11] G. Piccinini, “The First Computational Theory of Mind and Brain: A Close Look at Mcculloch and Pitts's ‘Logical Calculus of Ideas Immanent in Nervous Activity’”, *Synthese*, 2004, vol. 141, *(2)*, pp. 175-215.

[12] S. Sayogo, Djoko, and A. Pardo Theresa, “Exploring the Determinants of Scientific Data Sharing: Understanding the Motivation to Publish Research Data.” *Government Information Quarterly*, vol. 30, 2013, pp. S19-S31.

[13] C. Shen, and P. Monge, “Who connects with whom? A social network analysis of an online open source software community”, *First Monday*, 2011.

[14] C. Tenopir, S. Allard, K. Douglass, A.U. Aydinoglu , L. Wu, et al., “Data Sharing by Scientists: Practices and Perceptions”. *PLoS ONE* 6(6): e21101.https://doi.org/10.1371/journal.pone.0021101, 2011.

[15] D.A. Wheeler, “Why Open Source Software / Free Software (OSS/FS, FLOSS, or FOSS)? Look at the Numbers!”, http://www.dwheeler.com/contactme.html, 2015.

[16] WIPO, “What is Intellectual Property?”, http://www.wipo.int/about-ip/en/, 2018.

[17] Y. Wu, N. Wang, J. Kropczynski, and J. M. Carroll, “The appropriation of GitHub for curation”, *PeerJ Computer Science, 3*, 2017.

[18] K. Youngseek, and M. Adler. “Social Scientists’ Data Sharing Behaviors: Investigating the Roles of Individual Motivations, Institutional Pressures, and Data Repositories”. *International Journal of Information Management*, vol. 35, no. 4, 2015, pp. 408-418.