Consent Recommender System: A Case Study on LinkedIn Settings

Rosni K V, Manish Shukla, Vijayanand Banahatti, Sachin Lodha
TCS Research Labs, India
{rosni.kv,mani.shukla,vijayanand.banahatti,sachin.lodha}@tcs.com

Abstract

Privacy is an increasing concern in the digital world, especially when it has become a common knowledge that even high profile enterprises process data without data-subject's consent. In certain cases where data-subject's consent was taken, it was not linked to the proper purpose of processing. To address this growing concern, newer privacy regulations and laws are emerging to empower a data-subject with informed and explicit consent through which she can allow or revoke usage of her personal data. However, it has been shown that privacy self-management does not provide the expected results. This is mainly due to information overload as data-subjects use multiple services entailing variety of purposes, and hence, resulting in a very large number of consent requests. This may lead to *consent fatigue* as data-subject is now expected to provide informed consent for each associated purpose. The *consent fatigue* in data-subjects can lead to either incorrect decision making or opting for default values provided by the enterprise, and thus, defeating the purpose of new data privacy regulations.

In this work, we discuss the factors influencing the informed consent of a data-subject. Further, we propose a 'consent recommender system' based on Factorization Machines (FMs) to assist the data-subject and thereby avoiding *consent fatigue*. Our consent recommender system effectively models the interaction between the different factors which influence a data-subject's informed consent. We discuss how this setup extends for cold start data-subjects facing the decision problem with consent requests from multiple enterprises. Additionally, we demonstrate the scenario of consent recommendation as a prediction problem with minimum attributes available from LinkedIn's privacy settings.

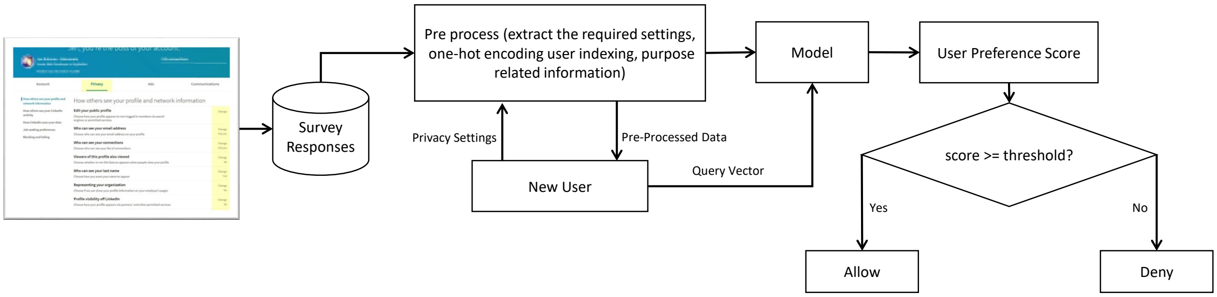
1 Introduction

With ever increasing digitalization we experience that enterprises capture consumer data for understanding their behavior and for offering better personalized services. More than often the captured data contains personal and sensitive information of the consumer (also referred to as 'data-subject'), and thus, leads to privacy concerns (Andrade, Kaltcheva, and Weitz 2002; Malhotra, Kim, and Agarwal 2004; Flavián and Guinalíu 2006). Till recently, the data privacy landscape was more enterprise centric with long and incomprehensible policy documents and default opt in for data sharing and usage (Cranor et al. 2013). In her work, Priya Kumar (Kumar 2016) discussed the specific ways in which vague or unclear language hinders the comprehension of enterprise practices. This paradigm represented one extreme of the data privacy management landscape where the data-subject had little or no control over her data with respect to its usage and sharing.

Some enterprises allowed data-subjects to access their data and provide consent for certain specific purposes such as sharing of personal email or demographic data with third party. However, such privacy preference controls provided by enterprises were either limited or there was a disconnect from privacy policy (Anthonysamy, Greenwood, and Rashid 2013) or it was hard to use them (Madden 2012). Further, these controls did not stop an enterprise from analyzing the data for gaining additional insights into data-subject's behavior. More recently, these concerns were addressed by newer privacy regulations and acts in different geographies, for example, GDPR in EU (Voigt and Von dem Bussche 2017) and CCPA in California (de la Torre 2018). These data protection regulations are designed to protect the personal information of individuals by restricting how such information can be collected, used and disclosed by having proper informed consent from data-subjects (Barnard-Wills, Chulvi, and De Hert 2016). For example, France's National Data Protection Commission (CNIL) penalized Google for not having a valid legal basis to process the personal data of the users of its services, especially for ads personalization purposes[1].

Informed consent is beginning to form the foundation of data protection law in many jurisdictions. It is intuitively considered as an appropriate method to ensure the protection of a data-subject's autonomy as it allows her to have control over her personal data (Voigt and Von dem Bussche 2017; Dwyer III, Weaver, and Hughes 2004). However, if a data-subject interacts with multiple services having consent requirement for many purposes (defined in Section 3) then it leads to information overloading while making decision, and hence, *consent fatigue*. In biomedical domain *consent fatigue* is a well discussed topic (Ploug and Holm 2013). Solove (Solove 2012) and Casteren (Casteren 2017) have studied about consumer's privacy self-management and their

[1] https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc



[redacted], you're the boss of your account.

[redacted]

Account | Privacy | Ads | Communications

How others see your profile and network information

How others see your LinkedIn activity

How LinkedIn uses your data

Job seeking preferences

Blocking and hiding

How others see your profile and network information

Edit your public profile
Choose how your profile appears to non-logged in members via search engines or permitted services
Change

Who can see your email address
Choose who can see your email address on your profile
Change
Only you

Who can see your connections
Choose who can see your list of connections
Change
Only you

Viewers of this profile also viewed
Choose whether or not this feature appears when people view your profile
Change
No

Who can see your last name
Choose how you want your name to appear
Change
Full

Representing your organization
Choose if we can show your profile information on your employer's pages
Change
Yes

Profile visibility off LinkedIn
Choose how your profile appears via partners' and other permitted services
Change
No

Figure 1: Recommender System Overview

ability to make meaningful decisions with information overload. A recent study (Degeling et al. 2018) discusses the impact of GDPR on web applications and services as well as new issues arising from the same. Two key takeaways from their work are: a) The majority of websites updated their privacy policies in the last two years, and, b) Average text length in policy document rose from a mean of 2,145 words in March 2016 to 3,044 words in March 2018 (+41% in 2 years) and increased another 18% until late May (3,603 words). The *consent fatigue* may either result in wrong decision making by data-subject or providing implicit consent by not taking any action.

In this work, we explore the problem of consent fatigue due to information overload and frequent decision making. To address this issue we proposed and implemented a consent recommender system for LinkedIn application. Our work enables a LinkedIn user in identifying appropriate privacy controls and its corresponding setting. It is especially useful for cold-starting a new user for whom no prior historical privacy preferences are available. The main contribution of our work consists of a novel combination of Factorization Machine (FM) (Rendle 2010; 2012) and factors affecting an individuals decision making process for predicting their privacy preference. That said, the details of our contribution are as follows:

- We conducted a survey on 50 data-subjects to identify factors that can influence their decision-making process. Further, we collected LinkedIn privacy setting data for each participant for building our recommendation model.
- In this work we have shown that the privacy recommendation problem can be modeled as a prediction problem. For that we used Factorization Machine (FM) (Rendle 2010; 2012) for consent recommendation. This also helped in analyzing the pairwise interaction of attributes for learning reliable weights. Further, we showed that the accuracy of our proposed model is around 88%. Also, we discussed the change in accuracy (in terms of precision, recall and F1-score) with respect to the different combination of features.

The rest of the paper is organized as follows. Related work is presented in section 2. Architecture and system description are given in section 3. The survey methodology, demography details and result analysis are discussed in section 4. The experimental results are shown in section 5. Section 6 describes the implication of our work, future research possibilities and the limitation of our work with some concluding remarks in section 7.

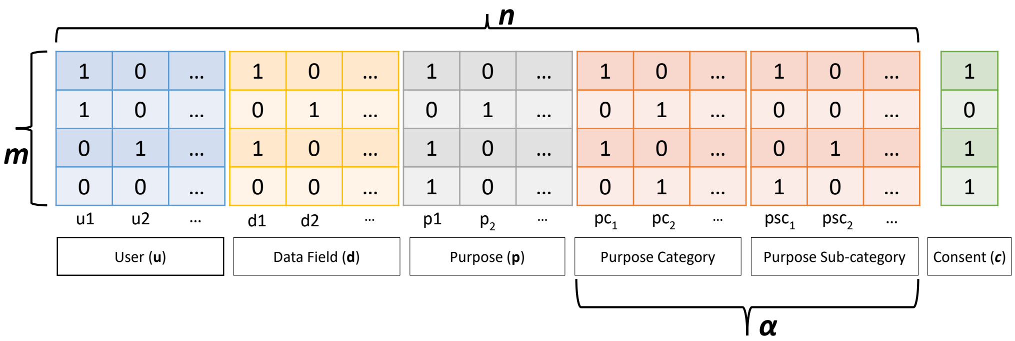
2 Related Work

Often services and applications capture more than required user data for analytics or generating profit by selling it to third party. An example of this was discussed in (Balebako et al. 2013) where they showed that even well-known mobile applications capture sensitive data of data-subjects and then share it with third party without their cognizance. However, with latest data privacy regulations a data-subject's consent becomes necessary to process her data. Substantial amount of work is done for understanding privacy concerns of data-subject (Liu et al. 2016; Olejnik et al. 2017; Knijnenburg 2014; Sadeh and Hong 2014; Liu, Lin, and Sadeh 2014; Sadeh et al. 2009; Wijesekera et al. 2017).

In their work, Sadeh et al analyzed the sensitive data requested by a mobile app and the purposes associated with it (Sadeh and Hong 2014). Liu et al, detected user profiles based on the user application permission settings (Liu, Lin, and Sadeh 2014). They further used Singular Value Decomposition (SVD) for addressing the issues related to sparsity and dimensionality. In (Wijesekera et al. 2017), authors reduce the burden on users by automating the decision making process in smartphones.

Researchers have also looked into the privacy preference recommender system for social networks. Ghainour et al (Ghazinour, Matwin, and Sokolova 2016) proposed a recommender system for privacy settings in social networks, particularly for Facebook. They modeled user's Facebook privacy settings of photo albums by independently considering different attributes, for example, personal profile and interests. In this paper, we also make use of the pairwise interaction of attributes. As it helps in learning reliable weights by taking the inner product of lower dimensional vectors.

In a recent work, (Naeini et al. 2017) focused on privacy expectations and preferences in IoT data collection scenarios. Naeini et al (2017) further showed that privacy preferences are diverse, context dependent and participants are more likely to consent to data if it benefits them. Additionally, they were able to predict data-subjects preferences after three data-collection scenarios. The work presented in (Naeini et al. 2017) comes closer to our work. However, their main focus is on improving the privacy notices for IoT



u1	u2	...	d1	d2	...	p1	p_2	...	pc_1	pc_2	...	psc_1	psc_2	...	
1	0	...	1	0	...	1	0	...	1	0	...	1	0	...	1
1	0	...	0	1	...	0	1	...	0	1	...	0	0	...	0
0	1	...	1	0	...	1	0	...	1	0	...	0	1	...	1
0	0	...	0	0	...	1	0	...	0	1	...	1	0	...	1
User (**u**)			Data Field (**d**)			Purpose (**p**)			Purpose Category			Purpose Sub-category			Consent (**c**)

Figure 2: Input Matrix to Factorization Model. Where, $\boldsymbol{\alpha}$ is the set of attributes, m is the number of samples and n is the number of features. For further description refer to Section 3 and 3.1.

devices and develop more advanced personal privacy assistants, whereas, we are addressing the problem of information overload, and hence, the issue of *consent fatigue* in post GDPR and CCPA era.

3 System Description

Definitions: Some basic definitions of the terms as per GDPR (Voigt and Von dem Bussche 2017):

1. *data-subject* is an individual whose personal data is collected, held or processed. In this paper terms consumer and data-subject are used interchangeably.
2. *personal data* shall mean any information relating to an identified or identifiable natural person ('data subject')
3. *consent* is defined as a data-subject's informed and unambiguous agreement to process her data.
4. *purpose* of processing data refers to the need and unambiguous reason for collecting, accessing and processing data-subject's data.

Problem Statement: Let U be the set of data-subjects such that $U = \{u_1, \dots, u_N\}$. Further, let S be a service provider (LinkedIn in our case), that processes large amount of data fields $D = \{d_1, \dots, d_K\}$. Let $P = \{p_1, \dots, p_X\}$ be the set of clear and unambiguous purposes under which S processes D. For a given purpose $p_i \in P$, there is an associated $D_i \subseteq D$. The service provider S will only process D_i for the purpose p_i. Similarly, a data field $d_j \in D$ could be linked to multiple purposes $P_j \subseteq P$. Also, purpose p_i is associated with a set of attributes (α_i) (e.g., description, purpose category, sensitivity of requested data field, etc.), such that $\boldsymbol{\alpha} = \bigcup_{i=1}^{X} \alpha_i$.

Figure 1 describes the overall flow of our proposed recommendation system. We selected LinkedIn for building our recommendation model because its a popular professional networking site and we found their privacy settings very comprehensive, including, handling of GDPR related concerns[2]. The modification in their policy was notified via a banner on their landing page. In case a data-subject keeps on using their service without modifying any settings then it is considered as implicit consent which is discussed by (Degeling et al. 2018). We extracted the privacy setting of each participant in our experiment. The collected data is processed to create a suitable feature vector for training the FM model using TensorFlow (Abadi et al. 2016). We tested the accuracy of model by splitting the collected data into training and testing and reported the results in Section 5.

3.1 Factorization Machines (FM)

Our data is described in the matrix format $X \in \mathbb{R}^{m \times n}$, wherein, $\mathbf{x}^i \in \mathbb{R}^n$ is the i^{th} row that represents the combination of a data-subject and a particular privacy setting with additional attributes as binary indicator variables. The response variable $y^i \in \mathbb{R}$ represents the consent value for i^{th} feature vector. Figure 2 shows the input matrix representation used in this work.

Why FM for Consent Recommendation? The Equation 1 shows the traditional linear regression model, where, $w_0 \in \mathbb{R}$ and $\mathbf{W} \in \mathbb{R}^n$ are bias and weights for features respectively. For any two given features we can independently learn the weight parameters using the model of Equation 1 with linear time complexity. However, this model is not suitable for learning the pairwise interaction of features as discussed in (Rendle 2010; 2012). A polynomial regression model with order 2 can capture the parameters for pairwise interaction, but, its time complexity is $O(n^2)$.

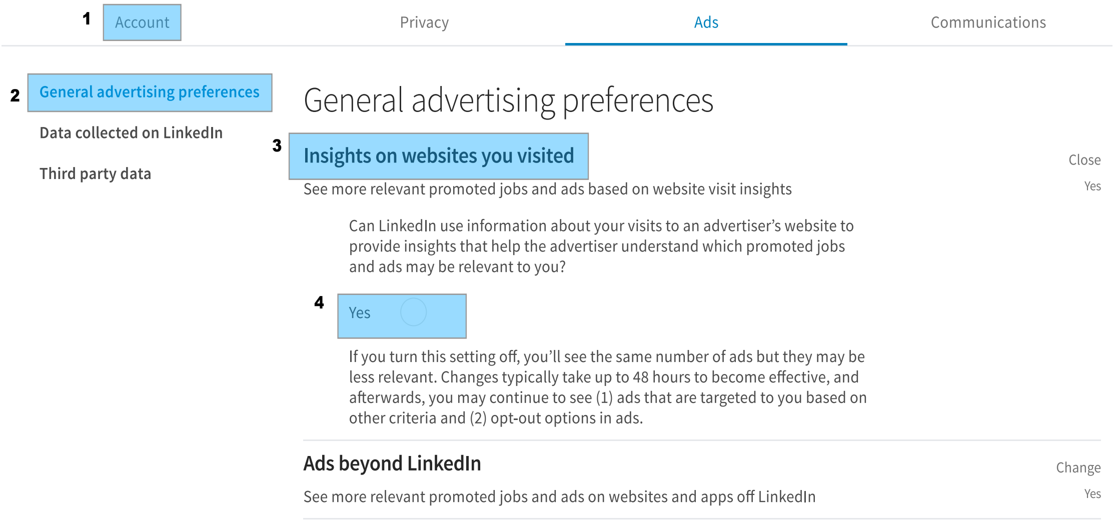
$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^{n} w_i x_i \tag{1}$$

In a *consent* recommendation system various factors interact and influence each other and that is why we have selected FM as our model. It solves the issue by factorizing the $\mathbf{W}$ as a lower dimensional factor matrix. The model equation from (Rendle 2012) is given below:

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{i'=i+1}^{n} x_i x_{i'} \sum_{j=1}^{k} v_{i,j} v_{i',j} \tag{2}$$

In Equation 2, model parameters are $w_0 \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n$ and $\mathbf{V} \in \mathbb{R}^{n \times k}$. Further, v_i and $v_{i'}$ in $\mathbf{V}$ represents the i^{th} and $(i')^{th}$ variables with k latent factors. The first part of the above equation models the linear interaction, and, second

[2] https://www.linkedin.com/help/linkedin/topics/6701/6702



General advertising preferences

Insights on websites you visited

See more relevant promoted jobs and ads based on website visit insights

Can LinkedIn use information about your visits to an advertiser's website to provide insights that help the advertiser understand which promoted jobs and ads may be relevant to you?

Yes

If you turn this setting off, you'll see the same number of ads but they may be less relevant. Changes typically take up to 48 hours to become effective, and afterwards, you may continue to see (1) ads that are targeted to you based on other criteria and (2) opt-out options in ads.

Ads beyond LinkedIn

See more relevant promoted jobs and ads on websites and apps off LinkedIn

Figure 3: LinkedIn's Privacy Settings. Example of purpose and related attribute is highlighted and numbered. 1. Purpose Category (e.g. Account), 2. Purpose Sub Category (e.g. General advertising preferences), 3. Purpose (e.g. Insights on websites you visited), 4. Setting Information comprises data field and consent value (e.g. toggle button representing 'yes')

part shows the pairwise interaction of variables with low rank(k) using their inner product. This effectively helps to estimate the parameters in highly sparse dataset. The Equation 2, is of order 2. We can have higher order variable interactions as shown below (Rendle 2010):

$$\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{l=2}^{d}\sum_{i_1=1}^{n}\cdots\sum_{i_l=i_{l-1}+1}^{n}\left(\prod_{j=1}^{l} x_{i_j}\right)\left(\sum_{f=1}^{k_l}\prod_{j=1}^{l} v_{i_j,f}^{(l)}\right) \tag{3}$$

Where, $\mathbf{V}^{(l)} \in \mathbb{R}^{n\times k_l}, k_l \in \mathbb{N}_0^+$ and, $\forall l \in \{2,\ldots,d\}$, with d as the order.

Prediction of Consent: Given a feature vector $\mathbf{x}$, Equation 3 quantifies the consent. The recommendation can be generated by thresholding the value of $\hat{y}(\mathbf{x})$. Therefore, the predicted consent $\mathcal{C}_p$ is defined as:

$$\mathcal{C}_p(\mathbf{x}) = \begin{cases} 1, & \text{allow if } \hat{y}(\mathbf{x}) \geq \theta \\ 0, & \text{deny if } \hat{y}(\mathbf{x}) < \theta \end{cases} \tag{4}$$

4 Methodology

This section describes the steps involved in our data collection procedure. We selected the participants with an active LinkedIn account with last login activity not older than 15 days. We presented a consent form prior to survey that explained to each participant about the collected data, its use in our study, and the retention period of the data. Those participants who gave consent for data collection and processing were allowed to volunteer further. The data collected from participants did not have any personally identifiable information. The study consisted of three sections: **a)** an online survey focused on understanding respondent's basic demographics, **b)** Internet User's Information Privacy Concern (IUIPC) survey(Malhotra, Kim, and Agarwal 2004), and **c)** some additional questions to support our design, so as to understand how active the participant is in social networking platforms, especially, in this case LinkedIn (refer to Section 4.2).

The participants were asked to provide us their privacy settings information from LinkedIn. We processed the settings information and related description for building binary indicator feature vectors ($\mathbf{x}^i \in \mathbb{R}^n$, refer to Section 3.1). We considered each section title as a purpose that comes under three categories (privacy, advertisement and communication) and 11 subcategories during our study. The purpose information comprised of one or more control buttons denoted as setting information (refer to Figure 3). Each type of variables such as setting, purpose and its attributes were encoded as one-hot vector.

4.1 Additional Survey Questions

Participants were asked to rate their comfort level with services using and sharing their personal information on a 5-point Likert scale: ***Q1:*** *I am comfortable with LinkedIn use/share my personal information or activity data for any purposes.* ***Q2:*** *I am comfortable with other social networks (example, Facebook, Twitter, Google+) use/share my personal information or activity data for any purposes*

To assess the change in a participant's behavior, we asked the question ***Q1*** and ***Q2*** as ***Q3*** and ***Q4*** respectively with the following updated scenario:

The enterprise explicitly says that for what purpose it is

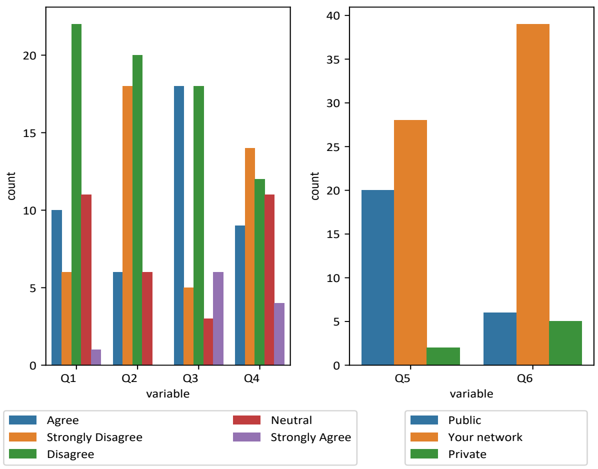


Figure 4: Survey Result

IUIPC score	Range	Mean	SD
Control	1-5	4.42	0.60
Awareness	1-5	4.65	0.54
Collection	1-5	4.29	0.68

Table 1: IUIPC Score Details

using the information and it's privacy practice is certified by a trusted organization.

'Q5' and ***'Q6'*** were formulated to understand participants opinion on visibility of their personal data on LinkedIn and other social networking sites. ***Q5:*** *If you are disclosing your personal information in LinkedIn, who can see your personal information?* ***Q6:*** *If you are disclosing your personal information in other social networks (example, Facebook, Twitter, Google+), who can see your personal information?*

4.2 Survey Result Analysis

Dataset Demographics. Sampled population from our research lab consists of data-subjects with an active LinkedIn account and an active user of at least one more social networking service. The number of participants who gave their consent for data collection experiment were 50. Out of these 50 participants 54% were Male and 46% were Female. 96% of the participants were from age group 22-30 years. The minimum educational qualification within the sample population was *under-graduate* degree, whereas, the highest qualification was *Doctor of Philosophy (PhD)*. Also, 68% of the participants were highly active (more than once in a week) on LinkedIn's social networking platform.

Findings. In the entry level survey the participants scored relatively well on IUIPC scale for control, awareness and collection of personal information as reported in Table 1. This indicates that participants have reasonably high level of privacy concerns. From the survey we found that 20% participants have modified their privacy settings only at the time of registration, 42% modify once in a quarter, 30% once in a year, and 8% never changed their setting and have given implicit consent for their data use. Figure 4 shows the results from our survey. It is apparent that the 'Agree, Disagree and Neutral' count value changes from ***'Q1'*** to ***'Q2'*** and from ***'Q3'*** to ***'Q4'***. We used this insight and included *purpose* and it's *attributes* for building our prediction model. In Figure 4, we can see that the most of the participants tend to make their personal information visible to their social network. However, some participants kept their information visible to the public in LinkedIn but not on other social networking sites. We conjecture that a participant could benefit by disclosing the professional information as it helps them building new professional connects, and hence, possibility of new job opportunities. This finding is coherent with the observation from Geffet et al (Zhitomirsky-Geffet and Bratspiess 2016). These insights suggest that the reputation of an enterprise and the potential benefits to the data-subject could influence consent decision.

5 Experiment Analysis

We surveyed 50 participants for LinkedIn with maximum of 174 privacy settings, 42 purposes, 4 purpose categories (3 values used here) and 11 purpose subcategories. Total we had 5584 samples (m) with 281 features ($n = 50 + 174 + 42 + 4 + 11$), for m and n refer to Section 3.1. If a participant gives her consent for a given data field and purpose then the state of the control is considered as '1', that is the control is selected, otherwise it will be '0'. Further, we utilized the TensorFlow implementation of FM algorithm (TFFM) with ADAM optimizer (Mikhail Trofimov 2016). Learning rate was kept as 0.001 and the threshold value (θ) was set as 0.5.

In our experiment, we randomly divided all the participants in 10 bins. We iterated over these 10 bins, using one bin for testing purpose and the remaining 9 bins for training our model. Finally, We averaged out the accuracy obtained from the 10 iterations, shown in Table 2. The sensitivity analysis of f1-score with respect to the rank is shown in Figure 5. It can be observed that there is change in accuracy with different degree of feature combination (order). Further, the size of the dataset is limited which may lead to the fluctuations in the line plot as rank increases. It would be interesting to use some contextual information such as text from purpose description to understand the meaning behind latent factors ($\mathbf{V} \in \mathbb{R}^{n \times k}$ in Equation 2). The complexity of different models is given in Table 3.

Mean Square Error, Precision and Recall: We analyzed the Mean Square Error (MSE), precision, recall and f1-score with different order and rank combinations. The results are shown in Table 2. Initially we considered all the purpose attributes in our TFFM model. Further, we assessed the impact of purpose attributes by removing each attribute one by one. From experiments we figured that rank(k) 17 gives better results in terms of accuracy. Moreover, we compared TFFM results with Linear Support Vector Machine (SVM) and polynomial SVM. Linear SVM showed marginal improvement over TFFM model as linear models work better with less amount of data. However, as explained in Section 3.1, TFFM can work as a consent recommendation system given its linear complexity, scalability with larger datasets

	Models	f1-score	precision	recall	MSE
No Rank	Linear SVM	**0.89**	0.87	0.94	-
	SVM (kernel='poly')	0.82	0.69	1.0	-
	TFFM (d=1)	0.88	0.87	0.89	0.135
TFFM	d=2	0.87	0.85	0.89	0.167
	d=3	**0.87**	**0.86**	**0.89**	**0.159**
	d=4	0.87	0.86	0.89	0.161
Order (d=3)	TFFM$_{x=A}$	0.80	0.85	0.76	0.231
	TFFM$_{x=B}$	0.84	0.84	0.84	0.274
	TFFM$_{x=A+B}$	0.72	0.85	0.64	0.313

Table 2: Evaluation in terms of f1-score, precision, recall and mean square error (MSE) for $rank = 17$ (where, $rank = k$ in Equation 2) and order d. TFFM$_x$ is the TFFM model without purpose attributes 'x'. Where 'x' can be Purpose Category (A), Purpose Sub Category (B) or both (A+B). Variants of TFFM model compared with SVM linear model and SVM with 'poly' kernel. It is observed that order d=3 performs better among other orders. Linear SVM performs slightly better than TFFM. Also, TFFM with all purpose attributes performs better than the model without purpose attributes

Model	**Order**	**Complexity**
FM	d	$O(k_d n^d)$ (straight forward)
FM	d	$O(kn)$ (reformulated)
FM	d	$O(k\bar{s}_D)$ (under sparsity)
SVM	2	$O(n^2)$

Table 3: Complexity of Models (Rendle 2010) with different cases, where k is the number of latent factors, d is the order, $\bar{s}_D$ denotes the non zero elements from the data ($\bar{s}_D$=2 for matrix factorization).

and can accommodate different contextual factors. It can be inferred from Table 2 that SVM with 'poly' kernel is overfitting with the data. Also, in his work Steffen Rendle (Rendle 2010) showed that SVM with 'poly' kernel fails with two way interactions.

Cold start vs warm start: The cold-start recommendation scenario appears when there are no prior preferences for users or items, whereas, warm-start arises when prior preferences are available.

FM model works with attributes or categories of input data represented as binary indicators (Rendle 2012). The flexibility of this model helps us to deal with cold-start users/items even when we lack prior preferences. Here, the purpose related attributes of input data are helpful for predicting the new data-subject's consent.

6 Discussion and Implication

Contributions. Our work makes some useful contributions in the context of information overload and resulting *consent fatigue* due to multiple purposes for whom consent is needed. We have shown that consent recommendation could be modeled as a prediction problem. Our recommender system has an accuracy of 87% for data-subjects with no prior preferences or usage history. For warm-start data-subjects the system is expected to perform even better. We also identified certain factors which may heavily influence a data-subject's decision making process for consent. Furthermore, the survey results showed that data-subjects are more comfortable in sharing information with enterprises providing professional services.

Future Work. Informed consent from data-subject is pivotal in data privacy regulations and safeguarding their interests. However, privacy policies are complex, and even with relevant educational qualification data-subjects find it difficult to make proper choices. Therefore, there is a need for personal digital assistant that can also help a data-subject in making consent decisions. For future work we will refer to (Liu et al. 2016; Naeini et al. 2017) as our baseline. As consent is pivotal concept in most of the regulations, therefore, we envision that it will be required even if the enterprise were to process homomorphically encrypted data (Gentry and Boneh 2009).

Implicit consent for data collection, sharing and processing is possible due to multiple reasons. Three main reasons contributing to *implicit consent* are: a) *consent fatigue*, b) data-subjects unawareness, and c) complex privacy policy document. This may lead to a sense of false compliance and security (Degeling et al. 2018). A potential area to explore is to identify possible breach of compliance regulations due to a data-subject's *implicit consent*.

In this work we built our recommender system by training our model on data gathered from LinkedIn. In post GDPR and CCPA era, all the service providers of varying type are expected to comply with them. However, more than often it is not feasible to gather sufficient data to build a model for each one of them. To address this issue *transfer learning* could be a possible area to look into. Assuming the consent requests from the other service has the same flavour of purposes and related attributes.

Apart from European Union's GDPR, many other countries are looking into their own version of data privacy laws and regulations. For example, Protection of Personal Information Act, 2013 (POPI Act) of South Africa, Personal Information Protection and Electronic Documents Act (PIPEDA) from Canada, Singapore Personal Data Protec-

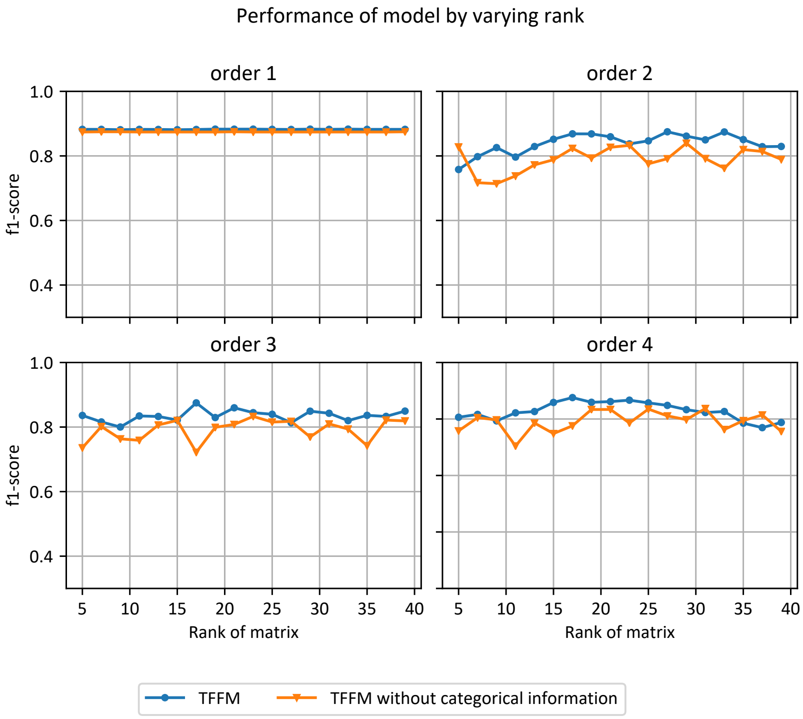


Figure 5: Performance of model by varying rank for different orders. Note that $order = 1$ is similar to linear models where there is no significance of latent factors.

tion Act, 2012, and Data Protection Act in India. In future we would like to do a user study and analyze the effect of their demographics on the decision making process.

Limitations. Our findings are based on study of privacy settings of a single web-application. This prediction model developed for LinkedIn might not be suitable for a dating site or a photograph sharing site. However, there is a possibility of exploring the application of *transfer learning* and checking the efficacy of our model on other applications.

We could collect only limited number of participant's privacy settings. In order to obtain a more reliable confidence metric, we will carry out experiments with more participants. Also, in this work we have not quantified the degree of *fatigue*. It will be interesting to see how it will affect the recommendation model. A possible way to assess it is to observe a data-subject's interaction with the application.

The information we obtained from the self reported responses of the participants may suffer from 'Privacy Paradox' (Norberg, Horne, and Horne 2007). Even though most of the participants were highly concerned about their privacy, but, their actual behavior towards consent request may change in real life. Further, we could not analyze whether the participants are going to change the privacy settings later or not.

We conclude that a lot of factors can affect a data-subjects consent depending on the purpose of processing data. However, the unavailability of factors in the real world setting challenged us in our experiments. For example, the time of consent request, benefit to a data-subject in exchange for consent, information about data field sensitivity and its retention period should matter, but it was hard to extract this information from the experimental setup.

7 Conclusion

In this work, we explored the issues pertaining to information overload and *consent fatigue* due to complex privacy policies and new regulations requiring consent for various purposes. We addressed this issue by implementing a consent recommender system for LinkedIn. Furthermore, we demonstrated that the recommendation problem could be modeled as a prediction problem. Our analysis of survey responses and LinkedIn data enabled us to identify some important factors which can influence a data-subject's decision making process. We hope that our work will be useful in identifying the issues pertaining to *consent fatigue* and build interest for further research in this area.

References

Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. 2016. Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, 265–283.

Andrade, E. B.; Kaltcheva, V.; and Weitz, B. 2002. Self-disclosure on the web: The impact of privacy policy, reward, and company reputation. *ACR North American Advances*.

Anthonysamy, P.; Greenwood, P.; and Rashid, A. 2013. Social networking privacy: Understanding the disconnect from policy to controls. *Computer* 46(6):60–67.

Balebako, R.; Jung, J.; Lu, W.; Cranor, L. F.; and Nguyen, C. 2013. "little brothers watching you": Raising awareness of data leaks on smartphones. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, 12:1–12:11. New York, NY, USA: ACM.

Barnard-Wills, D.; Chulvi, C. P.; and De Hert, P. 2016. Data protection authority perspectives on the impact of data protection reform on cooperation in the eu. *Computer Law & Security Review* 32(4):587–598.

Casteren, D. v. 2017. *Consent now and then*. Ph.D. Dissertation, Queensland University of Technology.

Cranor, L. F.; Idouchi, K.; Leon, P. G.; Sleeper, M.; and Ur, B. 2013. Are they actually any different? comparing thousands of financial institutions privacy practices. In *Proc. WEIS*, volume 13.

de la Torre, L. 2018. A guide to the california consumer privacy act of 2018. *Available at SSRN*.

Degeling, M.; Utz, C.; Lentzsch, C.; Hosseini, H.; Schaub, F.; and Holz, T. 2018. We value your privacy... now take some cookies: Measuring the gdpr's impact on web privacy. *arXiv preprint arXiv:1808.05096*.

Dwyer III, S. J.; Weaver, A. C.; and Hughes, K. K. 2004. Health insurance portability and accountability act. *Security Issues in the Digital Medical Enterprise* 72(2):9–18.

Flavián, C., and Guinalíu, M. 2006. Consumer trust, perceived security and privacy policy: three basic elements of loyalty to a web site. *Industrial Management & Data Systems* 106(5):601–620.

Gentry, C., and Boneh, D. 2009. *A fully homomorphic encryption scheme*, volume 20. Stanford University Stanford.

Ghazinour, K.; Matwin, S.; and Sokolova, M. 2016. Yourprivacyprotector, a recommender system for privacy settings in social networks. *arXiv preprint arXiv:1602.01937*.

Knijnenburg, B. P. 2014. Information disclosure profiles for segmentation and recommendation. In *SOUPS2014 Workshop on Privacy Personas and Segmentation*.

Kumar, P. 2016. Privacy policies and their lack of clear disclosure regarding the life cycle of user information. In *2016 AAAI Fall Symposium Series*.

Liu, B.; Andersen, M. S.; Schaub, F.; Almuhimedi, H.; Zhang, S. A.; Sadeh, N.; Agarwal, Y.; and Acquisti, A. 2016. Follow my recommendations: A personalized privacy assistant for mobile app permissions. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, 27–41. Denver, CO: USENIX Association.

Liu, B.; Lin, J.; and Sadeh, N. 2014. Reconciling mobile app privacy and usability on smartphones: Could user privacy profiles help? In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, 201–212. New York, NY, USA: ACM.

Madden, M. 2012. Privacy management on social media sites. *Pew Internet Report* 1–20.

Malhotra, N. K.; Kim, S. S.; and Agarwal, J. 2004. Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. *Information systems research* 15(4):336–355.

Mikhail Trofimov, A. N. 2016. tffm: Tensorflow implementation of an arbitrary order factorization machine. `https://github.com/geffy/tffm`.

Naeini, P. E.; Bhagavatula, S.; Habib, H.; Degeling, M.; Bauer, L.; Cranor, L.; and Sadeh, N. 2017. Privacy expectations and preferences in an iot world. In *Proceedings of the 13th Symposium on Usable Privacy and Security (SOUPS)*.

Norberg, P. A.; Horne, D. R.; and Horne, D. A. 2007. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs* 41(1):100–126.

Olejnik, K.; Dacosta, I.; Machado, J. S.; Huguenin, K.; Khan, M. E.; and Hubaux, J. 2017. Smarper: Context-aware and automatic runtime-permissions for mobile devices. In *2017 IEEE Symposium on Security and Privacy, SP 2017, San Jose, CA, USA, May 22-26, 2017*, 1058–1076.

Ploug, T., and Holm, S. 2013. Informed consent and routinisation. *Journal of Medical Ethics* 39(4):214–218.

Rendle, S. 2010. Factorization machines. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, 995–1000. IEEE.

Rendle, S. 2012. Factorization machines with libfm. *ACM Transactions on Intelligent Systems and Technology (TIST)* 3(3):57.

Sadeh, J. L. B. L. N., and Hong, J. I. 2014. Modeling users mobile app privacy preferences: Restoring usability in a sea of permission settings. In *Symposium on Usable Privacy and Security (SOUPS)*. Citeseer.

Sadeh, N.; Hong, J.; Cranor, L.; Fette, I.; Kelley, P.; Prabaker, M.; and Rao, J. 2009. Understanding and capturing peoples privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing* 13(6):401–412.

Solove, D. J. 2012. Introduction: Privacy self-management and the consent dilemma. *Harv. L. Rev.* 126:1880.

Voigt, P., and Von dem Bussche, A. 2017. *The EU General Data Protection Regulation (GDPR)*, volume 18. Springer.

Wijesekera, P.; Baokar, A.; Tsai, L.; Reardon, J.; Egelman, S.; Wagner, D.; and Beznosov, K. 2017. The feasibility of dynamically granted permissions: Aligning mobile privacy with user preferences. In *Security and Privacy (SP), 2017 IEEE Symposium on*, 1077–1093. IEEE.

Zhitomirsky-Geffet, M., and Bratspiess, Y. 2016. Professional information disclosure on social networks: The case of facebook and linked in in israel. *Journal of the Association for Information Science and Technology* 67(3):493–504.