Modeling Language Vagueness in Privacy Policies Using Deep Neural Networks

[1]Fei Liu, [2]Nicole Lee Fella, [1]Kexin Liao
[1]University of Central Florida, 4000 Central Florida Blvd., Orlando, Florida 32816
[2]Manhattan College, 4513 Manhattan College Parkway, Riverdale, NY 10471
feiliu@cs.ucf.edu, nfella01@manhattan.edu, ericaryo@knights.ucf.edu

Abstract

Website privacy policies are too long to read and difficult to understand. The over-sophisticated language undermines the effectiveness of privacy notices. People become less willing to share their personal information when they perceive the privacy policy as vague. The goal of this paper is to decode vagueness from a natural language processing perspective. While thoroughly identifying the vague terms and their linguistic scope remains an elusive challenge, in this work we seek to learn vector representations of words in privacy policies using deep neural networks. The vector representations are fed to an interactive visualization tool (LSTMVis) to test on their ability to discover syntactically and semantically related terms. The approach holds promise for modeling and understanding language vagueness.

Introduction

Website privacy policies represent a legally binding agreement between users and website operators. They are verbose, too long to read, and difficult to understand. Albeit the paramount importance, people tend to ignore these privacy notices unless a serious concern is raised, e.g., by the media. Research studies have explored various means to improve the effectiveness of privacy notice and choice. Cranor et al. (2002; 2006) introduce a standard machine-readable format for website privacy policies using the Platform for Privacy Preferences (P3P). The Usable Privacy Policy Project[1] aims to extract key privacy policy features for presentation to end-users in a structured and easily understandable format (Sadeh et al. 2013). These approaches allow the users to quickly navigate to the text passages that are related to certain key privacy practices. It alleviates the "too long to read" challenge brought by document length. On the other hand, studies that tackle the "difficult to understand" challenge have been largely absent from this space, partly because of the complexity and richness of natural language.

One might wonder why privacy notices need to adopt such a sophisticated language in the first place. Bhatia et al. (2016) suggest two causes in their recent work. First, the privacy policies need to be *comprehensive*, covering all possible cases such as the physical places (e.g., stores, offices) and web/mobile platforms. Second, the policy statements must be *accurate*, which means they are true to all data practices and systems. Clearly it will be difficult for the legal counsel to anticipate all the future needs, naturally they resort to generalization and sophistication to frame the statements, introducing vagueness to the text. An example statement is: "*The email address is used for sending account notifications and other system-related information as needed.*"

Condition (9): depending, necessary, appropriate, inappropriate, as needed, as applicable, otherwise reasonably, sometimes, from time to time
Generalization (12): generally, mostly, widely, general, commonly, usually, normally, typically, largely, often, primarily, among other things
Modality (8): may, might, can, could, would, likely, possible, possibly
Numeric Quantifier (11): anyone, certain, everyone, numerous, some, most, few, much, many, various, including but not limited to

Table 1: Table adopted from (Reidenberg et al. 2016). It includes a total of 40 vague terms that are manually identified by experts from 15 privacy policies. The terms are divided into four categories.

Vagueness is a linguistic phenomenon that is not yet fully studied in the natural language processing (NLP) community. A concept is considered vague if it lacks clarity or corresponds to borderline cases (e.g., tall, short). Even terms like "disability" raise questions such as "how much loss of vision is required before one is legally blind?"[2] Farkas et al. (2010) introduce a shared task on detecting uncertainty cues (i.e., hedges and weasels) from biological articles and Wikipedia pages. Reidenberg et al. (2016) manually analyze a set of 15 privacy policies and identify 40 vague terms (Table 1) which we also adopt in this study. Note that there appears to be a dilemma: if a collection of vague terms can be prespecified, classifying a piece of text as vague or not seems trivial; on the other hand, given the richness of nat-

[1] https://www.usableprivacy.org

[2] https://en.wikipedia.org/wiki/Vagueness

ural language, creating such a comprehensive list of vague terms can be highly challenging, if possible at all.

The main contribution of this work lies in learning vector representation of words in privacy policies using deep neural networks. There is one vector representation for each word token in the privacy policies. The vector representations are iteratively learned so that they would perform well in two tasks: predicting the next word given its context (e.g., "*we do not request any* ____" → "information") and predicting whether or not a word is in a list of prespecified vague terms (e.g., "*may*" → "Vague (V)", "*email*" → "Not Vague (N)". The 40 vague terms in Table 1 are used in this study. We hypothesize that certain dimensions of the vector representation encode the semantic meaning of the word, including vagueness. These vector representations are further fed to an interactive visualization tool (LSTMVis, Strobelt et al., 2016) to test on their ability to discover syntactically and semantically related terms. The approach holds promise for modeling vagueness of words within context. The visualization tool allows the privacy researchers to perform knowledge discovery on the website privacy policies.[3]

Related Work

We discuss related work along three dimensions: law, privacy policy, and natural language.

In the American Constitution, there is a "void for vagueness" doctrine. It states that the law should be clearly specified so that the average citizen would understand. If a rule is vague, it is unenforceable. Researchers in the law community have thus exploited the vagueness of legal language and interpretation of boundary decisions. In his seminal work, Waldron (1994) distinguishes ambiguity, contestability, vagueness, and introduces a general term "indeterminacy" to cover the three cases. Post (1994) argues that the legal rules cannot be simply rewritten to be more precise, since they are not in isolation but reflect the forms of social order. Jonsson (2009) suggests that vagueness in law does not call for specific interpretation of the law itself, but only for an application of the law on case-by-case basis. Studies in (Hernacki 2012) suggest that the decades-old antihacking statue Computer Fraud and Abuse Act (CFAA) is in need of a face-lift. Phrases such as "involve" and "other similar information" are not providing enough clarity. Liebwald (2013) concerns that the vagueness in combination with the elasticity of legal interpretation may affect the binding force of law. The paper introduces a theory called Hyperbola of Meaning. Raffman (2015) provides a characterization of linguistic vagueness. Vague words possess unclear boundaries, but are distinguished from ambiguity, underspecificity, and several forms of indeterminacy. Low et al. (2015) illustrate the application of the vagueness doctrine to four Supreme Court vagueness cases. They point out that when determining vagueness of statutes, it is important to take the intersection between state and federal law into account. Hunt (2015) studies "epistemicism", which states that vague statements are either true or false even though it is impossible to know which. The author suggests that vagueness should be explained within the theory of legal interpretation.

Vagueness has been studied within the scope of website privacy policies. In particular, Reidenberg et al. (2016) and Bhatia et al. (2016) introduce a theory of vagueness for privacy policy statements. The theory indicates how vague modifiers can be composed to increase or decrease the overall vagueness. They show that the increase in vagueness often decreases users' willingness to share personal information. Our work is different from these studies in that we do not attempt to generate a vagueness score for a given piece of text. Instead, we seek to exploit deep neural networks to learn word representations that encode semantic meanings and vagueness.

There are other studies that focus on improving the effectiveness of privacy policies. Vail et al. (2008) compare various ways to present privacy policy information to online users. Their findings suggest that users perceive paragraph-form policies to be more secure than others, however the user comprehension of such paragraph-form policies is poor. Kelley et al. (2010) develop a nutrition label approach for representing the key practices of privacy policies. They show that a standardized table format is effective in assisting users with their information needs. Micheti et al. (2010) aims to identify guidelines for privacy policies that children and teen can understand and accurately interpret. Phrases which cause misunderstanding and vagueness include "may," "except," and "aside from." Many users, especially young people, are aware that privacy policies are vague due to strategic reasons of service providers. The empirical study in (Lammel and Pek 2013) discusses Platform for Privacy Preferences (P3P). As a platform, P3P is used and interpreted by users to help automate decision making. While being one of the only widely used languages for privacy policies, P3P still has downfalls. More rigorous specifications in language and enforcement of correct use are necessary. In more recent studies, Reidenberg et al. (2015a; 2015b) study the effectiveness of privacy notice and choice framework and suggest that people do not agree with each other when interpreting privacy polices. Wilson et al. (2016b; 2016a) explore crowdsourcing for annotating privacy policies. They introduce a corpus of 115 privacy policies with manual annotations of fine-grained data practices.

Comprehensive studies have been missing for understanding vagueness in the natural language processing community. Farkas et al. (2010) focus on the detection of uncertainty cues and their linguistic scope in natural language texts. The motivation behind this task is to distinguish factual and uncertain information in text, which is of essential importance to information extraction. Much of the techniques involve sentence-level classifications using SVM, CRF, and maximum entropy. However, it remains to be seen if a word- or sentence-level classification formulation is well suited for this task. In (Alexopoulos and Pavlopoulos 2014), vagueness is considered to be a linguistic phenomenon. It arises with a lack of clear boundaries and conditions. These boundaries usually do not allow concrete distinction. Classifying text as vague or not vague can be subjective, making it important to look at agreement between interpretations and

[3] The code and data model are available at http://www.nlp.cs.ucf.edu

annotations. Using a naive Bayes classifier, the study shows that vague and not vague senses can be separated.

Data

Our dataset consists of 1,010 website privacy policies. These privacy policies are gathered using Amazon Mechanical Turk (`mturk.com`) from the most frequently visited websites across 15 categories, ranging from Arts, Business, Computers to Science, Shopping, and Sports. The privacy policy documents have been converted to XML format and are available for download publicly.[4] Liu et al. (2014) and Ramanath et al. (2014) perform studies using this dataset to align policy segments based on the issues they discuss. For example, text segments that discuss the usage of cookies (i.e., small data files transferred by the website to the user's computer) should be grouped together. They experiment with unsupervised hidden Markov models and demonstrate that the approaches are more effective than clustering.

In this study, we seek to learn a vector representation for each word token (i.e., occurrence of the word) in the dataset. Each privacy policy document is split into a set of sentences; each sentence is further split into a set of word tokens. All word tokens are lowercased. We remove sentences that contain 3 word tokens or less, since they are too short and often noisy (e.g., "Back to Top"). A special token $\langle /\mathsf{s} \rangle$ is used as the end-of-sentence symbol. A word token is deemed *vague* if it is included in a prespecified list of vague terms (see Table 1). Note that we consider word tokens such as "*among other things*" as vague, but an individual word (e.g., "*among*") is not vague when placed in other context (e.g., "*among consumers*").

Statistics of the dataset are illustrated in Table 2.

total # of web privacy policies	1,010
total # of sentences	107,076
total # of word tokens	2,534,094
total # and % of vague tokens	59,026 (2.3%)
total # and % of sentences that contain at least one vague token	41,033 (38.3%)

Table 2: Statistics of the dataset.

Modeling Language and Vagueness

So far we have demonstrated the needs for understanding language vagueness and described our dataset, we proceed by introducing a deep neural network for learning vector representation of words in privacy policies (see Figure 1 for illustration). Traditional approaches to building feature representation have been largely based on manual feature extraction (Farkas et al. 2010). The idea behind the deep neural network is that it learns to automatically construct a feature representation for each word, in the form of a dense continuous vector ($\boldsymbol{h} \in \mathbb{R}^d$). The feature representation is optimized so that it could perform well in two tasks: 1) predicting the next word given previous words in the sentence, and 2) predicting if the current word is vague or not given the context. This corresponds to a multi-task learning setting.

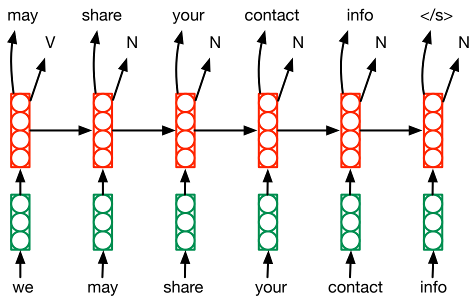


Figure 1: Modeling language and vagueness in website privacy policies using a recurrent neural network with two objectives. "Vague" → V, "Not Vague" → N.

Deep neural networks have seen considerable success in a range of natural language processing tasks. Our work is inspired by recent advances on learning word embeddings (Mikolov et al. 2013; Tang et al. 2014) and sequence-to-sequence models (Sutskever, Vinyals, and Le 2014; Cheng, Fang, and Ostendorf 2015; Luong et al. 2016).

Concretely, let $\boldsymbol{x} = \{x_1, x_2, \cdots, x_{\mathsf{N}}\}$ be an input sentence consisting of N word tokens. The word tokens come from a vocabulary $\mathcal{V}$ of size $|\mathcal{V}| = \mathsf{V}$. Each word is replaced by a pre-trained word embedding ($\boldsymbol{x}_i \in \mathbb{R}^{\mathsf{D}}$) before it is fed to the neural network. With a slight abuse of notation, we use x_i to represent the word token and $\boldsymbol{x}_i$ (boldface) to represent its embedding. We use the 300-dimension (D=300) word2vec embeddings pre-trained on Google News dataset with about 100 billion words[5]. Since the privacy policy language is considered to be domain-specific, we employ a vocabulary of 5,000 words in this study (V=5,000). They correspond to the most frequent words in the 1,010 privacy policies dataset. Among them, 602 words cannot find pre-trained word2vec embeddings, we thus randomly initiate the embeddings using a standard normal distribution.

Next we feed the sentence one word at a time to a recurrent neural network (see Figure 1). A recurrent neural network (RNN) corresponds to a language model, where the goal is to predict the next word given its previous words. The probability of the entire sequence $p(\boldsymbol{x})$ is represented in Eq.(1), whereas the individual probability $p(x_t|x_1, \cdots, x_{t-1})$ is calculated by RNN.

$$p(\boldsymbol{x}) = \prod_{t=1}^{\mathsf{N}} p(x_t|x_1, \cdots, x_{t-1}) \quad (1)$$

A recurrent neural network operates on a sequence of words and creates a hidden state representation $\boldsymbol{h}_t \in \mathbb{R}^d$ for the word at time step t. It learns a function of the form $\boldsymbol{h}_t = f(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t)$, where $\boldsymbol{h}_{t-1}$ is the hidden state representation of the previous time step and $\boldsymbol{x}_t$ is the input word

[4] `https://usableprivacy.org/data`

[5] https://code.google.com/archive/p/word2vec/. An alternative approach would be to learn domain-specific word embeddings using the privacy policy documents.

embedding of the current time step. Both the Long Short-Term Memory (LSTM) networks and Gated Recurrent Unit (GRU) networks are variants of the recurrent neural networks. They correspond to different gating mechanisms, hence different $f(\cdot)$. This work specifically focuses on using GRU to produce the hidden state representations, where $\boldsymbol{h}_t = \text{GRU}(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t)$. GRUs have seen considerable success in recent NLP applications (Luong et al. 2016). It uses two neural gates to control the flow of information, where $\boldsymbol{i}_t \in \mathbb{R}^d$ and $\boldsymbol{r}_t \in \mathbb{R}^d$ respectively represent the *input* and *reset* gate. $\boldsymbol{c}_t \in \mathbb{R}^d$ is sometimes referred to as the *cell* value and $\boldsymbol{h}_t \in \mathbb{R}^d$ is the *hidden* state representation we are interested in.

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}^i \boldsymbol{x}_t + \boldsymbol{U}^i \boldsymbol{h}_{t-1} + \boldsymbol{b}^i) \quad (2)$$

$$\boldsymbol{r}_t = \sigma(\boldsymbol{W}^r \boldsymbol{x}_t + \boldsymbol{U}^r \boldsymbol{h}_{t-1} + \boldsymbol{b}^r) \quad (3)$$

$$\boldsymbol{c}_t = \tanh(\boldsymbol{W}^c \boldsymbol{x}_t + \boldsymbol{U}^c \boldsymbol{h}_{t-1} + \boldsymbol{b}^c) \quad (4)$$

$$\boldsymbol{h}_t = \boldsymbol{i}_t \odot \boldsymbol{c}_t + (1 - \boldsymbol{i}_t) \odot \boldsymbol{h}_{t-1} \quad (5)$$

In the above equations, $\boldsymbol{W}^i, \boldsymbol{W}^r, \boldsymbol{W}^c$ and $\boldsymbol{U}^i, \boldsymbol{U}^r, \boldsymbol{U}^c$ are parameters, $\boldsymbol{b}^i, \boldsymbol{b}^r, \boldsymbol{b}^c$ are biases; $\odot$ corresponds to the element-wise product of two vectors; $\sigma(\cdot)$ is the sigmoid function; $\tanh(\cdot)$ is the hyperbolic tangent function. They are applied element-wise to the vectors.

The hidden state $\boldsymbol{h}_t$ is expected to carry over semantic information from the beginning of the sentence to the current time step; Using the vector representation $\boldsymbol{h}_t$, we learn to complete two tasks: first, $\boldsymbol{h}_t$ is used to predict the next word using a softmax activation function (Eq.(9)), where $p(y_t = j|\boldsymbol{h}_t)$ is the probability that the next word y_t is predicted as the j-th word in the vocabulary; second, $\boldsymbol{h}_t$ is employed to predict if the current word is vague or not, where $p(c_t = k|\boldsymbol{h}_t)$ is the probability of the current word being vague ($k = 1$) or not ($k = 2$).

$$p(y_t = j|\boldsymbol{h}_t) = \frac{\exp(\boldsymbol{w}_j \boldsymbol{h}_t)}{\sum_{j'=1}^{\mathsf{V}} \exp(\boldsymbol{w}_{j'} \boldsymbol{h}_t)} \quad (6)$$

$$p(c_t = k|\boldsymbol{h}_t) = \frac{\exp(\boldsymbol{w}_k \boldsymbol{h}_t)}{\sum_{k'=1}^{\mathsf{C}} \exp(\boldsymbol{w}_{k'} \boldsymbol{h}_t)} \quad (7)$$

We use θ to represent all the trainable parameters in the aforementioned deep neural network. The above model can be trained in an end-to-end fashion using stochastic gradient descent. In particular RMSProp (Tieleman and Hinton 2012) is used for parameter estimation, which has been shown to perform well in sequence learning tasks. During training, the model parameters are iteratively updated so as to minimize the negative log likelihood of the training data $\mathcal{L}(\theta)$.

$$\begin{aligned}\mathcal{L}(\theta) = & -\alpha \sum_{i=1}^{\mathsf{S}} \sum_{t=1}^{\mathsf{N}} \sum_{j=1}^{\mathsf{V}} \log p(y_t = j|\boldsymbol{h}_t; \theta) \\ & -\beta \sum_{i=1}^{\mathsf{S}} \sum_{t=1}^{\mathsf{N}} \sum_{k=1}^{\mathsf{C}} \log p(c_t = k|\boldsymbol{h}_t; \theta) \quad (8)\end{aligned}$$

where S=107,076 is the total number of sentences in our dataset, N=50 is set to be the maximum number of words per sentence, V=5,000 is the vocabulary size, C=2 is the number of categories (i.e., vague or not). α and β are scalar coefficients used to indicate the weights of the components in the leanring objective. They are empirically set to α=1 and β=1 in our study. This means that the system is subject to equal penalty when it mispredicts the next word or vagueness of the current word. We set the dimensionality $d = 512$. The deep neural network finally produces a 512-dimension vector representation for each word in the dataset. The model is trained for 30 epochs. The accuracy of predicting the identity of the next word ("Accuracy-Word") and accuracy of predicting the word vagueness ("Accuracy-Vagueness") are plotted in Figure 2. The "Accuracy-Vagueness" curve saturates after the first couple of epochs, suggesting word-level binary prediction is not a difficult task, whereas the "Accuracy-Word" curve increases steadily across all the training epochs. The "Accuracy-Word" yields a much lower accuracy since language modeling—predicting the next word (1 out of V selection) given the history—remains a genuinely challenging task.

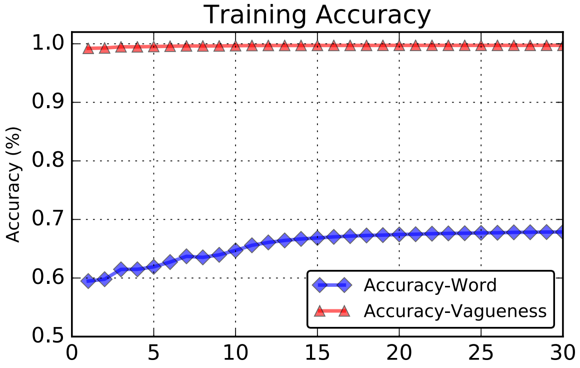
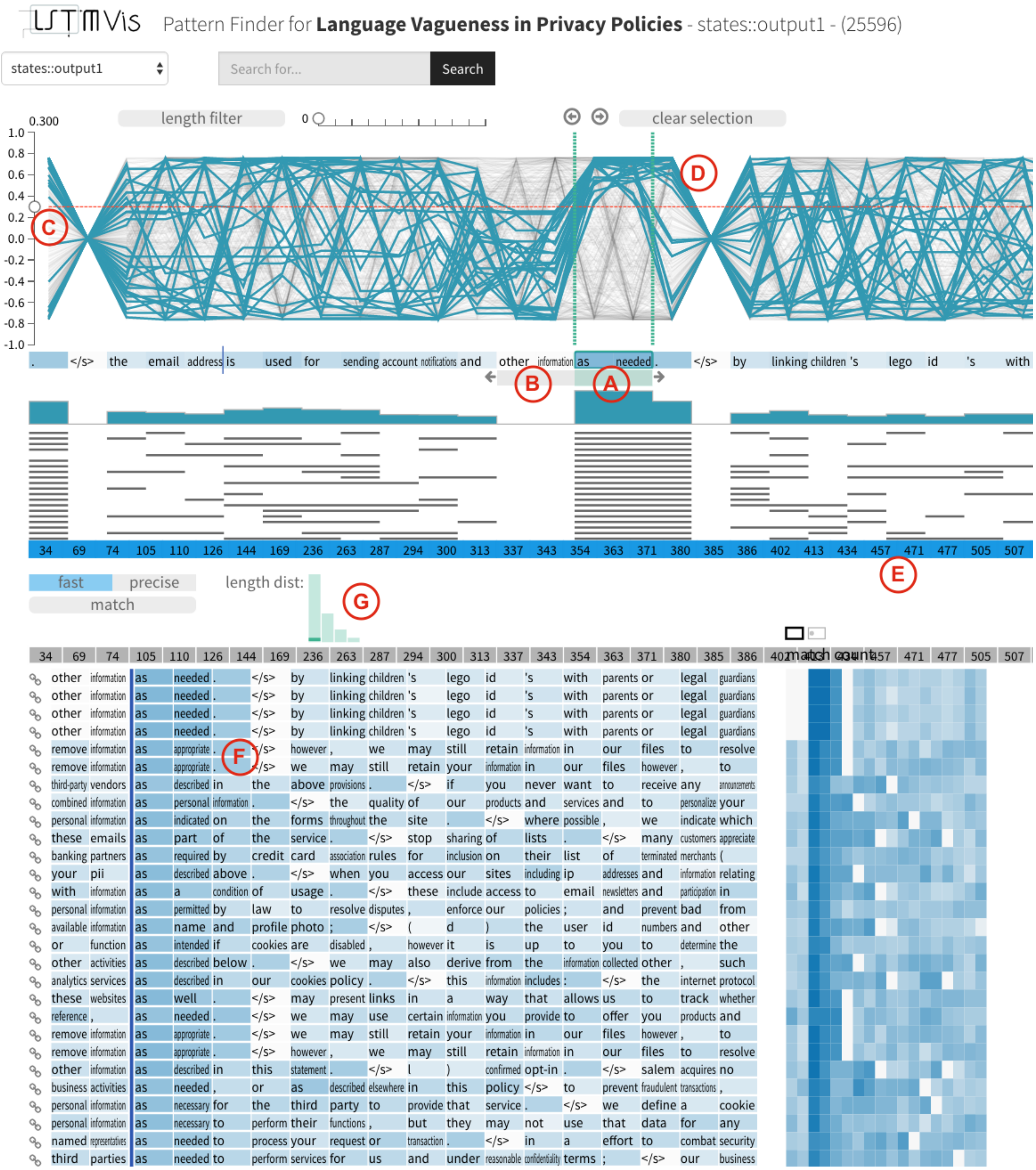


Figure 2: Training accuracy across 30 epochs.

Visualization

The neural network presented in the previous section creates a 512-dimension vector representation for each word in the privacy policy dataset. The vectors are colored in red in Figure 1. These vector representations resemble the feature vectors we normally obtain through a linear model (e.g., SVM or maximum entropy) or dimensionality reduction approach (e.g., SVD). They could be used in downstream classification tasks such as predicting if a piece of text is vague or not. However, because of the lack of such large-scale datasets and consistent annotation guidelines for vagueness prediction of privacy policies, we choose not to perform empirical evaluation on the datasets. Instead, we seek to interpret the learned vector representations and explore what information is encoded in the 512-dimension vectors.

Researchers strive to understand the neural models in natural language processing. Very recently, Li et al. (2016) develop strategies to understand the model compositionality. That is, how sentence meanings are built from the meanings of words and phrases. The approach measures the "salience" of each dimension based on how much it contributes to the final decision, which is approximated using first-order



```
. </s> the email address is used for sending account notifications and other information as needed . </s> by linking children 's lego id 's with
```

```
other information as needed . </s> by linking children 's lego id 's with parents or legal guardians
other information as needed . </s> by linking children 's lego id 's with parents or legal guardians
other information as needed . </s> by linking children 's lego id 's with parents or legal guardians
other information as needed . </s> by linking children 's lego id 's with parents or legal guardians
remove information as appropriate . </s> however , we may still retain information in our files to resolve
remove information as appropriate . </s> we may still retain your information in our files however , to
third-party vendors as described in the above provisions . </s> if you never want to receive any announcements
combined information as personal information . </s> the quality of our products and services and to personalize your
personal information as indicated on the forms throughout the site . </s> where possible , we indicate which
these emails as part of the service . </s> stop sharing of lists . </s> many customers appreciate
banking partners as required by credit card association rules for inclusion on their list of terminated merchants (
your pii as described above . </s> when you access our sites including ip addresses and information relating
with information as a condition of usage . </s> these include access to email newsletters and participation in
personal information as permitted by law to resolve disputes , enforce our policies ; and prevent bad from
available information as name and profile photo ; </s> ( d ) the user id numbers and other
or function as intended if cookies are disabled , however it is up to you to determine the
other activities as described below . </s> we may also derive from the information collected other , such
analytics services as described in our cookies policy . </s> this information includes : </s> the internet protocol
these websites as well . </s> may present links in a way that allows us to track whether
reference , as needed . </s> we may use certain information you provide to offer you products and
remove information as appropriate . </s> we may still retain your information in our files however , to
remove information as appropriate . </s> however , we may still retain information in our files to resolve
other information as described in this statement . </s> ( ) confirmed opt-in . </s> salem acquires no
business activities as needed , or as described elsewhere in this policy </s> to prevent fraudulent transactions ,
personal information as necessary for the third party to provide that service . </s> we define a cookie
personal information as necessary to perform their functions , but they may not use that data for any
named representatives as needed to process your request or transaction . </s> in a effort to combat security
third parties as needed to perform services for us and under reasonable confidentiality terms ; </s> our business
```

Figure 3: Visualization of the vector representations using LSTMVis.

under the same circumstances	**necessary or appropriate to**	**personally-identifying information**
under the following conditions	necessary to	personal information
under the following circumstances	required to	access information
under the circumstances	otherwise permitted by	financial information
in any case	your right to	aggregate information

Table 3: Example similar phrases identified by the visual tool. The given phrases are shown in bold. Note that the similar phrases are hand-picked. Not all system identified phrases are closely related to the given phrases.

derivatives. Strobelt et al. (2016) present a visual analysis tool named LSTMVis[6]. The tool explores the hidden state dynamics of a recurrent neural network. It allows the user to select an input phrase and find similar phrases in the dataset that demonstrate similar hidden state patterns. We adopt LSTMVis in our study and import the vector representations produced in the previous section. The visualization is presented in Figure 3.

The interface consists of two views: the *select* view corresponds to the upper panel ((A) to (D)) and the *match* view corresponds to the lower panel ((E) to (G)). All sentences in the dataset are concatenated into a meta word sequence and delimited by the special symbol $\langle/\mathsf{s}\rangle$. Each word is represented using a fixed-width box; if words do not fit into the box, they are squeezed. Users are provided with buttons to move forwards or backwards with the word sequence, as well as a search box (disabled for now) to directly jump to certain text region. Each vector dimension corresponds to a line in the *select* view. Because our vector representation contains 512 dimensions, there are 512 lines in the figure, numbered from 0 to 511.

The user starts by selecting a phrase in the word sequence (e.g., "as needed," see (A)). This action turns on a set of vector dimensions (represented as S_1), where "turn on" means the cell value of the dimension, in both of the selected word positions, is greater than a threshold (default to 0.3, see (C)). The gray slider (see (B)) further allows the user to select a few context words (e.g., "other information") that surround the current selected phrase. Similarly, this action turns on a second set of vector dimensions (represented as S_2). Note that our goal is to identify the dimensions that uniquely characterize the selected phrase ("as needed") but not the surrounding words. As a result, the intersection of the two sets of dimensions $|S_1 \cap S_2|$ are the ones we wish to focus on. These dimensions are listed in the interface (see (E)).

In the *match* view, the visual tool continues to search for text regions where the same set of vector dimensions ($|S_1 \cap S_2|$) have been turned on. The text regions are further ranked by the inverse of number of additional "on" cells $|S_1 \cup S_2|$ and the length of the text region. The top phrases are listed on the interface (see (F)) with length distribution plotted (see (G)). The color intensity is used to signal the value of $|S_1 \cap S_2|$. For the selected phrase ("as needed"), we observe that several syntactically and semantically similar phrases have been selected, including "as appropriate," "as indicated on", "as required by", "as described above/below/in," and "as necessary to/for." Several similar examples are presented in Table 3. These findings suggest that even in the relatively restricted domain of website privacy policies, a large number of text variations exist. They use different text expressions to represent the same or similar meanings. It is thus left to be seen if creating a comprehensive list of vague terms is feasible given the richness and complexity of natural language.

Discussion

This study explores a recurrent neural network (RNN) model with two objectives to jointly perform language modeling and vagueness identification. We experiment with various network structures and settle with the current model given its simple structure and proven success (Cheng, Fang, and Ostendorf 2015). The deepness of the model comes from the backpropagation through time. Other variants of the neural network may serve the purpose. The lack of annotated datasets remains a hurdle towards understanding language vagueness using a data-driven approach. This work primarily focuses on knowledge discovery using an interactive visualization tool. Future work will consider a more integral approach to evaluating the language vagueness such as with case-based reasoning (Rissland, Ashley, and Branting 2006). Judging a legal phrase as vague or not is challenging in that it requires more than identifying the vagueness of individual words. In the future we wish to assist the legal counsels to clarify the privacy text, as well as raise public awareness of the vague terms as presented in the website privacy policies.

Conclusion

In this work we attempt to computationally model the vagueness of privacy policies using deep neural networks. The neural networks learn to generate vector representations for words in the privacy policies. We explore visualization of the learned vector representations, identify dimensions that could capture language specific characteristics, and present example phrases that potentially signal vagueness. Our learned model and visualization allow researchers to explore the vagueness of natural language and perform knowledge discovery.

Acknowledgements

We are grateful to the three reviewers for their helpful suggestions. Nicole Lee Fella was supported by the NSF REU program (1560302) during the completion of this work.

[6] `lstm.seas.harvard.edu`

References

Alexopoulos, P., and Pavlopoulos, J. 2014. A vague sense classifier for detecting vague definitions in ontologies. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*.

Bhatia, J.; Breaux, T. D.; Reidenberg, J. R.; and Norton, T. B. 2016. A theory of vagueness and privacy risk perception. In *Proceedings of the IEEE International Conference on Requirements Engineering (RE)*.

Cheng, H.; Fang, H.; and Ostendorf, M. 2015. Open-domain name error detection using a multi-task RNN. In *Proceedings of the Conference Empirical Methods in Natural Language Processing (EMNLP)*.

Cranor, L. F.; Guduru, P.; and Arjula, M. 2006. User interfaces for privacy agents. *ACM Transactions on Computer-Human Interaction (TOCHI)* 13(2):135–178.

Cranor, L. F. 2002. *Web Privacy with P3P*. O'Reilly & Associates.

Farkas, R.; Vincze, V.; Mora, G.; Csirik, J.; and Szarvas, G. 2010. The CoNLL-2010 shared task: Learning to detect hedges and their scope in natural language text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL)*.

Hernacki, A. 2012. A vague law in a smartphone world: Limiting the scope of unauthorized access under the computer fraud and abuse act. *American University Law Review* 61(5):1543–1584.

Hunt, L. W. 2015. What the epistemic account of vagueness means for legal interpretation. *Law and Philosophy* 35(1):29–54.

Jonsson, O. P. 2009. Vagueness, interpretation, and the law. *Legal Theory* 15:193–214.

Kelley, P. G.; Cesca, L.; Bresee, J.; and Cranor, L. F. 2010. Standardizing privacy notices: An online study of the nutrition label approach. In *Proceedings of CHI*.

Lammel, R., and Pek, E. 2013. Understanding privacy policies (a study in empirical language usage analysis). *Empirical Software Engineering* 18:310–374.

Li, J.; Chen, X.; Hovy, E.; and Jurafsky, D. 2016. Visualizing and understanding neural models in NLP. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.

Liebwald, D. 2013. Law's capacity for vagueness. *International Journal for the Semiotics of Law* 26(2):391–423.

Liu, F.; Ramanath, R.; Sadeh, N.; and Smith, N. A. 2014. A step towards usable privacy policy: Automatic alignment of privacy statements. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*.

Low, P. W., and Johnson, J. S. 2015. Changing the vocabulary of the vagueness doctrine. *Virginia Law Review* 101(8):2051–2116.

Luong, M.-T.; Sutskever, I.; Le, Q. V.; Vinyals, O.; and Kaiser, L. 2016. Multi-task sequence to sequence learning. In *Proceedings of International Conference on Learning Representations (ICLR)*.

Micheti, A.; Burkell, J.; and Steeves, V. 2010. Fixing broken doors: Strategies for drafting privacy policies young people can understand. *Bulletin of Science Technology Society* 30(2):130–143.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*.

Post, R. C. 1994. Reconceptualizing vagueness: Legal rules and social orders. *California Law Review* 82(3):491–507.

Raffman, D. 2015. Precis of unruly words: A study of vague language. *Philosophy and Phenomenological Research* 90(2):452–456.

Ramanath, R.; Liu, F.; Sadeh, N.; and Smith, N. A. 2014. Unsupervised alignment of privacy policies using hidden Markov models. In *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Reidenberg, J. R.; Breaux, T.; Cranor, L. F.; French, B.; Grannis, A.; Graves, J. T.; Liu, F.; McDonald, A. M.; Norton, T. B.; Ramanath, R.; Russell, N. C.; Sadeh, N.; and Schaub, F. 2015a. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Law Technology Journal* 30(1).

Reidenberg, J. R.; Russell, N. C.; Callen, A.; Qasir, S.; and Norton, T. B. 2015b. Privacy harms and the effectiveness of the notice and choice framework. *I/S Journal of Law Policy for the Information Society* 11:485–524.

Reidenberg, J. R.; Bhatia, J.; Breaux, T. D.; and Norton, T. B. 2016. Ambiguity in privacy policies and the impact of regulation. *Journal of Legal Studies* 45(2).

Rissland, E. L.; Ashley, K. D.; and Branting, L. K. 2006. Case-based reasoning and law. *The Knowledge Engineering Review* 20(3):293–298.

Sadeh, N.; Acquisti, A.; Breaux, T.; Cranor, L.; McDonald, A.; Reidenberg, J.; Smith, N.; Liu, F.; Russel, C.; Schaub, F.; and Wilson, S. 2013. The usable privacy policy project: Combining crowdsourcing, machine learning and natural language processing to semi-automatically answer those privacy questions users care about. Technical Report CMU-ISR-13-119, Carnegie Mellon University.

Strobelt, H.; Gehrmann, S.; Huber, B.; Pfister, H.; and Rush, A. M. 2016. Visual analysis of hidden state dynamics in recurrent neural networks. In *arXiv:1606.07461*.

Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to sequence learning with neural networks. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*.

Tang, D.; Wei, F.; Yang, N.; Zhou, M.; Liu, T.; and Qin, B. 2014. Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*.

Tieleman, T., and Hinton, G. 2012. Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning.

Vail, M. W.; Earp, J. B.; and Anton, A. I. 2008. An empirical study of consumer perceptions and comprehension of web site privacy policies. *IEEE Transactions on Engineering Management* 55(3):442–454.

Waldron, J. 1994. Vagueness in law and language: Some philosophical issues. *California Law Review* 82(3):509–540.

Wilson, S.; Schaub, F.; Dara, A. A.; Liu, F.; Cherivirala, S.; Leon, P. G.; Andersen, M. S.; Zimmeck, S.; Sathyendra, K. M.; Russell, N. C.; Norton, T. B.; Hovy, E.; Reidenberg, J. R.; and Sadeh, N. 2016a. The creation and analysis of a website privacy policy corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Wilson, S.; Schaub, F.; Ramanath, R.; Sadeh, N.; Liu, F.; Smith, N. A.; and Liu, F. 2016b. Crowdsourcing annotations for websites' privacy policies: Can it really work? In *Proceedings of the 25th International World Wide Web Conference (WWW)*.