The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Neural Ideal Point Estimation Network

Kyungwoo Song, Wonsung Lee, Il-Chul Moon
Korea Advanced Institute of Science and Technology
291 Daehak-ro, Yuseong-gu
Daejeon 34141, South Korea
{gtshs2,aporia,icmoon}@kaist.ac.kr

Abstract

Understanding politics is challenging because the politics take the influence from everything. Even we limit ourselves to the political context in the legislative processes; we need a better understanding of latent factors, such as legislators, bills, their ideal points, and their relations. From the modeling perspective, this is difficult 1) because these observations lie in a high dimension that requires learning on low dimensional representations, and 2) because these observations require complex probabilistic modeling with latent variables to reflect the causalities. This paper presents a new model to reflect and understand this political setting, NIPEN, including factors mentioned above in the legislation. We propose two versions of NIPEN: one is a hybrid model of deep learning and probabilistic graphical model, and the other model is a neural tensor model. Our result indicates that NIPEN successfully learns the manifold of the legislative bill's text, and NIPEN utilizes the learned low-dimensional latent variables to increase the prediction performance of legislators' votings. Additionally, by virtue of being a domain-rich probabilistic model, NIPEN shows the hidden strength of the legislators' trust network and their various characteristics on casting votes.

Introduction

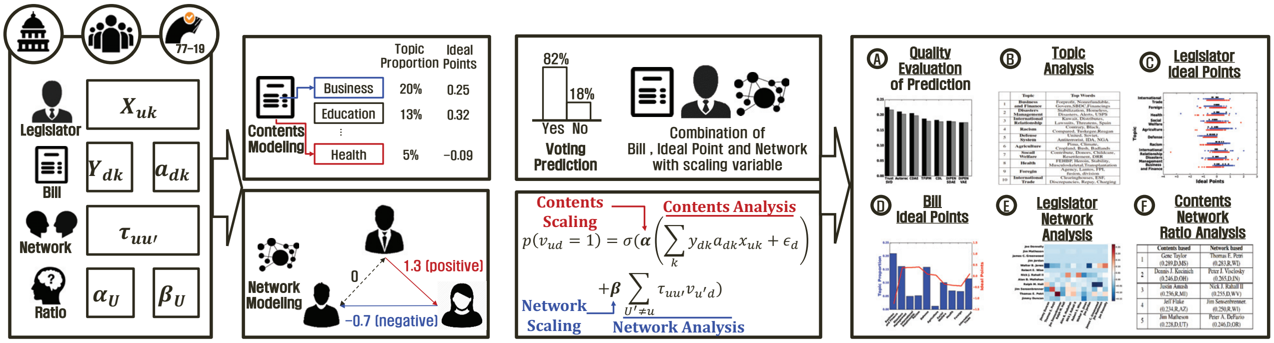
Recent developments in machine learning have enabled a deeper understanding of human behavior in diverse contexts. These advances include divulging intentions and sentiments in dialogs (Bertero et al. 2016); predicting purchases from online markets (Chong et al. 2017); recommending movies to friends (Shah, Rao, and Ding 2017); and discovering social network links between individuals (Guo, Zhang, and Yorke-Smith 2015). The recent machine learning models provide the contexts of these behaviors, which have been regarded as the latent aspects of human behavior.

One latent modeling of human behavior can be a form of complex Bayesian probabilistic models, a.k.a. probabilistic graphical model (PGM). The modelers used graphical notations, embedding the probabilistic variables and their causalities, to represent the key factors and their relations. For instance, latent Dirichlet allocation (LDA) models the generative process of documents, i.e. the composition of topics at large, a main topic of documents, and a word selection when describing a topic (Blei, Ng, and Jordan 2003).

Another effort in modeling the latent variable is improving the quality of the latent representation of the data. While the above probabilistic models focused on the contextual modeling, the latent variables reside in a high dimensional and nonlinear space, so the learning of the latent variables have been limited. For example, the stacked de-noising autoencoder (SDAE) (Vincent et al. 2010) learns this manifold space through encoding the noised inputs into the low dimensional latent representations; and reconstructing the original inputs with the latent representations with neural network layers. Further advances have made through casting this autoencoding mechanism to the variational inference approaches, and a variational autoencoder (VAE) (Kingma and Welling 2014) optimizes the variational distribution of the latent representations with neural networks.

Supported by the two research advances, one distinct research direction has been merging the latent representation learning and the probabilistic graphical model on human behavior. Collaborative deep learning (CDL) (Wang, Wang, and Yeung 2015) is one example merging SDAE with a probabilistic model of matrix factorization that often used to explain and predict the human behavior of recommendations. Whereas CDL gives a clear passway on how we can further develop various models of human behavior with support from the deep learning, different application domains require different latent modeling, so the model structure needs to be further customized and expanded.

This paper introduces Neural Ideal Point Estimation Network (NIPEN) which models the generative process of political voting by estimating ideal points in diverse legislative aspects with learning the low dimensional representations from neural networks. Specifically, we propose two versions of NIPEN. The first version, NIPEN-PGM is a hybrid model by representing the contextual causalities as a PGM, and by learning the low dimensional representations with multi-layered perceptron (MLP) autoencoders, i.e. SDAE and VAE. The second version, NIPEN-Tensor, is a neural tensor model that substitutes the PGM part with the neural tensor model. NIPEN-Tensor could be viewed as a generalized version of NIPEN-PGM. NIPEN-Tensor models the legislative voting with the tensor composition and the nonlinear operations between diverse legislative factors



$$p(v_{ud}=1)=\sigma(\alpha\left(\sum_k y_{dk}a_{dk}x_{uk}+\epsilon_d\right)+\beta\sum_{U'\neq u}\tau_{uu'}v_{u'd})$$

Figure 1: The summarized procedure of NIPEN. NIPEN predicts the votes with the combination of contents and network analyses. We can interpret not only an individual legislator's ideal points but also trust networks between legislators

while NIPEN-PGM assumes the marginalization and the linearized operation in the same modeling part.

Second, NIPEN is the most comprehensive model in the latent modeling of the political domain. Assuming that we model a voting process of legislators, NIPEN is the first model of unifying 1) the voting behavior, 2) the network influence between congressmen, 3) the political ideal point of bills and congressmen, 4) the textual topic of bills, and 5) the relative strength of network influence and ideal points when casting a vote. Some of these latent variables have been seen in other models, (Gerrish and Blei 2012; Gu et al. 2014; Chaney, Blei, and Eliassi-Rad 2015), but not as the unified model to depict a whole political picture. Since diverse factors, such as the contents of the bill and the human relations, greatly influence the voting (Cohen and Malloy 2014), an effective modeling of the legislative voting requires an integrated model, such as NIPEN. We show that NIPEN recorded significant performance improvements in all metrics compared to existing models. We also show various qualitative analyses that can only obtained via this comprehensive model. The entire procedures and analyses of NIPEN is summarized by Figure 1.

Previous Research

Modeling Political Network and Ideal Points

Network analyses and ideal point estimation have been widely studied in computer science and quantitative political science for its importance. In the line of political network analyses, most studies analyzed co-sponsorship data (Faust and Skvoretz 2002; Fowler 2006). Faust and Skvoretz (2002) clarified the topological structures in the network of the U.S. Senate (1973-1974), and they found that the network among U.S. Senator in 93rd Congress is O-star, I-star and Trans structure (Faust and Skvoretz 2002). Fowler (2006) inferred the relationship in U.S. Congress (1973-2004) by measuring the centrality to find the most central legislators (Fowler 2006). In the community of ideal point estimation, Poole and Rosenthal (1985) proposed a nonlinear logit model to account for political choices of legislators (Poole and Rosenthal 1985). However, it was a one-dimensional estimation, and the analysis could not identify what the ideal dimension stands for. To overcome the limitation, Clinton et al. (2004) proposed a multi-dimensional ideal point estimation model, but these models still remained at the simple logit model extensions (Heckman and Snyder Jr 1996; Clinton, Jackman, and Rivers 2004).

With the advance of topic modeling, multi-dimensional ideal point models were developed, and these models provide more accurate interpretations on the ideal points. Gerrish and Blei (2012) proposed an issue-adjusted model (Gerrish and Blei 2012) with the labeled LDA (Ramage et al. 2009), and Yupeng et al. (2014) proposed a topic-factorized ideal point model (TFIPM) (Gu et al. 2014) with probabilistic latent semantic analysis (PLSA) (Hofmann 1999) to estimate the ideal points of legislators based on roll-call data. Further extensions of TFIPM have made through including available domain data. For instance, Islam et al. (2016) proposed SCIPM by including co-sponsorship networks between judges in the supreme court (Islam et al. 2016). These works have remained in the extension of the probabilistic graphical model without the innovation from the deep learning community, which our work extends 1) the probabilistic graphical model with variational autoencoders and 2) the neural tensor model for the causality modeling of the legislative voting.

Collaborative Filtering and Deep Learning

Collaborative Filtering is a recommendation algorithm that considers the relationship between users and items (Koren, Bell, and Volinsky 2009). One of representative approach is a matrix factorization which factorizes the rating matrix as user latent and item latent factors.Recently, the deep learning has initiated two theoretic developments. First, the matrix factorization itself is a low-dimensional representation method because of its latent vector learning, so does the autoencoding in the deep learning. For example, Sedhain et al. (2015) proposed Autorec (Sedhain et al. 2015), a basic autoencoder based CF algorithm, and Autorec outperforms other state-of-the-art MF algorithms like LLORMA (Lee et al. 2013). Wu et al. (2016) expand Autorec by concatenating a user latent variable to the rating input information in the encoder part of Autorec (Wu et al. 2016). Li et al. (2015) adopted two autoencoders corresponding to users and items (Li, Kawale, and Fu 2015), and they showed the interaction mechanism between the two autoencoders by using the marginalized SDAE (Chen et al. 2012). Second, the matrix factorization is related to the low-dimensional feature representation by adding the representation of the model as the distilled version of the side information. For instance, Wang et al. (2015) proposed a collaborative deep learning (CDL) which combines SDAE with MF (Wang, Wang, and Yeung 2015). Furthermore, Ying (2016) proposed a model of collaborative deep ranking which combines ranking with algorithm and SDAE (Ying et al. 2016). Wang et al. (2017) proposed the relational deep learning with SDAE to link prediction between items (Wang, Shi, and Yeung 2017).

Method

This section introduce the detailed descriptions of NIPEN-PGM and NIPEN-Tensor in turn.

NIPEN with Probabilistic Graphical Model and Autoencoders

Figure 2 describes the model structure of NIPEN-PGM. We start the detailed description from the bill low dimension modeling part, which is the bill plate with the $d \in D$ subscript. We apply either VAE or SDAE to learn the low dimensional representation, or topic, of z_{dk}[1] with the observed bill's text w_{dv}. z_{dk} can be extracted through the probabilistic encoder, q_ϕ with parameter ϕ and decoder, p_θ with parameter θ. The bill's latent representation has two components:the bill's topic proportion z_{dk} and the latent offset ξ_{dk}, and we model the combination of the two component as the below.

$$y_{dk} = \xi_{dk} + z_{dk}, \quad \xi_{kd} \sim N(0, \lambda_y^{-1})$$

Since the bill itself and the bill's text may have two different latent variables, ξ_{dk} becomes the *offset* between the bill's latent representation and the bill's topic proportion.

From the defined bill's latent representation y_{dk}, we model how the bill's latent representation generates the voting observation r_{ud}. Here, $u \in U$ is the dimension of the legislators. We assumed that a legislator cast votes considering three latent factors: the bill's latent representation y_{dk}, the bill's ideal point a_{dk}, and the legislator's ideal point x_{uk}.

$$a_{dk} \sim N(0, \lambda_u^{-1}), \quad x_{uk} \sim N(0, \lambda_u^{-1})$$

Now, we define NIPEN-PGM without the network factor. This voting procedure is modeled as Eq. (1) where η_d is a bias value of a legislative bill, and σ is a sigmoid function. Eq. (1) is designed to increase the probability of voting *YEA* when the ideal points of the bill and the legislator have the same sign; and when an ideal-aligned dimension of the y_{dk} is high. Additionally, η_d indicates whether the bill is more broadly accepted or not, regardless of ideal points.

$$p(r_{ud} = 1) = \sigma(\sum_{k=1}^{K} y_{dk} a_{dk} x_{uk} + \eta_d) \tag{1}$$

[1] d, u, and k mean each document, legislator, topic respectively. Small subscripts indicate the row and column index in order.

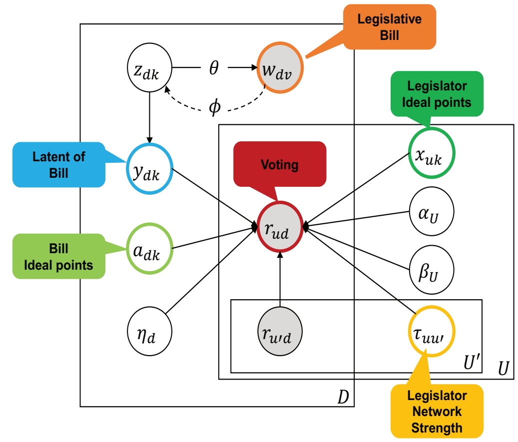


Figure 2: Graphical model representation of NIPEN-PGM

Finally, we add the network component to NIPEN-PGM. The interest of a particular legislative group could be an important factor in the voting process. Following this implication, we modeled the network between two legislators as below. Before the network modeling, we limited the network influence between the legislators sharing the same term, and this neighbor set, I_u, is defined as a neighborhood of legislator, u.

$$\tau_{uu'} \sim N(0, \lambda_\tau^{-1}) \quad \alpha_u \sim N(0, \lambda_\alpha^{-1}) \quad \beta_u \sim N(0, \lambda_\alpha^{-1})$$

The legislator u's voting is affected by two terms. The first term is the ideal alignment modeled in Eq. (1). The second term is the voting record of the neighbor legislator, $r_{u'd}$, and the second term is also weighted by the network strength, $\tau_{uu'}$, between the two legislators. Since this is a linear summation, $\tau_{uu'}$ will model the degree of voting agreement between two legislators. These two terms are unified with scaling parameters α_u and β_u. The purpose of modeling α_u and β_u is analyzing whether a certain legislator is influenced more either from the bill or from the network in casting votes. Eq. 2 is the overall voting formulation of NIPEN-PGM.

$$\begin{aligned} p(r_{ud} = 1) = \sigma(\alpha_u(\sum_{k} y_{dk} a_{dk} x_{uk} + \eta_d) \\ + \beta_u(\sum_{u' \in I_u} \tau_{uu'} r_{u'd})) \end{aligned} \tag{2}$$

NIPEN with Neural Tensor Model

Existing models, including NIPEN-PGM, do not directly model the relationships between the topics, which means that there is no cross-operiation between the dimension of K. Some cases, i.e. correlated topic model (Lafferty and Blei 2006), model the correlation between topics via the logistic normal distribution, but this is not an operation modeling of topic influences, rather the variable modeling of topic covariance. The recent introduction of neural tensor models (Socher et al. 2013) enable the cross-operations between the latent topic dimension. This topic cross-operation can model the legislator's ideal point non-linear influences when two

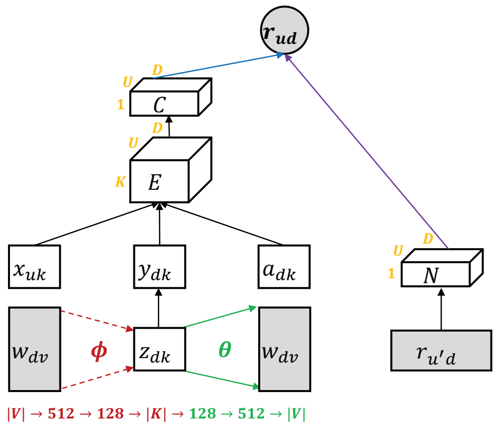


Figure 3: Neural network view of NIPEN-Tensor. The contents part is connected with the blue line (with content scaling parameter α_u), and the network part is connected with the purple line (with the network scaling parameter β_u).

topics are combined within a bill. Here, we propose NIPEN-Tensor to incorporate the cross-topic influence in casting a vote, which could not be modeled in NIPEN-PGM. NIPEN-Tensor and NIPEN-PGM are similar in the parts of document and influence network modeling. The only different part is the voting decision modeled as Eq. 2 which multiplies the factors per a topic and marginalizes. NIPEN-Tensor considers that the multiplication per a topic should be changed to consider the nonlinear effect from the topic set, not a single topic. Therefore, we represent the previous topic-wise multiplcaiton of $y_{dk}a_{dk}x_{uk}$ as a tensor E, and this tensor still treats the topic dimension to be independent. Then, we apply a fully-connected layer to cross-operate the topic dimension of E, and the neural network has C that is the output of the cross-operation. The overall structure and formulation for the NIPEN-Tensor are shown in Figure 3 and Eq. 3, respectively.

$$\begin{aligned} E_{udk} &= x_{uk} y_{dk} z_{dk} \\ \widetilde{E}_{udl} &= \tanh\Big(\sum_k E_{udk} W^{(T_1)}_{kl} + b^{(T_1)}_l\Big) \\ C_{ud} &= \sum_k \widetilde{E}_{udl} W^{(T_2)}_{l1} + \eta_d \\ N_{ud} &= \sum_{u' \in U} \tau_{uu'} v_{u'd} \\ p(r_{ud} = 1) &= \sigma\Big(\alpha_u C_{ud} + \beta_u \sum_{u' \in I_u} N_{u'd}\Big) \end{aligned} \tag{3}$$

$W^{(T_1)}, b^{(T_1)}, W^{(T_2)}$ are weights and biases applied to E_{udk}, $\widetilde{E}_{udl}$ tensor. In particular, $W^{(T_1)} \in R^{K \times K}$ models the correlation between topics, and $W^{(T_2)} \in R^{K \times 1}$ models the influence of each topic on the voting. Since the signs of x_{uk}, y_{dk}, and a_{dk} are important, we use $\tanh$ instead of ReLU (Rectified linear unit) to transform the outputs nonlinearly.

Parameter Inference of NIPEN

The parameters of both NIPENs are enumerated in the previous section, and we learn the parameters in two folds: learning the autoencoder to represent the bill's topic and the CF, alternatively. The first set of parameters related to autoencoders is $\psi^{(1)} = (\theta, \phi)$; and the second set of parameters related with the legislative-CF is $\psi^{(2)} = (y, a, \eta, x, W^{(T_1)}, W^{(T_2)}, b^{(T_1)}, \tau, \alpha, \beta)$.

The overall inference algorithm of both NIPENs follows the maximization of variational evidence lower bound with two assumptions. Following CDL, the first assumption is connecting the autoencoder and CF through ξ, and the strength is controlled by the variance of ξ, which is λ_y. When learning $\psi^{(1)}$, we apply the stochastic gradient variational Bayes (SGVB) estimator.

Second, we assumed that the variational distribution of $\psi^{(2)}$ as a point mass for simplicity, so the parameters of the variational distribution are updated by each casted vote record, which is traditional Bayesian belief updates. Specifically, the likelihood of the posterior is presented as the lower bound in the below. Then, the lower bound, which has realized values of $q_\phi(\mathbf{z}|\mathbf{w})$, $p_\theta(\mathbf{z})$ and an observed input, has only $\psi^{(2)}$, so the gradient method can find the *maximum a-posteriori*, or *MAP*, of $\psi^{(2)}$. As a summary, the objective function of both NIPENs is specified as follows:

$$\begin{aligned} \mathcal{L}_{NIPEN} = &-\mathrm{D}_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{w}) \| p_\theta(\mathbf{z})) + \frac{1}{L}\sum_{l=1}^{L} \log p_\theta(\mathbf{w}|\mathbf{z}^l) \\ &+ \frac{\lambda_f}{2} \sum_{(u,d), r_{ud} \neq 0} \frac{1 + r_{ud}}{2} \log p(r_{ud} = 1) \\ &+ \frac{\lambda_f}{2} \sum_{(u,d), r_{ud} \neq 0} \frac{1 - r_{ud}}{2} \log p(r_{ud} = -1) \\ &- \frac{\lambda_y}{2} \sum_{d=1}^{D} \|y_d - z_d\|_2^2 - \frac{\lambda_u}{2}(\|a\|_F^2 + \|x\|_F^2) \\ &- \frac{\lambda_\tau}{2}(\|\tau\|_F^2) - \frac{\lambda_\alpha}{2}(\|\alpha\|_2^2 + \|\beta\|_2^2) \end{aligned}$$

Similar to (Wang and Blei 2011; Wang, Wang, and Yeung 2015), the parameters related with the autoencoder and the legislative-CF are infered by coordinate ascents which maximizes $\mathcal{L}_{NIPEN}$. For legislative-CF related parameters $\psi^{(2)}$, we take the gradient of $\mathcal{L}_{NIPEN}$ w.r.t each parameters given the current θ and ϕ. Given the legislative-CF related parameters $\psi^{(2)}$, we infer the autoencoder related parameters by computing $\nabla_{\psi^{(1)}} \mathcal{L}_{NIPEN}$. We utilized the Tensorflow library (Abadi et al. 2016) to optimize the parameters.

NIPEN-PGM and NIPEN-Tensor are only different in the vote casting process, and the related term in the objective function is the third and the fourth terms with $\log p(r_{ud} = 1)$. These terms could be computed as the conventional gradient descent in two variants of NIPEN, so there is no change in the learning mechanism.

In the original definition, the network, τ, is a $|U|$-by-$|U|$ matrix, and the number of parameters becomes large given

$O(U^2)$. To reduce the squared complexity, τ is approximated by the product of $\widetilde{\tau_1}$ and $\widetilde{\tau_2}$ where $\widetilde{\tau_1} \in \mathbb{R}^{U \times G}$, $\widetilde{\tau_2} \in \mathbb{R}^{G \times U}$. We assume that $\widetilde{\tau_1}$ and $\widetilde{\tau_2}$ are not related. G can be interpreted as the number of groups containing the legislators. This approximation results in $O(GU)$ for the network parameter inference.

Table 1: Attributes of *Politic2013* and *Politic2016* dataset

	Politic2013	*Politic2016*
# of legislators ($\|U\|$)	1,540	1,537
# of bills ($\|D\|$)	7,162	7,975
# of votings ($\|D\|$)	2,779,703	2,999,844
# of House	1,299	1,266
# of Senator	241	271
# of Republican	767	778
# of Democrat	767	752
# of unique word ($\|V\|$)	10,000	13,581
Average # of unique word for each bill ($\frac{\sum_{d,v}(I_{w_{dv}>0})}{V}$)	192.77	378.66
# of bills less than 10 unique words	65	0
Period	1990-2013	1989-2016
Source	THOMAS	GovTrack
Data type	1 (YEA), -1 (NAY)	

Results

Datasets on Political Ideal Points

We used two roll-call datasets[2]. Table 1 provides the descriptive statistics of the two datasets: *Politic2013* and *Politic2016*. *Politic2013* limits the number of a unique word to 10,000, and there are 65 bills which have less than ten words, while *Politic2016* chooses 13,581 unique words, and there are no bills with less than ten words. *Politic2013* is a more sparse dataset than *Politic2016* in the ratings and the vocabulary sizes.

Baselines and Implementation Details

The variations of NIPEN were compared to five baseline models as follows:

- **TFIPM**: Topic Factorized Ideal Point estimation Model (Gu et al. 2014) is specialized in politics to analyze the roll-call data.
- **Autorec**: A simple autoencoder model which is utilized to predict the ratings. Autorec (Sedhain et al. 2015) encodes and reconstructs the rating matrix. We used Item-based Autorec.
- **Trust SVD**: Trust SVD (Guo, Zhang, and Yorke-Smith 2015), a type of trust-based matrix factorizations, is built on SVD++ with trust information.
- **CDAE**: Collaborative Denoising Autoencoder (Wu et al. 2016) used a denoising autoencoder with user latent variables.

[2] For the research community, we released the dataset on https://github.com/gtshs2/NIPEN (*Politic2013* was collected from (Gu et al. 2014))

- **CDL**: Collaborative Deep Learning (Wang, Wang, and Yeung 2015) used the deep learning and the CF, jointly. CDL improves performance by using document information additionally, and CDL uses SDAE to learn document manifold.

Quantitative Evaluations

We performed the five-fold cross-validation to quantitatively evaluate the variations of NIPENs, and the performance measures are RMSE, MAE, accuracy, and negative average log-likelihood (NALL) measures. We compared nine models: five baseline models in section 4.2, and four NIPEN variations, which are NIPEN-PGM(SDAE), NIPEN-PGM(VAE,approx.), NIPEN-PGM(VAE), and NIPEN-Tensor. NIPEN-PGM has three variants by choosing either SDAE or VAE as the autoencoder for the text modeling, and by choosing either using the whole matrix for the influence or the low-rank approximated matrix of the influence.

Table 2 statistically confirms that the best performance model in every metric is always a variation of NIPEN, which is confirmed with statistical significance. In detail, first, we compare NIPEN-PGM(VAE) and NIPEN-PGM(SDAE), and their performance gap is larger in *Politic2013* than in *Politic2016* which is a relatively sparse setting as shown in Table 1. We conjecture that NIPEN-PGM(VAE) is better in handling the sparse dataset than NIPEN-PGM(SDAE). Second, NIPEN-Tensor is a model that considers the correlation between topics, and NIPEN-Tensor may have a better performance when a bill's text has multiple topics with complex and rich textual information. As discussed in Section *Datasets on Political Ideal Points*, *Politic2016* has richer textual information than *Politic2013*, and we conjecture that this is the reason why NIPEN-PGM(VAE) in *Politic2013* and NIPEN-Tensor in *Politic2016* show better performances. Third, while the accuracy improvement is relatively small, the improvements on other metrics, particularly RMSE and MAE, are relatively large. Already, the baseline models achieve the accuracy higher than 95%, so the accuracy improvement could seem minimal. However, our likelihood estimation of *YEA* and *NAY* is considerably improved given the RMSE and the MAE improvement.

Qualitative Evaluations

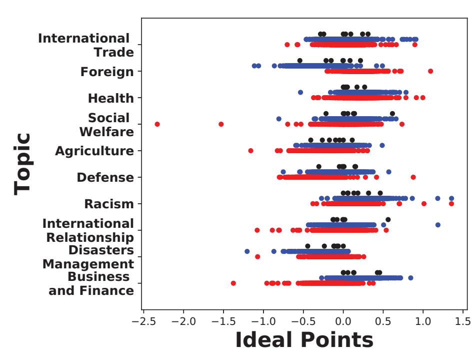


Figure 4: Individual legislator's ideal points for each topic

Table 2: Quantitative evaluation on *Politic2013* and *Politic2016* datasets. Two-standard deviation is shown in parentheses

	Politic2013				*Politic2016*			
	RMSE	MAE	Accuracy	NALL	RMSE	MAE	Accuracy	NALL
Trust SVD	0.2253 ($\pm$0.0007)	0.1399 ($\pm$0.0011)	0.9408 ($\pm$0.0003)	0.1866 ($\pm$0.0011)	0.2168 ($\pm$0.0011)	0.1353 ($\pm$0.0010)	0.9463 ($\pm$0.0009)	0.1782 ($\pm$0.0015)
Autorec	0.2110 ($\pm$0.0099)	0.0975 ($\pm$0.0136)	0.9411 ($\pm$0.0056)	0.1466 ($\pm$0.0177)	0.2031 ($\pm$0.0015)	0.0886 ($\pm$0.0110)	0.9454 ($\pm$0.0007)	0.1349 ($\pm$0.0125)
CDAE	0.2059 ($\pm$0.0007)	0.0831 ($\pm$0.0009)	0.9428 ($\pm$0.0006)	0.1450 ($\pm$0.0009)	0.1977 ($\pm$0.0037)	0.0802 ($\pm$0.0052)	0.9475 ($\pm$0.0023)	0.1357 ($\pm$0.0046)
TFIPM	0.1872 ($\pm$0.0002)	0.0682† ($\pm$0.0002)	0.9526 ($\pm$0.0003)	0.1213 ($\pm$0.0007)	0.1794 ($\pm$0.0010)	0.0625† ($\pm$0.0006)	0.9566 ($\pm$0.0005)	0.1121 ($\pm$0.0016)
CDL	0.1834† ($\pm$0.0008)	0.0786 ($\pm$0.0019)	0.9554† ($\pm$0.0004)	0.1147† ($\pm$0.0018)	0.1780† ($\pm$0.0013)	0.0769 ($\pm$0.0012)	0.9583† ($\pm$0.0008)	0.1106† ($\pm$0.0017)
NIPEN-PGM(SDAE)	0.1801** ($\pm$0.0014)	0.0591** ($\pm$0.0012)	0.9566** ($\pm$0.0006)	0.1155 ($\pm$0.0018)	0.1779 ($\pm$0.0005)	**0.0560**** ($\pm$0.0004)	0.9581 ($\pm$0.0003)	0.1173 ($\pm$0.0015)
NIPEN-PGM(VAE, approx.)	0.1804 ($\pm$0.0089)	0.0611* ($\pm$0.0065)	0.9565 ($\pm$0.0047)	0.1165 ($\pm$0.0086)	0.1791 ($\pm$0.0076)	0.0599 ($\pm$0.0057)	0.9571 ($\pm$0.0039)	0.1152 ($\pm$0.0070)
NIPEN-PGM(VAE)	**0.1753**** ($\pm$0.0007)	**0.0588**** ($\pm$0.0008)	**0.9587**** ($\pm$0.0006)	**0.1075**** ($\pm$0.0011)	0.1753** ($\pm$0.0017)	0.0570** ($\pm$0.0012)	0.9590** ($\pm$0.0010)	0.1112 ($\pm$0.0024)
NIPEN-Tensor	0.1818** ($\pm$0.0008)	0.0663** ($\pm$0.0003)	0.9556** ($\pm$0.0003)	0.1155 ($\pm$0.0020)	**0.1729**** ($\pm$0.0015)	0.0608** ($\pm$0.0006)	**0.9600**** ($\pm$0.0008)	**0.1057**** ($\pm$0.0022)
Improvement	4.41%	13.78%	0.35%	6.27%	2.87%	10.40%	0.18%	4.43%

NALL : Negative Average Log Likelihood
Improvement : Relative improvement of the best version of NIPEN compared to the best model, which is marked by †, among the baselines
$P^* < 0.05$; $P^{**} < 0.01$ (Student's one-tailed t-test against the † model)

Table 3: Selected top-five words for each topic. The number of listed topics was set to ten.

	Politic2013	*Politic2016*
# of legislators ($\lvert U\rvert$)	1,540	1,537
# of bills ($\lvert D\rvert$)	7,162	7,975
# of votings ($\lvert D\rvert$)	2,779,703	2,999,844
# of House	1,299	1,266
# of Senator	241	271
# of Republican	767	778
# of Democrat	767	752
# of unique word ($\lvert V\rvert$)	10,000	13,581
Average # of unique word for each bill ($\frac{\sum_{d,v}(I_{w_{dv}}>0)}{V}$)	192.77	378.66
# of bills less than 10 unique words	65	0
Period	1990-2013	1989-2016
Source	THOMAS	GovTrack
Data type	1 (YEA), -1 (NAY)	

In addition to the quantitative results, we interpret the latent variables of NIPEN-PGM(VAE) on *Politic2016*. First, to comprehend the dataset and the qualitative results, we computed the word-topic matrix from well-learned VAE variables, ψ_1, as shown in Table 3. This table provides a snapshot of topics in the bills. Then, we relate this topic to the bill's ideal points, a_{dk}. The latent dimension, k, becomes the common dimension of an ideal point value and a topic weight for each topic in the bill. Figure 5 shows an example of the topic weight as the bar chart and the ideal point value as the line chart. The illustrated bill, or H.Res.794 (114th), has the largest absolute value, $|a_{dk}\widetilde{z}_{dk}|$ in a 'Business and Finance' topic where $\widetilde{z}_{dk}$ denotes the normalized z_{dk}. This bill's ideal point is correlated with the legislator's

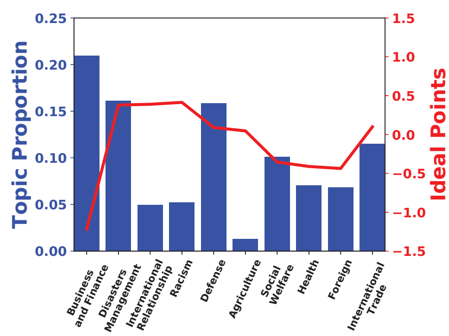


Figure 5: Topic proportion and ideal points of H.Res.794 (114th) bill

ideal point, x_{uk}, to generate the vote records. Here, the dimension, k, is the same latent dimension of the topic in Table 3, and we provide the scatter plot of the legislator's ideal points per topic in the Figure 4. The prior mentioned bill (H.Res.794 (114th)) considers the appropriations for financial services and general government, and the major topic is *Business and Finance*, and the bill's ideal point in *Business and Finance* is -1.217. Together, the vote casting will be determined by the legislator's view on *Business and Finance*, and this topic shows the greatest disagreement between the Republicans and the Democrats according to the Figure 4. In the real world, the voting results were same as expected: 1) the voting was very partisan, 92.2% Republican voted *YEA* and the 90.3% Democrat voted *NAY*. The second qualitative interpretation focuses on the legislator's net-

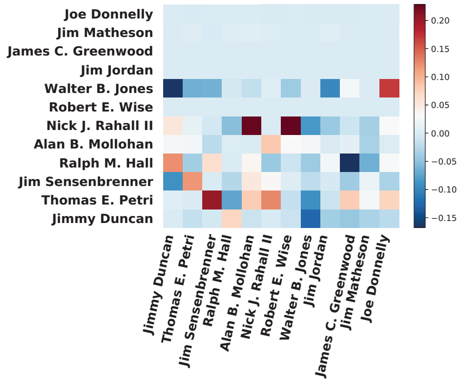


Figure 6: Trust network between legislators

work. We selected 12 legislators who have either strongly positive or negative relationships with each other, shown in the Figure 6. In general, the legislators have a strong positive relationship when they have the same district and the party. Among the top-five positive relationships, four of them have the same party and the same district, i.e. 'Thomas E. Petri↔Jim Sensenbrenner', 'Nick J. Rahall II→Robert E. Wise', and 'Nick J. Rahall II→Alan B. Mollohan'[3]. The closest relations are 'Thomas E. Petri' and 'Jim Sensenbrenner'. They were both republican representatives from Wisconsin, and they share similar voting patterns. They have voted 6,288 times for the same bill, and the 5,764 votes were same (91.6%). Especially, they voted *NAY* for H.R.730 (111th) which is a "suspension of the rules", and 397 legislators votes *YEA*.

The third qualitative analysis concentrates on the interaction between the contents and the network parts. We used two scaling variables α_u and β_u, which controls the strengths of contents factor and network factor, respectively. Table 4 shows the top-five legislators who were affected by either contents or network factors. Since the variations of NIPEN is an integrated model of network modeling as well as the textual bill modeling, the NIPENs should better perform than the baseline models, i.e. CDL, which only models the text, and Figure 7 confirms this hypothesis.

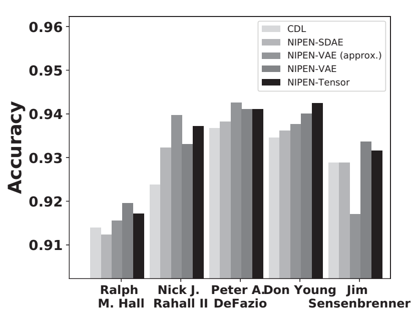


Figure 7: Accuracy of top five legislators who are affected by network factor

[3] τ_{uu} is asymmetric matrix. arrow('→') indicates the direction of the trust

Table 4: Top-five legislators who are affected by contents or network factors a lot. The scaling variable (α_u for contents based, and β_u for network based), political party, and district of the member are indicated in parentheses.

	Contents based	**Network based**
1	Ron Paul (0.260, R, TX)	Ralph M. Hall (0.304, R, TX)
2	Virgil H. Goode (0.220, R, VA)	Nick J. Rahall II (0.250, D, WV)
3	Dennis J. Kucinich (0.218, D, OH)	Peter A. DeFazio (0.247, D, OR)
4	Henry Cuellar (0.198, D, TX)	Don Young (0.228, R, AK)
5	Walter B. Jones (0.195, R, NC)	Jim Sensenbrenner. (0.227, R, WI)

Conclusion

We proposed two versions of machine learning models, NIPEN-PGM and NIPEN-Tensor, to analyze the ideaology in the legislation process. The variations of NIPEN show the state-of-the-art performance in all measures on *Politic2013* and *Politic2016*. Furthermore, NIPEN provides various interpretations in why *YEA* or *NAY* is casted by illustrating 1) the ideal point estimation of individual legislators and bills; 2) the trust network between legislators; and 3) the content and network influence for each legislator. These supervised and unsupervised tasks could be critical insights into quantitatively understanding politics in the legislative process.

Acknowledgments. This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No. R7117-17-0219, Development of Predictive Analysis Technology on Socio-Economics using Self-Evolving Agent-Based Simulation embedded with Incremental Machine Learning)

References

Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; and Others. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Bertero, D.; Siddique, F. B.; Wu, C.-S.; Wan, Y.; Chan, R. H. Y.; and Fung, P. 2016. Real-Time Speech Emotion and Sentiment Recognition for Interactive Dialogue Systems. *ACL*.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent Dirichlet Allocation. *The Journal of Machine Learning Research* 3:993–1022.

Chaney, A. J.; Blei, D. M.; and Eliassi-Rad, T. 2015. A probabilistic model for using social networks in personalized item recommendation. *Proceedings of the 9th ACM Conference on Recommender Systems* 43–50.

Chen, M.; Xu, Z.; Weinberger, K.; and Sha, F. 2012. Marginalized Denoising Autoencoders for Domain Adaptation. *Proceedings of the 29th International Conference on Machine Learning (ICML)* 767—-774.

Chong, A. Y. L.; Ch'ng, E.; Liu, M. J.; and Li, B. 2017. Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. *International Journal of Production Research* 55(17):5142–5156.

Clinton, J. D.; Jackman, S.; and Rivers, D. 2004. The Statistical Analysis of Roll Call Data. *The American Political Science Review* 98(2):355–370.

Cohen, L., and Malloy, C. J. 2014. Friends in high places. *American Economic Journal: Economic Policy* 6(3):63–91.

Faust, K., and Skvoretz, J. 2002. Comparing Networks Across Space and Time, Size and Species. *Networks* 32(2002):267–299.

Fowler, J. H. 2006. Connecting the congress: A study of cosponsorship networks. *Political Analysis* 14(4):456–487.

Gerrish, S., and Blei, D. M. 2012. How they vote: Issue-adjusted models of legislative behavior. *Advances in Neural Information Processing Systems* 25(1):2762–2770.

Gu, Y.; Sun, Y.; Jiang, N.; Wang, B.; and Chen, T. 2014. Topic-factorized ideal point estimation model for legislative voting network. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014.* 183–192.

Guo, G.; Zhang, J.; and Yorke-Smith, N. 2015. TrustSVD : Collaborative Filtering with Both the Explicit and Implicit Influence of User Trust and of Item Ratings. *Proceedings of the Twenty-ninth AAAI Conference on Artificial Intelligence (AAAI)* 123–129.

Heckman, J. J., and Snyder Jr, J. M. 1996. Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *National bureau of economic research* 28(0).

Hofmann, T. 1999. Probabilistic latent semantic indexing. *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* 50–57.

Islam, M. R.; Hossain, K. T.; Krishnan, S.; and Ramakrishnan, N. 2016. Inferring Multi-dimensional Ideal Points for US Supreme Court Justices. *Proceedings of the 30th Conference on Artificial Intelligence (AAAI)* 4–12.

Kingma, D. P., and Welling, M. 2014. Auto-encoding variational bayes. *In Proceedings of the International Conference on Learning Representations (ICLR).*

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix Factorization Techniques for Recommender Systems. *Computer* 42(8):42–49.

Lafferty, J. D., and Blei, D. M. 2006. Correlated topic models. In *Advances in neural information processing systems*, 147–154.

Lee, J.; Kim, S.; Lebanon, G.; and Singer, Y. 2013. Local Low-Rank Matrix Approximation. *ICML* 28.

Li, S.; Kawale, J.; and Fu, Y. 2015. Deep collaborative filtering via marginalized denoising auto-encoder. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 811–820. ACM.

Poole, K. T., and Rosenthal, H. 1985. A spatial model for legislative roll call analysis. *American Journal of Political Science* 357–384.

Ramage, D.; Hall, D.; Nallapati, R.; and Manning, C. D. 2009. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* 1(August):248–256.

Sedhain, S.; Menon, A. K.; Sanner, S.; and Xie, L. 2015. AutoRec : Autoencoders Meet Collaborative Filtering. *Proceedings of the 24th International Conference on World Wide Web (WWW)* 111–112.

Shah, V.; Rao, N.; and Ding, W. 2017. Matrix Factorization with Side and Higher Order Information. *arXiv preprint arXiv:1705.02047*.

Socher, R.; Chen, D.; Manning, C. D.; and Ng, A. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems*, 926–934.

Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; and Manzagol, P.-A. 2010. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of Machine Learning Research* 11:3371–3408.

Wang, C., and Blei, D. M. 2011. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 448–456. ACM.

Wang, H.; Shi, X.; and Yeung, D.-y. 2017. Relational Deep Learning : A Deep Latent Variable Model for Link Prediction. *AAAI*.

Wang, H.; Wang, N.; and Yeung, D.-Y. 2015. Collaborative Deep Learning for Recommender Systems. *KDD* 1235–1244.

Wu, Y.; DuBois, C.; Zheng, A. X.; and Ester, M. 2016. Collaborative Denoising Auto-Encoders for Top-N Recommender Systems. *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining - WSDM '16* 153–162.

Ying, H.; Chen, L.; Xiong, Y.; and Wu, J. 2016. Collaborative deep ranking: A hybrid pair-wise recommendation algorithm with implicit feedback. *Pacific-Asia Conference on Knowledge Discovery and Data Mining* 9652 LNAI:555–567.