Understanding the ecospace of philanthropic projects

H. Lamba
M. E. Helander
M. Singh
N. Lethif
A. Bhamidipaty
S. A. Baset
A. Mojsilović
K. R. Varshney

U.S. citizens donated an estimated $373.25 billion to charity in 2015. These donations came from individuals, corporations, and various foundations. Most of the funds were used for launching projects focused on topics in specific regions of the world. However, there is infrequent formal knowledge transfer between the wide array of projects, and often no singular, unifying historical database. Therefore, an organization initiating a new project may not be aware of what organizations it can partner with, what the estimated value (or the budget) should be, and what learning can be derived from projects that have happened in the past. In this paper, we study publicly available data from the Clinton Global Initiative's Commitment to Action directory, a philanthropic project portfolio comprising 3,200 projects, 10,000 organizations, and multiple topics such as healthcare and education (as of June 2016). We propose a kernel-based tensor factorization approach that provides recommendations to organizations starting on a new project, based on the lessons learned from previous projects.

Introduction

In this paper, we consider the topic of philanthropic project effectiveness. These global projects are facilitated by governments, nongovernmental organizations, and the private sector, including private foundations, with intentions such as economic development and humanitarian causes. In 2015, estimated donations from U.S. individuals, estates, foundations, and corporations totaled approximately $373.25 billion [1]. Over the past 60 years, government development aid from wealthy countries specifically to Africa exceeded $1 trillion. Interestingly, per-capita income in Africa today is lower than that in the 1970s, and currently, as many as 350 million people live on less than a dollar a day [2].

According to a report published in *Stanford Social Innovation* [3], there have been numerous instances of duplicative philanthropic projects, purportedly due to a failure within the field of philanthropy to share fundamental field research. An example of an organization repeating work occurred in Malawi, where a gravity-fed water pipe system commissioned by the Canadian government failed 10 years after the same system, commissioned by the United States, broke down [4]. Improved sharing of information about philanthropic projects with similar characteristics might have helped to avoid the second failure.

Recently, there has been significant criticism of the effectiveness of development aid [5–7], arguing that such aid may perpetuate the cycle of poverty and hinder economic growth in targeted regions [8]. Authors such as Dichter [9] have noted that the aid industry has sometimes become one in which the utility of the project is inversely proportional to the amount spent. Banerjee et al. [10] rigorously studied the successes and failures of various aid programs. They concluded that very few interventions—including those targeting illiterate adults, education vouchers for school uniforms and textbooks, and vaccination and HIV/AIDS programs—were effective.

To facilitate the sharing of knowledge and suggestions from past philanthropic projects, we propose a system that can take into account the details of past projects and provide suitable recommendations. Currently, an organization initiating a new project on a certain topic will not have an accurate perception about what its estimated value (budget) or duration should be, which may be recommended from previous projects. In fact, there is much information that

Digital Object Identifier: 10.1147/JRD.2017.2717098

can influence these types of recommendations, including details such as the type of facilitating or leading organization, the project topic, and the targeted region. Note that it is ineffective to treat each organization type, topic, or targeted region as independent. For instance, Pakistan and India might not be independent of each other due to geographic similarities; similarly clean water and health might be related issues. Recommendation clearly needs to take these kinds of dependencies into account.

We define the research problem as follows:

> *Given a dataset of philanthropic projects and additional information about the dependencies between different attributes of those projects, predict the potential value and duration to an organization that is planning to conduct a new philanthropic project, taking into account its topic and targeted region.*

To solve this problem, we considered publicly available data from the website of the Clinton Global Initiative that has managed a philanthropic project portfolio over the past 13 years. We analyzed 3,200 of its community's projects, initiated from around the world, covering various topics such as water sanitation, education, and global health. These projects are multifaceted. They include multiple dimensions—organizations, targeted countries, and topics. We refer to these multiple dimensions as *modes*. Additionally, there is vast similarity information that exists on each of these modes, which can be utilized. We refer to this information as *side-information.*

The large number of modes and the presence of side information delineated in the research problem leads to a unique system formulation for recommendation that is not typically encountered in applications such as e-commerce (the typical domain for recommendation). Although various simplifications and relaxations may be applied to the underlying machine learning task, the full complexity of the problem is only captured by fully integrated formulations. We propose a kernel-based tensor factorization approach to recommend an estimated value and estimated duration of a philanthropic project based on the knowledge of previous projects. This technically challenging paradigm of kernel-based tensor factorization was recently introduced to the literature by Lamba et al. [11]. Note that the proposed approach works with a variety of general modes and side information. The recommendation algorithm robustly takes into account all of these external data present, to produce final recommendations.

Our contributions are as follows:

- *Past knowledge*: Our paper uses computational methods to analyze the knowledge from past philanthropic projects to make recommendations for new projects.
- *Holistic information*: Our approach uses all the information available from past projects such as the region information, organization information, targeted topic, and the similarities between regions and organizations to provide the suitable recommendation.
- *Recommendation engine*: The proposed recommendation engine, unlike a standard two-dimensional collaborative-filtering-based recommendation engine, is multi-dimensional; i.e., it takes into account various modes of information. It also captures the similarities within each possible entry (organizations and regions) in a particular domain to recommend the output variable (budget and duration). In other words, it captures the relations between different countries, different organizations, and different topics. We show that the given approach performs better than the standard baseline approaches.

The remainder of the paper is structured as the follows. We first discuss the related work and follow this with a discussion of the dataset. Next, we introduce our approach for the recommender system. We discuss the experiment and results and conclude with a discussion of future work.

Related work

There has been a considerable amount of work to understand the effectiveness of philanthropic projects and understand what are successful practices. After years of continued philanthropic projects in the underdeveloped-economy parts of the world such as Africa, it has become essential to understand the effectiveness of development aid. Mosley [12] concluded that there is no correlation between aid and growth rate of gross national product in developing countries. A similar result was presented by Boone [13], who concluded that aid is ineffective because it leads to direct consumption of the aid rather than investment in the country. Moyo, in her book *Dead Aid* [5], criticizes the existing development aid efforts by providing multiple anecdotal evidences and potential reasons for the failures of initiatives and campaigns. She suggests a bottom-up approach instead of a top-down approach, which means understanding the needs of the countries and investing money in the nation's market rather than giving aid without a proper understanding. Burnside and Dollar [14] provide further empirical evidence that the effect of aid on gross domestic product growth is positive and significant in developing countries with "sound" institutions and economic policies, but less so in poorer economies. There is very little work in trying to understand the projects, past or ongoing, in a data-intensive manner. Our proposed recommender systems approach can assist in understanding the past projects and use this data to further assist organizations starting a new philanthropic project.

Recommender systems have been defined as software tools and techniques providing suggestions of valuable items or information to users [15]. They have been used extensively on e-commerce websites to recommend what product a user should buy, music websites to suggest what music a user should listen to, and on movie websites such as Netflix** to suggest what movies a user should watch, and so on. Many data-driven techniques have been proposed for recommender systems [16], but due to space constraints, we do not discuss all of them here. It is well known that latent-factor (LF) matrix factorization methods perform well for the recommendation problem [17]. Recent advances include clustering-based matrix approximation methods [18] and stable matrix approximation techniques [19] to improve generalizability of the algorithms.

Mnih et al. [20] introduced a probabilistic framework for matrix factorization (PMF) that has been used to solve the recommendation problem. A fully Bayesian treatment of PMF was shown to be efficient and accurate for the problem of recommendations [21]. PMF has been extended and applied in various applications. A major extension of PMF has been kernel PMF (KPMF), where side-information of users and items can be represented in a kernel and used to provide better recommendations [22]. KPMF has been extensively tested on image restoration and recommender systems [23–25]. Although KPMF has been useful in traditional recommendation system settings that involve the two modes—users and items—there are additional modes that KPMF cannot handle, such as temporal effects. However, there are specialized matrix completion methods that have been designed to account for specific side information such as social networks and time [26–29]. There also exist methods that are domain-independent [30].

Recommendation in the presence of a temporal mode can be viewed as a tensor completion problem (in contrast to matrix completion) problem [31]. Tensor completion techniques have been widely used to solve the recommendation problem. Xiong et al. [31] proposed a Bayesian version of probabilistic tensor factorization that also considers the time domain. Shashua and Hazan [32] proposed a non-negative tensor factorization approach, which also has application to computer vision problems. Acar et al. [33] proposed a scalable version of tensor factorization. Similarly, as in the case of matrices, the tensor completion problem can be enhanced by using kernels to capture the intra-mode relations. There have been attempts to use side information for tensor completion problems by changing the objective function of the tensor decomposition and adding regularization terms that take into account the side information [34]. None of these approaches takes into account the similarities available in all modes of the tensor as side information to assist tensor completion. Our work is closest to an earlier approach [11] that proposed kernelized tensor factorization for improved recommendations on e-commerce websites. We employ a similar approach to solve two tensor factorization problems: one for recommending estimated duration and the other for recommending estimated value of philanthropic projects.

There is very little work in trying to understand philanthropic projects, past or ongoing, in a data-based manner. Such an understanding entails analyzing the targeted regions, the durations of these projects, their budgets, and the partnerships among different organizations. We believe our work is the first to thoroughly analyze a comprehensive collection of recent philanthropic projects to obtain insights and analysis in this space.

Dataset

We use public data from the Clinton Global Initiative, with a philanthropic project portfolio comprising of 3,228 projects from 2005 to 2016, which is searchable at [35]. Based on data provided by implementing organizations, the projects are estimated to have a combined total impact on 430 million people over 180 countries. The projects are diverse, covering topics such as education, environment, energy, global health, and food systems. We believe that the diversity of this particular project portfolio gives a robust approximation of the entire philanthropic space.

While the foundation is not directly involved with the funding, implementation, or measurement of a given project, it is involved with developing the commitments, monitoring the projects, and providing opportunities for organizations to network with other like-minded organizations. For our analysis, we refer to a philanthropic project as a *commitment*, which is the terminology used by the foundation in managing its portfolio. The data obtained is self-reported to the foundation by each commitment maker, and we assume it to be correct.

Each commitment has the following fields:

Title: The title of the commitment indicates the goal of the project, and this can be useful to determine the topic/area of the project along with the proposed methodology.

Start year: The year in which the commitment was made.

Maker: The commitment maker is one or multiple organizations that are initiating the commitment, and will be primarily responsible to ensure that the commitment reaches its desired goal.

Partner: The partnering organizations collaborate with the makers to ensure that the commitment is successful.

Estimated duration: The estimated number of years the commitment will take to achieve its goal.

Estimated value: The estimated amount (generally, in U.S. dollars) that the commitment maker thinks will be spent on this commitment.

Region: The regions that will receive the benefit of the given commitment.

Countries: The exact list of countries in the above-mentioned region that will be targeted by the commitment.

Background: The background gives information about the commitment. It contains details about the problem the commitment is intended to solve, a brief overview of the approach, and the potential impacts the commitment maker is aiming for.

Partner information: Partner information contains information about the organization that the commitment maker is partnering with on this commitment or the desirable qualities of an organization it is willing to partner with.

Progress information: This information is updated, yearly, with progress of the previous year.

The following is a list of some of the values associated with our dataset: number of unique commitments (3,228), number of unique commitment maker organization (2,555), and number of unique partner organizations (7,822). The earliest and latest commitment launching year in the dataset were 2005 and 2016, respectively. There were 211 unique countries and regions targeted. The estimated duration is 1 year (minimum), 3.11 (mean), 3 (median), and 91 years (maximum).

To better understand the data, we further analyzed each factor of the commitment (i.e., Organizations, Regions, Durations, Values, and Launch Year). We present basic statistics of each of these factors in the remainder of this section.

Organizations

Organizations form a critical part of a commitment. The organizations are responsible for leading the commitments, and making sure that the commitments achieve their goal. They are responsible for providing the funding for ground-work operations. Organizations can be involved as partners in the commitment. The partners further assist commitment makers in operations. Most of the commitments (2,855) have one organization as commitment maker; 255 commitments had 2 organizations as makers, 44 commitments had 3 organizations, 24 commitments had 4, and 19 commitments had 5 organizations as makers. There is only one commitment that has 8 organizations working together to achieve the goal. The number of organizations as commitment makers follows a power law distribution. For partner organizations, we see a similar trend: the number of partner organizations per commitment follows power-law. This can be shown by seeing in the data that number of partner organizations $[X = (1, 2, 3, 4, 5, 6, 7, 8, \text{and} \geq 9)]$ have $[Y = (1{,}345, 395, 344, 263, 200, 135, 86, 73, \text{and } 387)]$ commitments, respectively.

Geographies

When we analyzed each commitment on the basis of its targeted geographies, we found that North America was the most popular region, followed by Africa. Analyzing the geographic distribution shows that North America is the predominant region with 1,442 commitments, followed by Africa (847). Asia and Latin America comprise 783 and 704 commitments respectively, followed by Middle East with 257 commitments. Europe, and Oceania had 220 and 95 commitments respectively. There were 17 commitments that targeted the entire world.

Estimated duration

Estimated duration is one of the vital fields in the commitment. Based on our discussions with experts overseeing the operations at the philanthropic organization, we learned that this value is difficult for commitment makers to estimate. A high number of commitments (858) had an estimated duration of only 1 year, while 683 were 2 years. The number of commitments running for (3, 4, 5, 6, 7, 8, 9, and 10) years was 660, 342, 323, 135, 77, 42, 22, and 52, respectively. There were only 26 commitments with estimated duration greater than 10 years. We can clearly see that the commitment distribution follows a power-law distribution, but it has a small peak again at 10 years. This might be due to the rounding-up by the organizations creating the report or making the commitment.

Estimated value

Similar to the estimated duration, estimated value is difficult to predict accurately as most of the organizations are unaware of the costs of the entire project. The inaccuracies in the estimated value often lead to failed projects. Currently, estimated value is an estimate provided by the commitment makers; however, this requires intensive manual work. The value distribution appears to be Gaussian, especially if the value is represented in the natural log-scale.

Proposed approach

In this section, we describe our proposed approach of recommending estimated value and estimated duration. Before presenting our approach, we will discuss some preliminary notations, and definitions.

Preliminaries

A tensor is a multidimensional array that generalizes vectors (i.e., one-way tensors) and matrices (i.e., two-way tensors) to higher dimensions. We provide a list of symbols and their definitions in **Table 1**. We use Q to denote the order (or the number of modes) of the tensors. We use N_i to denote the number of entries/fibers in the ith mode of tensor $\mathbf{R}$.

Tensor factorization is a method of decomposing a tensor into a core tensor and factor matrices such that their

Table 1 List of symbols.

Symbol	*Definition*
$\mathbf{R}$	$N_1 \times N_2 \times \ldots \times N_Q$ data tensor
$\mathbf{R}_{i,j,k}$	(i,j,k)-th entry of tensor $\mathbf{R}$. (Assuming order $Q = 3$)
Q	Order of tensor $\mathbf{R}$
N_q	Length of q-th mode of $\mathbf{R}$
D	Dimensions of the latent factors
$\mathbf{U}^{(q)}$	$N_q \times D$ latent matrix for q-th mode of $\mathbf{R}$
$\mathbf{U}_{i(q)}{}^{(q)}{}_{,:}$	$i_{(q)}$-th row of $\mathbf{U}^{(q)}$
$\mathbf{U}_{:,d}{}^{(q)}$	d-th latent factors for all rows of q-th mode of $\mathbf{R}$
$\mathbf{K}^{(q)}$	$N_Q \times N_Q$ covariance matrix for q-th mode of $\mathbf{R}$
$\mathbf{S}^{(q)}$	Inverse covariance matrix of $\mathbf{K}^{(q)}$

products approximate the original tensors. We use the famous canonical polyadic decomposition (CP) tensor decomposition method [36], which assumes that the core tensor has only super-diagonal entries. Mathematically, tensor decomposition is an optimization problem where the goal is to find factor matrices $\mathbf{U}^{(q)}$ to minimize the reconstruction error.

Probabilistic interpretation: Tensor decomposition

CP decomposition can be expressed as a graphical model. **Figure 1** (left) shows the independent model that previously has been explored [19, 21]. In this independent model, for each factor, each row is generated independently from the given distribution.

Since the independent model assumes that each row is generated independently, it misses the information that might exist between different modes. For example, the given model will assume that *clean water* and *clean environment* projects are independent.

Tensor decomposition with side information

A popular way of relaxing the independence assumption in matrix factorization problems has been to introduce kernels [23]. We extend the same idea to a general N-way tensor factorization. Lamba et al. [11] proposed that kernel matrices can be introduced to a general N-way tensor factorization approach as shown in Figure 1 (right). Each column of the factor matrices is generated from the kernel matrices, and hence each row of the factor matrix can be estimated based on the correlation with other factor matrices captured via kernel matrices. The generative model for the kernel-based tensor decomposition is as follows:

- For each factor q, generate each column d of the factor, $\mathbf{U}^{(q)}_{:,\mathrm{d}} \sim \mathrm{GsP}(0, \mathbf{K}^{(q)})$, where GP is a Gaussian process with covariance $\mathbf{K}^{(q)}$ and the colon notation indicates all rows.
- For each non-missing entry, $\mathbf{R}_{(\mathbf{I}_1\mathbf{I}_2 \ldots \mathbf{I}_Q)}$, generate $\mathbf{R}_{(I_1 I_2 \ldots I_Q)} \sim N(\sum_{d=1}^{D} \prod_{q=1}^{Q} U^{(q)}_{I_q,d}, \sigma^2)$.

In the first step, we are generating the factor matrix, where each column of the factor matrix is being sampled from a Gaussian process. The correlation across rows in the columns is according to the kernel. Once factors are generated, we consider their outer product and then add Gaussian noise with zero mean and given variance to each outer product term.

We need to estimate the parameters of this approach i.e. the factor matrices. This can be done via the maximum log-likelihood approach. The likelihood of the given process is obtained as follows. Given the factor matrices, $\mathbf{U}^{(q)}$, and the observed elements of $\mathbf{R}$, we have the following for observed data:

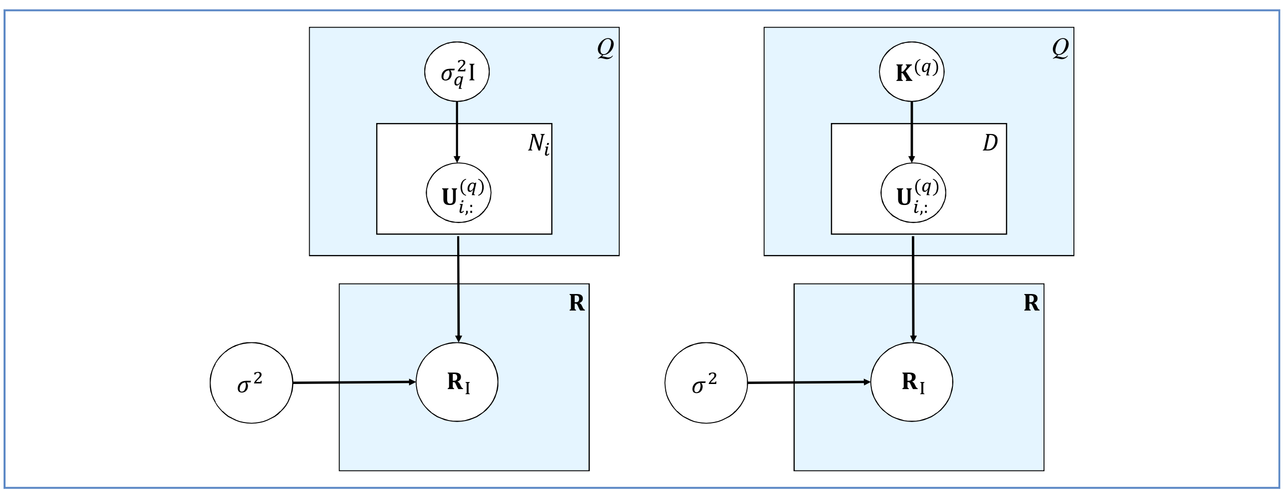

Figure 1

Graphical model for probabilistic tensor decomposition. *Left*: without kernel (independent), *Right*: with kernel.

$$p(\mathbf{R}|\mathbf{U}^{(1)},\mathbf{U}^{(2)}\ldots,\ \mathbf{U}^{(Q)},\ \sigma^2) = \prod_{I}\prod_{q=1}^{Q}\left[N\left(\mathbf{R}_{\mathrm{I}}|\mathbf{U}^{(q)}_{i_{q},:},\sigma^2\right)\right]^{\delta^I}, \quad (1)$$

where I is the set of all indices for $\mathbf{R}$, and $\delta^{\mathbf{I}}$ is indicator function which takes value 1 for when the particular value is observed, and 0 if it is not observed. Also, on the other hand, we have the following:

$$p(\mathbf{U}^{(q)}|\mathbf{K}_{\mathrm{U}^{(q)}}) = \prod_{d=1}^{D} GP\left(\mathbf{U}_{:,d}|0,\mathbf{K}_{\mathrm{U}^{(q)}}\right) \quad (2)$$

As a reminder, GP is a Gaussian process. From this, we can compute the log-posterior:

$$\begin{aligned}\log p(\mathbf{U}^{(1)},\mathbf{U}^{(2)}\ldots,\ \mathbf{U}^{(\mathrm{Q})}|\mathbf{R},\sigma^2,\mathbf{K}^1,\mathbf{K}^2\ldots\mathbf{K}^Q) &= \frac{-1}{2\sigma^2}\sum_{I}\delta_I\left(\mathbf{R}_I - \sum_{d=1}^{D}\prod_{q=1}^{Q}\mathbf{U}^{(q)}_{\mathrm{I}_{\mathrm{q}},d}\right)\\ &- A\log\sigma^2 - \frac{1}{2}\sum_{q=1}^{Q}\sum_{d=1}^{D}\left(\mathbf{U}^{(q)}_{:,d}\right)^{\mathrm{T}}\mathbf{S}^{(q)}\mathbf{U}^{(q)}_{:,d}\\ &- \frac{D}{2}\left(\sum_{q=1}^{Q}\log\left|\mathbf{K}^{(q)}\right|\right) + C,\end{aligned} \quad (3)$$

where A is the total number of non-missing entries, T is the transpose, $|\mathbf{K}|$ is the determinant of $\mathbf{K}$, and C is a constant. By differentiating the above log-posterior equation with respect to factor matrices, we obtain the update equations, which can be used in a stochastic gradient descent-like algorithm for finding the ideal parameters. We use learning rate $(\eta = 0.001)$ for stochastic gradient descent and regularization $(\sigma = 0.1)$ for the standard deviation.

Recommendation

We use a three-way tensor to represent the commitments data. The first mode of the given tensor represents the organizations involved; the second mode represents the concept or the topic of the commitment, and the third mode represents the targeted country of the commitment. Based on what we are recommending, the values of the tensor are either the estimated value or estimated duration. Given the number of unique organizations, the number of countries targeted, and the huge space of potential concepts, the tensor generated in this fashion may be high-dimensional, making the entire tensor factorization approach non-scalable. Therefore, to reduce the dimensionality of the tensor, we cluster the topic of the commitments based on similarities.

For additional scalability, we represent the values in the tensor (estimated value or estimated duration) using log-bucketization (i.e., the bucket for value of 6 with log base 2 will be 2). This allows us to perform the inference quickly and to provide an estimate for the final value. One of the limitations of representing commitments in tensor format is that for every commitment, some of the given modes are not unique; i.e., there might be multiple organizations or multiple countries in a single commitment. If a set is difficult to represent in a single value of the tensor, we expand the set to obtain multiple single values. For example, if commitment C_1 is (Organization *A*, Concept 1, Country *X,* Country *Y*, Duration *D*) then we represent C_1 as two values in the tensor as (Organization *A*, Concept 1, Country *X*, Duration *D*) and (Organization *A*, Concept 1, Country *Y*, Duration *D*). In the next section, we provide details about the tensors we used for recommendation, modes in tensors and how were they generated, and kernels (for side information).

Experiments

As defined earlier, our problem statement is as follows:

Given a dataset of philanthropic commitments and additional information about the dependencies between different attributes of those commitments, recommend an estimated value and estimated duration to an organization that is planning on conducting a new philanthropic project, taking into account its topic and targeted region.

We first define the modes of the tensors, i.e., the organization mode, concept mode, and country mode. Then, we go on to define the target values, i.e., estimated value and estimated duration. Subsequently, we define the kernels for each mode and finally discuss the experimental settings and the experiments performed.

Modes and tensor values

Organization mode

For our experiments, we consider commitment makers as the organization mode. This decision is to reduce sparsity in the tensor, thus making the tensor decomposition approach computationally tractable. Currently, we do not cluster organizations by attributes, as it is difficult to consistently obtain such information about the organizations in our dataset due to their variety, e.g., small, large, private, non-profit, public, etc.

Concept mode

The given dataset does not inherently have any defined variable for the concept or the topic of the commitment. Therefore, we design tools to infer it from the text that is associated with the commitment, i.e., the *background text*, which provides information about the commitment, and progress information. It is not trivial to find coherent concepts and assign each commitment to one of the concepts without defining a proper ontology of the terms used in this domain (which is a time-consuming task). Therefore, we apply latent Dirichlet allocation (LDA) for the task [37]. Since the text corpus is relatively small, we

Table 2 Five of the 20 topics (Clusters) found by LF-LDA [35].

Topic	*Top words*
1	Leaders, peace, society, international, leadership
2	Financial, services, access, income, credit
3	Children, education, school, early, child
4	Water, rural, access, areas, sanitation
5	Health, care, HIV, aids, treatment, medical

use latent representations of each word trained from larger corpora to apply to our corpus to enhance the coherence of the results of LDA. This is done by using a framework proposed to enhance LDA by appending it with the latent factors, also known as latent factor latent Dirichlet allocation (LF-LDA) [38]. Some of the topics such formed, along with top keywords in that topic, are shown in **Table 2**. We can observe that the topics are coherent, and each topic might align with a unique area of development aid (for instance, topic 5 might be related to healthcare). For every document, we obtain a probability of belonging to a topic. We assign each document to the topic with the maximum probability of belonging.

Country mode
The country mode is very straightforward. We use the exact country as a categorical variable in the tensor.

Kernels

Country similarity
As one concrete instantiation, we compute similarities between pair of countries by computing the geographic distance between them. We assume that countries that are closer to each other might be similar; for instance, *India* and *Pakistan* are geographical neighbors and very similar in terms of development aid. Once similarities are computed, we create a network by retaining the $N = 100$ most similar pairs. On this newly formed network, we define the country-based kernel as the commute time (CT) kernel i.e., $\mathbf{K}_{\mathrm{CT}} = \gamma \mathbf{L}^{-1}$, where γ is a regularization constant and set to 0.1 for our experiments, and $\mathbf{L}$ is the Laplacian of the adjacency matrix defined over the network we obtained earlier.

Concept similarity
To obtain an adjacency matrix, we compute similarities between the topics based on similarity between the probability vectors of the documents assigned to the topics. The same process is used to obtain the kernel as for the country level kernel.

To summarize, we use the modes and kernels for two tasks: estimated value prediction and estimated value prediction. The dimensions for both of these tensors by

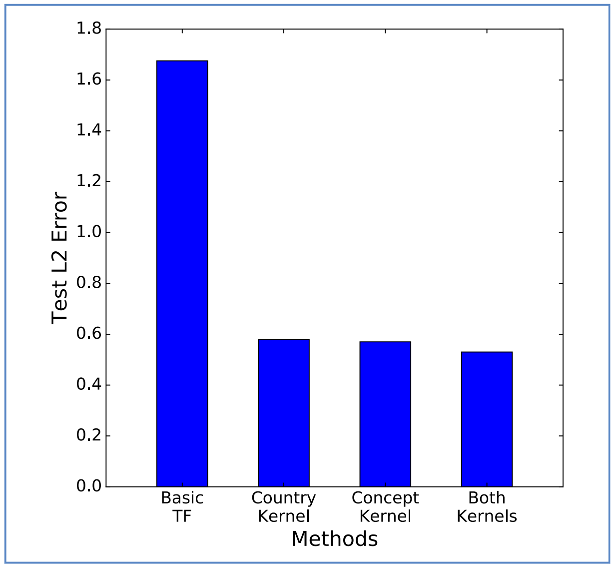

Figure 2

Performance of proposed approach for estimating duration.

modes are: organization mode ($N_1 = 2{,}461$), country mode ($N_2 = 213$), and concept mode ($N_3 = 20$).

Methodology
For experimentation, we define our baseline to be tensor-decomposition-based recommendation that does not include any side-information. Specifically, we use a basic tensor factorization approach (TF) as our baseline. To measure the performance of the methods, we use the L2 Norm error, which we define as follows:

$$L2\ Error = \sqrt{\sum_{i \in I} \delta_i \left(\mathbf{R}_i - \sum_{d=1}^{D} \prod_{q=1}^{Q} \mathbf{U}^{(q)}_{I_q,d} \right)^2}, \quad (4)$$

where δ_i is the indicator function that takes value 1 for when the particular value is observed, and 0 if it is not observed. We further define the recommendation tensors as $\mathbf{T}$, $\mathbf{V}$ for estimated duration and for estimated value tensors, respectively. Instead of using exact values, we log-bucketize the estimated value and estimated duration for each commitment to ameliorate scalability issues. Therefore, the results are an approximate bucket suggestion for any incoming query. We present the results of our experiments in the next section.

Results
We ran all our methods for 3,000 iterations, with a learning rate ($\eta = 0.001$) and regularization ($\sigma = 0.1$). The experiment was conducted with 90% training set and 10% testing set. We report the L2-error on the testing set. In **Figure 2**, we show the results of the task of estimating

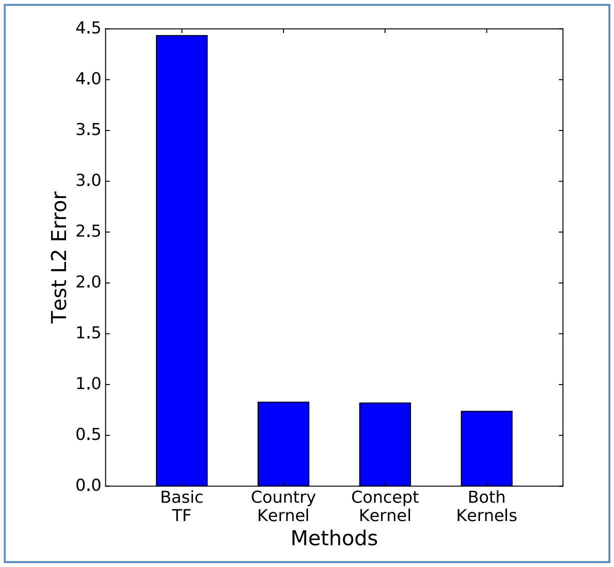

Figure 3

Performance of proposed approach for estimating value.

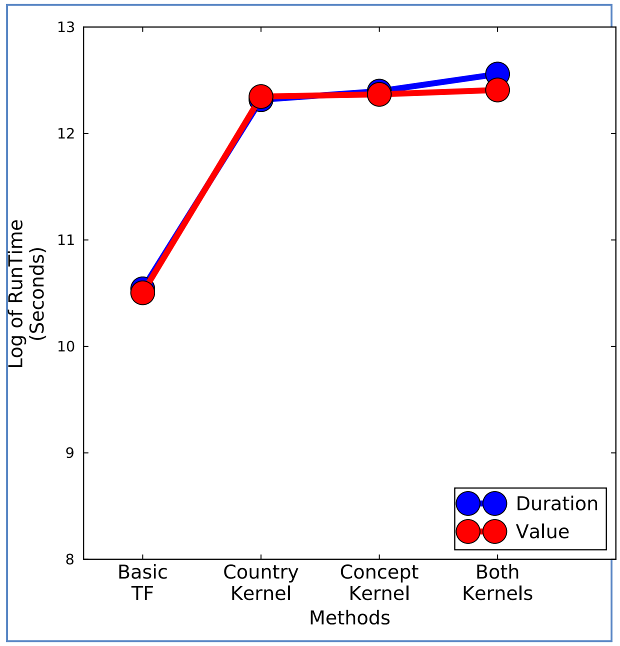

Figure 4

Time taken by the proposed approach for both the tasks.

value. We can see that the proposed approach with the both the kernels—country level and concept level kernel—perform slightly better than when either of the country kernel and concept kernel are applied. All of the kernel-based approaches perform significantly better than the baseline tensor factorization approach. The results indicate that adding concept kernel is slightly more informative than the country kernel. A similar trend can also be noted in **Figure 3** for the estimating duration task.

We also measure the program runtime (i.e., the wall clock time) of our approaches with respect to the baseline approach, and we use a 2.3-GHz i7 machine with 16 GB of RAM. The results are shown in **Figure 4**. Notably, the kernel-based approaches take approximately 10 times longer to converge; however, with that compromise regarding the runtime, the error is decreased by approximately 85%.

Conclusions

In this work, we analyzed 3,228 commitments, coming from 2,500 organizations across various topics. We analyzed a diverse set of commitments to extract a knowledgeable computer model from the data. We used this past knowledge as basis for our recommendation system. We propose a recommendation model which can work well with the given modes in the dataset, and can assist organizations launching new commitments by recommending the value and duration of the project. The proposed model takes into account additional information that might not be available in the dataset. Additional information can take into account the knowledge which might be useful for future commitments, such as the similarities between targeted countries, or similarities between the topic of the commitment. We show that the proposed approach performs better than the baseline method.

Limitations and future work

In this work, we tested our approach using the commute time kernel for concepts and countries. There can be many kernels, such as regularized Laplacian, and many other graph kernels that could be tested. Currently, the similarities between the countries are based on their geographic distance. In the future, they can be replaced with similarities due to their economic variables. Kernels for organizations can be proposed if attribute data of organizations can be extracted from the web or the commitment text. We assume that the estimating value and estimating duration are independent tasks, but it is very likely that duration is correlated with value, and therefore using the recommendation for duration could be further used to make inference about the value of the commitment.

The work addresses the problem of recommendation based on the assumption that all commitments in the dataset have been fulfilled. A natural extension will be to measure whether the commitments are fulfilled or unfulfilled and based on that, follow cues only from the fulfilled commitments. Furthermore, although the Clinton Global Initiative's commitments portfolio is publicly available on its website, an accessible user interface to the recommender engine will allow philanthropies from around the world to learn even more.

Acknowledgments

We thank the Clinton Global Initiative for suggesting this problem and for discussions about the data. This project was conducted under the auspices of the IBM Science for Social Good initiative.

**Trademark, service mark, or registered trademark of Netflix, Inc., or OpenStack Foundation in the United States, other countries, or both.

References

1. Giving USA, "The Annual Report on Philanthropy for the Year 2015," 2016. [Online]. Available: https://givingusa.org/product/2016-paperback-book-package
2. D. Moyo, "Why foreign aid is hurting Africa." [Online]. Available: http://www.wsj.com/articles/SB123758895999200083
3. C. Sherry, "Stop funding duplicative projects." [Online]. Available: https://ssir.org/articles/entry/stop_funding_duplicative_projects
4. D. Damberger, "Learning from failure." [Online]. Available: http://singjupost.com/transcript-david-damberger-on-learning-from-failure-at-tedxyyc/
5. D. Moyo, *Dead Aid: Why Aid is Not Working and How There is Another Way for Africa*. New York, NY, USA: Farrar, Straus and Giroux, 2010.
6. M. M. Miller, *Poverty, Inc*. [Documentary]. Online: Available: http://www.povertyinc.org, Dec. 2014.
7. N. Elzenga, "Gestrand ontwikkelingswerk," *Kijk Mag.*, pp. 42–47, Oct. 2008.
8. S. Majhanovich and G.-J. A. Macleans, Ed., *Economics, Aid and Education: Implications for Development*. Dordrecht, The Netherlands: Sense Publishers, 2013, pp. 17–18.
9. T. Dichter, *Despite Good Intentions*. University of Massachusetts Press, 2003.
10. A. V. Banerjee, A. H. Amsden, R. H. Bates, J. N. Bhagwati, A. Deaton, and N. Stern, *Making Aid Work*. Cambridge, MA, USA: MIT Press, 2007.
11. H. Lamba, V. Nagarajan, K. Shin, and N. Shajarisales, "Incorporating side information in tensor completion," in *Proc. Int. Conf. World Wide Web* , 2016, pp. 65–66.
12. P. Mosley, *Foreign Aid: Its Defense and Reform*. Lexington, KY, USA: Univ. of Kentucky Press, 1987.
13. P. Boone, *The Impact of Foreign Aid on Savings and Growth*. London, U.K.: London School of Econ. Political Sci., 1994.
14. C. Burnside and D. Dollar, "Aid, policies, and growth," *Amer. Econ. Rev.*, vol. 90, no. 4, pp. 847–868, 2000.
15. P. Resnick and H. R. Varian, "Recommender systems," *Commun. ACM*, vol. 40, no. 3, pp. 56–58, 1997.
16. F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, *Recommender Systems Handbook*. New York, NY, USA: Springer-Verlag, 2003.
17. Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *IEEE Comput. Soc.*, vol. 42, no. 8, pp. 30–37, Aug. 2009.
18. C. Chen, D. Li, Q. Lv, J. Yan, and S. Chu, "GLOMA: Embedding global information in local matrix approximation models for collaborative filtering," in *Proc. AAAI Conf. Artif. Intell.*, 2017.
19. D. Li, C. Chen, Q. Lv, J. Yan, L. Chang, and S. Chu, "Low-rank matrix approximation with stability," in *Proc. Int. Conf. Mach. Learn.*, 2016, vol. 48, pp. 295–303.
20. A. Mnih and R. Salakhutdinov, "Probabilistic matrix factorization," in *Proc. Neural Inf. Process. Syst.*, 2007, pp. 1–8.
21. R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 880–887.
22. T. Zhou, H. Shan, A. Banerjee, and G. Sapiro, "Kernelized probabilistic matrix factorization: Exploiting graphs and side information," *Proc. SIAM Int. Conf. Data Min.*, vol. 12, pp. 403–414, 2012.
23. M. Gonen and S. Kaski, "Kernelized Bayesian matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intll.,* vol. 36, no. 10, pp. 2047–2060, 2014.
24. K. Shin and U. Kang, "Distributed methods for high-dimensional and large-scale tensor factorization," in *Proc. Int. Conf. Data Min.*, 2014, pp. 989–994.
25. J. Mohsen and E. Martin, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proc. Int. Conf. Recommender Syst.*, 2010, pp. 135–142.
26. I. Porteous, A. U. Asuncion, and M. Welling, "Bayesian matrix factorization with side information and Dirichlet process mixtures," in *Proc. 24th AAAI Conf. Artif. Intell.*, 2010, pp. 563–568.
27. Z. Lu, D. Aggarwal, and I. Dhillon, "A spatio-temporal approach to collaborative filtering," in *Proc. Int. Conf. Recommender Syst.*, 2009, pp. 13–20.
28. A. Karatzoglou, X. Amatrian, L. Baltrunas, and N. Oliver, "Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering," in *Proc. Int. Conf. Recommender Syst.*, 2010, pp. 79–86.
29. N. D. Lawrence and R. Urtasun, "Non-linear matrix factorization with Gaussian processes," in *Proc. Int. Conf. Mach. Learn.*, 2009, pp. 601–609.
30. J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," in *Proc. Int. Conf. Comput. Vis.*, 2009, pp. 208–220.
31. L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. Carbonell, "Temporal collaborative filtering with Bayesian probabilistic tensor factorization," in *Proc. SIAM Int. Conf. Data Min.*, 2010, pp. 211–222.
32. A. Shashua and T. Hazan, "Non-negative tensor factorization with applications to statistics and computer vision," in *Proc. Int. Conf. Mach. Learn.*, 2005, pp. 792–799.
33. E. Acar, D. Dunlavy, T. Kolda, and M. Morup, "Scalable tensor factorization with missing data," in *Proc. SIAM Int. Conf. Data Min.*, 2010, pp. 701–712.
34. A. Narita, K. Hayashi, R. Tomioka, and H. Kashima, "Tensor factorization using auxiliary information," *Data Min. Knowl. Discovery*, vol. 25, no. 2, pp. 298–324, 2012.
35. Clinton Global Initiative Commitments. [Online]. Available: https://www.clintonfoundation.org/clinton-global-initiative/commitments
36. T. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
37. D. Blei, A. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
38. D. Nguyen, R. Bilingsley, L. Du, and M. Johnson, "Improving topic models with latent feature word representations," *Trans. Assoc. Comput. Linguist.*, vol. 3, pp. 299–313, 2015.

Received January 14, 2017; accepted for publication February 11, 2017

Hemank Lamba *Carnegie Mellon University, Pittsburgh, PA 15213 USA (hlamba@cs.cmu.edu).* Mr. Lamba is a graduate student at School of Computer Science at Carnegie Mellon University (CMU). Prior to this, he was a research software engineer in Telecom and Mobile Analytics team at IBM Research - India. His research interests include data mining, user behavior modeling, and social networks analysis. He is a 2016 CMU Presidential Fellow. He has also been involved with the Data Science for Social Good program at University of Chicago and the Science for Social Good Initiative at IBM. He completed his bachelor's program from Indraprastha Institute of Information Technology Delhi (IIIT Delhi) in 2012.

Mary E. Helander *IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (helandm@us.ibm.com).* Dr. Helander is a math scientist in the Data Science Department at the IBM T. J. Watson Research Center. She received a B.A. degree in computer information systems and mathematics from SUNY Potsdam in 1983, an M.S. degree in industrial engineering and operations research from Syracuse University in 1984, and a Ph.D. degree in industrial engineering and operations research from the University of Buffalo in 1992. Dr. Helander was previously a Research Fellow and faculty member at Linköping University, in Sweden. She has published in journals such as *Networks*, *Transportation Science*, *Computers and Operations Research*, and *IEEE Transactions on Software Engineering*.

Moninder Singh *IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (moninder@us.ibm.com).* Dr. Singh has been with IBM Research since 1998. He received his Ph.D. degree in computer and information science from the University of Pennsylvania in 1998. Dr. Singh is a member of the Data Science Department and is primarily interested in developing and deploying solutions for interesting problems in business analytics and decision support. His main research areas are machine learning and data mining, data privacy, information retrieval, probabilistic modeling and reasoning, and text mining.

Nizar Lethif *IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (nlethif@us.ibm.com).* Since 2014, Mr. Lethif has been a Software Engineer at the IBM T. J. Watson Research Center.

Anuradha Bhamidipaty *IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (abhamidi@us.ibm.com).* Ms. Bhamidipaty is a Technical Staff Member at IBM Research since 2003. She works in the Service Delivery Process Management group on business process innovations for service delivery. Her interests include process compliance and mining, application of machine learning techniques in process business management, and studying process variations and process improvements for delivery excellence. She has published papers in leading conferences including Knowledge Discovery and Data Mining and Very Large Databases. She received her B.Tech. degree in computer science from Sri Venkateswara University, Tirupati, India, and her M.Tech. degree in information technology from Indian Institute of Technology (Bombay) in 2000 and 2002, respectively.

Salman A. Baset *IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (sabaset@us.ibm.com).* Dr. Baset is a Research Staff Member at IBM T. J. Watson Research Center. His research is at the intersection of security, cloud, and performance. He has developed novel approaches for managing application security on cloud, which are now part of IBM cloud. He is also active in standardization and open source communities. As chair of SPEC Cloud subcommittee, he led the design and implementation of SPEC Cloud IaaS 2016, the first industry standard cloud benchmark. He is a co-author of RELOAD protocol (RFC6940, published by Internet Engineering Task Force) for building peer-to-peer communication systems. His work has been recognized through a SPEC Presidential Award in 2016, Young Scholars Award by Marconi Society in 2008, a best paper award at IPTCOMM in 2010, and several IBM awards. From 2012-2014, he served as the chair of Distributed and Fault Tolerant Computing (DFTC) Professional Interest Community (PIC) at IBM.

Aleksandra Mojsilović *IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (aleksand@us.ibm.com).* Dr. Mojsilović is an IBM Fellow and scientist at the IBM T. J. Watson Research Center. She received her Ph.D. degree in electrical engineering from the University of Belgrade, Belgrade, Serbia in 1997. She was a Member of Technical Staff at the Bell Laboratories, Murray Hill, New Jersey (1998-2000), and then joined IBM Research, where she currently leads the Data Science Department. She is a founder and co-director of the IBM Science for Social Good initiative. Her research interests include multidimensional signal processing, predictive modeling, and pattern recognition. She has applied her skills to problems in computer vision, healthcare, multimedia, finance, human resources, public affairs and economics. She is one of the pioneers of business analytics at IBM and in the industry; throughout her career she championed innovative uses of analytics for business decision support. For her technical contributions and the business impact of her work, Dr. Mojsilović was appointed an IBM Fellow, the company's highest technical honor. She is the author of over 100 publications and holds 16 patents. Her work has been recognized with several awards including IEEE Signal Processing Society Young Author Best Paper Award, Institute for Operations Research and the Management Sciences (INFORMS) Wagner Prize, IBM Extraordinary Accomplishment Award, IBM Gerstner Prize, and Best Paper awards at the European Conference on Computer Vision (ECCV) and the Conference on Service Operations and Logistics, and Informatics (SOLI). She is an IEEE Fellow and a member of INFORMS and Society of Women Engineers (SWE).

Kush R. Varshney *IBM Research, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (krvarshn@us.ibm.com).* Dr. Varshney is a Research Staff Member and manager in the Data Science Department at the IBM T. J. Watson Research Center. He received his Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 2010. He applies data science and predictive analytics to human capital management, healthcare, olfaction, public affairs, and international development. He conducts academic research on the theory and methods of statistical signal processing and machine learning. His work has been recognized through best paper awards at the 2009 International Conference on Information Fusion, 2013 IEEE Conference on Service Operations and Logistics, and Informatics (SOLI), 2014 ACM SIGKDD (Association for Computing Machinery's Special Interest Group on Knowledge Discovery and Data Mining), Conference on Knowledge Discovery and Data Mining, and 2015 SIAM (Society for Industrial and Applied Mathematics) Conference on Data Mining (SDM). Dr. Varshney is codirector of the IBM Science for Social Good initiative.