Political Dimensionality Estimation Using a Probabilistic Graphical Model

Yoad Lewenberg
The Hebrew University
of Jerusalem, Israel
yoadlew@cs.huji.ac.il

Yoram Bachrach
Microsoft Research,
Cambridge, United Kingdom
yobach@microsoft.com

Lucas Bordeaux
Microsoft Research,
Cambridge, United Kingdom
lucasb@microsoft.com

Pushmeet Kohli
Microsoft Research,
Redmond, WA, USA
pkohli@microsoft.com

Abstract

This paper attempts to move beyond the left-right characterization of political ideologies. We propose a trait based probabilistic model for estimating the manifold of political opinion. We demonstrate the efficacy of our model on two novel and large scale datasets of public opinion. Our experiments show that although the political spectrum is richer than a simple left-right structure, peoples' opinions on seemingly unrelated political issues are very correlated, so fewer than 10 dimensions are enough to represent peoples' entire political opinion.

1 Introduction

The problem of best describing political variation has been a key issue in social sciences for over a century, and many models have been proposed over the years. The most prominent system for classifying political positions, ideologies and parties is the Left-Right classification. The notions of "Left" and "Right" in politics originate from the seating arrangements in the French legislative body during the French Revolution of 1789: the aristocracy sat on the right of the Speaker and the commoners sat on the left. The key ideological point of contention was the "old order", with the Right supporting aristocratic interests and the church, and with the Left supporting republicanism, secularism, and civil liberties.

However, is using the left-right terminology justified by *data* about political opinions? The terms left-wing and right-wing have evolved to capture a somewhat different meaning in modern day USA than their meaning in the years following the French Revolution. The left-wing generally refers to egalitarianism, social policies supporting the working class, and multiculturalism, typically including socialists and libertarians, identifying with the Democratic party; The right-wing refers to conservative Christian values, support for a free-market system, and traditional family values, and including conservatives and free-market supporters, identifying with the Republican party.

Opinions regarding a *single* issue can easily be expressed on a single axis. Parties or people can be placed on the axis depending on the degree of support to a stance relating to this issue; those whose strongly support a position can be placed on one side, followed by those with who slightly support it, and so on, ending with those strongly opposing the stance. For example, we can place parties who strongly support heavy regulation of businesses on the far left, those who want complete business freedom under any circumstances on the far right, and those who support light regulation of some businesses in the center.

However, most political parties take an ideological stand regarding *many* issues: immigration, free medical care, minimum wage, regulating banks, religious freedom, abortions, gay and lesbian rights, regulating drugs and alcohol, and many others. When the ideology of parties spans multiple issues, representing these ideologies requires a *political spectrum* — a system for classifying different political positions using *several* geometric axes, each representing an independent political dimension. For example, a spectrum with two axes may include one axis for sociocultural issues (relating for example to supporting or opposing a heavy investment in welfare) and one for economic issues (for example, supporting or opposing de-regulation of business).

One way to define a political spectrum is by using a dimension for each of the important issues the people and parties care about. Given a political spectrum of K dimensions we can express people's opinions and the political ideology of parties as K-dimensional vectors. However, a large dimensionality makes it cumbersome for people to explain their views, so clearly we would like to use the smallest possible dimensionality that can fully express the opinions of most people and parties. Political scientists have noted that the wide popularity of the left-right identification stems from the surprising fact that in many countries, it *is* possible to map parties into a single left-right axis [39, 17]. For example, Von Beyme [38] categorized European parties into nine "families" that described most parties, and was able to linearly order seven such families from left to right: communist, socialist, green, liberal, Christian democratic, conservative and right-wing extremist. Although a single left-right axis can describe many parties, in many countries parties may take any combination of several issues [24, 39]. Common examples are the issues of economic freedom (taxation, free trade, free enterprise) and personal freedom (drug legalization, abortion and conscription). Some populations require a large dimensionality to represent, as individuals or parties may take any combination of positions regarding many issues, whereas other populations can be represented using less dimensions, due to strong correlations between stances on many issues.

Currently, 98 out of the 100 members in the United States Senate are affiliated with either the Democratic party (left) or the Republican party (right). This low-dimensional political landscape contrasts to other countries. For instance, in the United Kingdom, the House of Commons is composed of 11 parties. This disparity in the number of parties suggests that the political spectrum in the USA has fewer dimensions than the political spectrum in the UK.

Political scientists have to infer the key dimensions in a political spectrum representation using *data mining* techniques. For example, Ferguson [10] has used a set of questions pertaining to many issues: birth control, capital punishment, censorship, communism, evolution, law, patriotism, theism, treatment of criminals, and war. He used People's responses to questions pertaining to these issues as the input to a *factor analysis process*, trying to describe the variability between the correlated responses on different issues, using a low number of latent factors. His analysis showed that three dimensions, which he called Religionism, Humanitarianism, and Nationalism, were sufficient to capture much of the variability in the data. In other words, most people in the dataset could be described using three numbers, so that their position regarding *all* issues could be predicted given these numbers with high accuracy.

Such factor analysis based techniques for building a political spectrum can be thought of in terms of *dimensionality reduction*. This process transforms the data in the high-dimensional space into a representation in a space of fewer dimensions. Many such techniques are based on algebraic methods, such as SVD (singular value decomposition) or PCA (principal component analysis). Given the data and a target number K of dimensions, they find a good representation of the original data in a Kdimensional space. In this sense, these techniques can be thought of as a lossy compression technique. They receive peoples' opinions on many questions, and attempt to characterize both people and questions using very concise descriptions (vectors in a low dimension space). The original responses of the people regarding the political stances can then be reconstructed approximately. However, the approximation quality depends on the compression ratio: with a high dimensionality it is possible to represent people on the spectrum so that the full opinion regarding any issue can be determined accurately, and with a low dimensionality individuals may be represented very concisely, but with a higher error.

Although factor analysis can be a useful tool in inferring possible dimensions on the political spectrum, it is still unclear what the dimensionality of the political spectrum should be. How can we use data mining and machine learning tools in order to determine the true dimensionality of the data? How accurately can we infer peoples' political opinions using their location in this political spectrum?

1.1 Our Contribution:

We analyze the optimal number of political dimensions to use. In contrast to earlier data driven approaches for analyzing data so as to construct a political spectrum, we do not use an algebraic dimensionality reduction technique. Rather, we use a *Bayesian* model selection approach.

We use a *probabilistic graphical model* for dimensionality reduction, representing both users and questions regarding political stances as feature vectors in a low dimensional space, where similarity is measured by the inner product. Thus, each coordinate in the low dimensional space relates to one political dimension. Any choice of the number of dimensions in the low dimensional space results in a slightly different model. We choose the most plausible number of dimensions given the data, by taking the model with the *maximal evidence*.

We apply the technique on two datasets, each containing People's responses regarding various questions about their political stances. The first dataset consists of the responses of 38,000 UK users who ranked political issues by their importance. The second dataset consists of the responses of 1,500 users from the USA, sourced from Amazon's Mechanical Turk, who rated the degree to which they agree to 56 sentences representing a political stance. In each of these we use our approach to determine the optimal dimensionality of the political spectrum.

On the one hand, our results indicate the political spectrum is richer than simple "left-right" structure represented using a single dimension. On the other hand, they indicate that peoples' opinions on seemingly unrelated political issues are very correlated, so fewer than 10 dimensions are enough to represent peoples' entire political opinion.

2 Methodology

The key issue we focus on is determining how many dimensions underly the political positions of people with respect to a broad range of questions. We use statistical tools akin to factor analysis. Factor analysis methods attempt to represent a set of observed correlated variables in terms of several 'common' factors. The common factors are not directly observed in the data and thus are sometime called latent variables. Existing approaches use factor analysis to identify the main factors in political values, but use an algebraic factor analysis, where the number of political dimensions is an *input* of the factor analysis algorithm [10, 9]. In contrast, we use a Bayesian approach, which chooses the optimal dimensionality to use.

We first describe our high level methodology. Our technique receives an input dataset which contains the responses of many participants, P, regarding a set of political questions, Q. Each such question represents a political stand, and the participant is asked to express the degree to which they agree or disagree with the stand, on a numerical scale. For example, such an item may be "Alcohol and cigarettes should be heavily taxed.", and a participant must rate the item on a seven-point scale between strongly disagreeing (1) and strongly agreeing (7) with the statement.[1] When the dataset relates to $|P|$ participants responding to $|Q|$ questions, the dataset is thus a matrix of $|P| \cdot |Q|$ numbers.

Given the dataset, we apply a dimensionality reduction procedure that models both participants and questions as low dimensional vectors. If the dimensionality used in our procedure is chosen to be K, each participant and question are represented by vectors in $\mathbb{R}^K$. Let $p' \in \mathbb{R}^K$ be the representation chosen for the participant, and let $q' \in \mathbb{R}^K$ be the representation chosen for the question. The procedure is devised so that the *predicted* rating the participant $p \in P$ would give to a question $q \in Q$ is the inner product of these two vectors $\langle p', q' \rangle$. The dimensionality reduction is based on a probabilistic graphical model. Any choice of a dimensionality K for the political spectrum results in a slightly different such model $\mathcal{H}_K$. We use the dimensionality for which the evidence under the model is maximized.

3 PMPS Model

The dimensionality reduction model we use is a Probabilistic Model for expressing the Political Spectrum, or PMPS for short. PMPS is a probabilistic graphical model which resembles other Bayesian models for matrix factorization [2, 28, 33, 6].

Graphical models were first introduced by Pearl [27]. We use the more general framework of factor graphs (see e.g., [19]) in order to describe the factored structure of the assumed joint probability distribution among the variables. Once the graphical model is defined and the values of the observed variables are set, inference algorithms (such as approximate message-passing methods) can be used in order to infer the marginal probability distribution of the unknown variables [19].

[1] For a ranking of political issues, we assign integer scores for these topics in consecutive order.

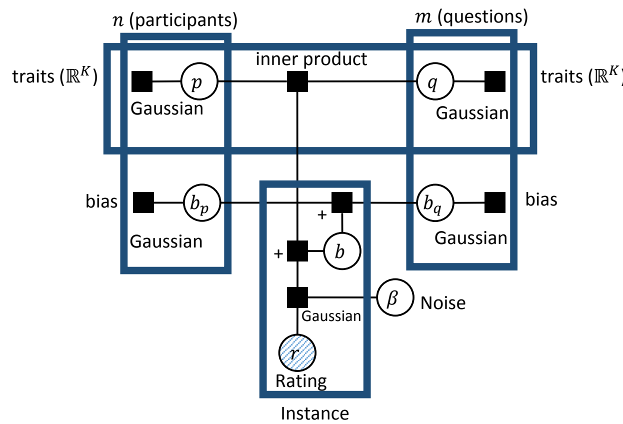


Figure 1: **Factor graph for the PMPS model.** The large plates indicate parts of the graph which are repeated with repetition indexed by the participant, question or trait.

The data fed to the model is a set of observations of the form (p, q, r) where $p \in P$ is the participant, $q \in Q$ is the political stance question, and $r \in \mathbb{N}$ is the rating given by the participant, expressing the degree to which they agree to the stance presented in the question.

The model assumes that the participants and questions can be characterized by K underlying "traits". These traits might be, for instance, the user's opinion on economy related issues. Note that the number of dimensions K of the model, i.e., the size of the participant and question vector has to be determined before the construction of the model. We model the process by which a participant $p \in P$ produces an answer r to a question $q \in Q$ as inner product between two K dimensional vectors of unobserved (latent) variables, one for the participant and one for the question. Thus we assume that: every participant has a latent trait vector and a latent bias; and Each question has a latent trait vector and a latent bias. The bias captures the fact that some participants give higher ratings than others, and that some questions receive higher ratings than others on average. Information such as the latent vector and bias of the participants and questions are modeled as unobserved variables, whereas the given response to a question by a user is modeled as an observed variable.

PMPS is a joint probabilistic model whose factor graph is given in Figure 1. Namely, the rating r that a participant with a latent vector $p \in \mathbb{R}^K$ gives to a question with a latent vector $q \in \mathbb{R}^K$ is modeled as

$$\Pr(r|t, s, b) = \mathcal{N}\left(r|p^T q + b_p + b_q, \beta^2\right), \tag{1}$$

where β is the standard deviation of the observation noise, and $b_p + b_q$ are the participant and question bias, respectively.

In order to do inference on the model, we must first define prior distributions for the variable of interest. We assume factorizing Gaussian priors for the participant vector traits $p_i \sim \mathcal{N}(\mu_p, \sigma_p^2)$ and bias $b_p \sim \mathcal{N}(\mu_p, \sigma_p^2)$, and question vector traits $q_i \sim \mathcal{N}(\mu_q, \sigma_q^2)$ and bias $b_q \sim \mathcal{N}(\mu_q, \sigma_q^2)$. The Gaussian prior was chosen as it allows us to specify a range of plausible values using two parameters, and to admit a relative simple approximation inference. Also, it characterizes the assumption that a priory we expect that extreme ratings would be uncommon. In this work we set $\mu_p = \mu_q = 0$ and $\sigma_p = \sigma_q = \beta = 1$.

Inference in this model is done using message passing algorithms. We implemented the model using the Infer.NET [25] framework for probabilistic graphical models. Inference was done approximately, using the Expectation Propagation (EP) algorithm [26]. EP calculates marginal distributions on a given factor graph by iteratively computing messages along edges that propagate information across the factor graph. EP runs iteratively until convergence, so its runtime is linear in the model's size, which in the case of PMPS is of size $O(|P| \cdot |Q|)$. We note that while in our dataset all observations were present (i.e. we had the rating of every participant to every question), PMPS can also handle a partial set of observations (i.e. the case where we cannot observe the response of some participants to some of the questions).

3.1 Model Selection

For a specific K the model $\mathcal{H}_K$ was built, however different K values produce different models. The task of selecting the most plausible statistical model from a set of candidates $\mathcal{H}_1, \ldots, \mathcal{H}_N$ given the data is called *model selection*.

An important aspect of model selection is that we should not compare models solely based on how well it fits the data, but also based on their simplicity.[2] In other words, a good model selection technique should be balanced, achieving a good trade-off between *goodness of fit* and *simplicity* [40]. A key advantage of a Bayesian approach is the existence of a well-accepted methodology for achieving such a trade-off.

The posterior probability, $\Pr(\mathcal{H}_K|D)$ of the model $\mathcal{H}_K$ given the data D, is given by Bayes's theorem:

$$\Pr(\mathcal{H}_K|D) = \frac{\Pr(D|\mathcal{H}_K)\Pr(\mathcal{H}_K)}{\Pr(D)}. \tag{2}$$

[2]As an example, consider a dataset with five points in 2D space. One model that has a perfect fit is to use a fourth-degree polynomial; however if we look at the points and find that they are approximately on a straight line, we will favour a much simpler linear model that simply assumes some amount of noise.

This gives the following probability ratio between model $\mathcal{H}_i$ and model $\mathcal{H}_j$ [23]:

$$\frac{\Pr(\mathcal{H}_i|D)}{\Pr(\mathcal{H}_j|D)} = \frac{\Pr(D|\mathcal{H}_i)}{\Pr(D|\mathcal{H}_j)}\,\frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_j)}. \tag{3}$$

If we have a uniform prior over models, i.e. no a priori belief that either $\mathcal{H}_i$ or $\mathcal{H}_j$ is more probable, Equation 3 simplifies to $\Pr(D|\mathcal{H}_i)/\Pr(D|\mathcal{H}_j)$, this ratio is known as the *Bayes factor* [18]. We thus wish to find the model that maximizes $\Pr(D|\mathcal{H}_K)$. The density $\Pr(D|\mathcal{H}_K)$ is obtained by integrating over the unknown parameters values in that model:

$$\Pr(D|\mathcal{H}_K) = \int_{\theta_K} \Pr(D|\mathcal{H}_K, \theta_K) \Pr(\theta_K|\mathcal{H}_K)\, \mathrm{d}\theta_K, \tag{4}$$

where θ_K are the parameters under $\mathcal{H}_K$, $\Pr(\theta_K|\mathcal{H}_K)$ is the prior density and $\Pr(D|\mathcal{H}_K, \theta_K)$ is the probability of the data given the model $\mathcal{H}_K$ with parameters θ_K. The quantity in Equation 4 is called the *evidence* for model $\mathcal{H}_K$.

Simple models tend to make precise predictions and complex models, by their nature, are capable of making a greater variety of predictions. Therefore, in the case where the data are compatible with both models, the simpler model will turn out as more probable. The dimension K that results in the highest model evidence, suggests that the latent dimension of the data is that K.

4 Results

We now present our results, produced by applying PMPS learning and alternative methods on the UK and US datasets. We first describe the datasets, then discuss the empirical evaluation and the results.

4.1 Datasets

In both UK and USA datasets, users were directly asked to give their opinion regards several issues.[3]

4.1.1 UK Dataset

The website 38DEGREES[4] [1] has conducted a poll in which users from across the UK were asked to rank 18 issues according to their priorities. The issues were: The

[3]Asking users directly for their political opinions is one route for obtaining data. An alternative is mining social network data to infer political opinions [37, 35]. These could then be correlated with other inferred user psycho-demographic traits [21, 20, 4, 36] or socio-economic perceptions [22, 29, 11] (similarly to our analysis here). We used a direct survey as it offers less noisy observations than inferred traits (and although this limited the amount of data we collected, we believe the size of our dataset is sufficient for our analysis).

[4]http://www.38degrees.org.uk/

- Having an abortion is the choice of the mother. No one else has the right to decide this for her.
- Life begins at conception. Babies are people and deserve to be protected from abortion by law.
- It is morally a role of the state to provide basic medical care for everyone.
- Immigrants help grow our economy.
- Access to guns should be severely controlled.
- Owning a gun should be a fundamental right.
- My country has the duty to bring democracy to the world.
- Alcohol and cigarettes should be heavily taxed to discourage their use.
- Education including higher education should be free to all.
- The government should extend paid maternity leave to 3 months for every working mother.

Figure 2: An example of a few questions from the questionnaire

	Immigration	Poverty	Environment	Cost of living	The EU
Climate change	-0.3535	0.01949	0.58	-0.3272	-0.2592
Immigration		-0.3583	-0.2245	0.0636	0.4757
Poverty			-0.0804	0.0669	-0.3491
Environment				-0.2956	-0.1842
Cost of living					-0.0628

Table 1: The correlation between the issues

NHS[5], Banking, Privacy, Tax dodging, The economy, Animal welfare, Climate change, Poverty, Human rights, Education, Transport, Energy, Immigration, Environment, The EU, Housing, Cost of living and Privatisation of public services.

The results of this poll were generously shared with us. The dataset consists of 38,000 records of users, where every user is identified with his/her postcode.

4.1.2 USA Dataset

Using Amazon Mechanical Turk [34], We asked 1,500 users from different states in the USA to fill a questionnaire with 56 statements. For each statement, each user was asked give his/hers opinion and select one option from the seven-point scale from disagree strongly to agree strongly.

The statement were about different topics such as: religious, abortion, gay and lesbian rights, public health care and immigration. An example of a few questions can be found in Figure 2, the complete questionnaire is available online.[6]

[5]NHS is the British National Health Service

[6]http://tinyurl.com/h7t339v

4.1.3 Correlations Between Responses to Different Questions

Unsurprisingly, when examining the issue ratings in the UK dataset we note that some issues' rankings are correlated. Table 1 shows the correlations between the ratings of various issues. For example, as can be seen in Table 1, a user that ranked *Environment* at a high place is likely to do the same for *Climate change*, and a user who ranked *Immigration* at a high place is likely to rank *Climate change* at a low place.

Similar correlations between the responses to different political stance questions also occur in the USA dataset, indicating that many political issues are inter-correlated.

4.2 Predictive Performance

The high correlation between various items indicates that it should be possible to predict the responses of a participant to some of the questions based on the way their responded to other questions. Given a specific single question $q \in Q$, we can train a linear regression model to predict how a person $p \in P$ would respond to that question given their responses to some other subset of questions $S \subset Q$. If we are interested in predicting the responses to *multiple* questions,

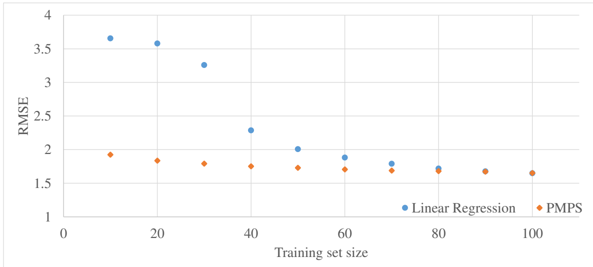


Figure 3: The prediction error (RMSE) of linear regression and PMPS.

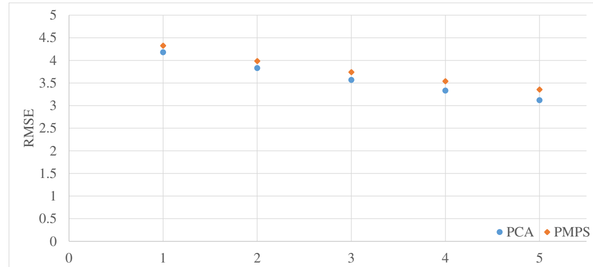


Figure 4: The error (RMSE) of PCA and PMPS against the number principal components on the UK data.

$T \subset Q$ based on the responses to some of the other questions, $S \subset Q$ (where $S \cap T = \emptyset$), we can train $|T|$ linear regression models, each taking as features the responses to the questions in S and predicting the responses to one of the questions $t_i \in T$.

An alternative approach is to use our PMPS model to generate the predictions. In this approach we learn to express questions as a posterior distribution over vectors in a low-dimensional space. Then, given a partial set of responses given by the person p to a subset S of the questions, we also obtain a representation of that person as a posterior distribution over vectors in the low dimensional space. Thus, PMPS would also produce a posterior distribution over the possible responses p would have to all the remaining questions, $Q \setminus S$ (including of course the questions in T). We can then take the mode of the posterior distribution over the responses to an unobserved question $t_i \in T$ and use them as a prediction for the response p would give to t_i.

We first designed an experiment to contrast the predictive performance of PMPS with that of linear regression models, using the USA dataset. The experiment goal was to predict the responses of users to questions in our USA dataset. In each trial we randomly select a subset $P^b \subset P$ of the participants to use for training. We let the model (either the linear regression model or PMPS) to observe the responses of these training participants to all the questions. We then select a subset of test participants $P^t \subset P$ (such that $P^b \cap P^t = \emptyset$). We also select a set $T \subset Q$ of target questions and a set of predictor questions $S \subset Q$ (where $S \cap T = \emptyset$). We first train the model using the training participants P^b, then let the model observe the responses of each test participants $p_j \in P^t$ to the target questions S (and *only* these questions). Next, we use the model to predict the responses of each test participant to each of the target questions, and examine the prediction error. In our experiments we used $|S| = 30$ predictor questions and $|T| = 26$ questions (so S and T form a partition of the question set). In our experiments we fixed $|P^t| = 20$ and we varied the number of training participants $|P^b|$ between 10 participants and 100 participants. We used 1,000 trials to measure the average prediction error of each model (linear regression or PMPS).

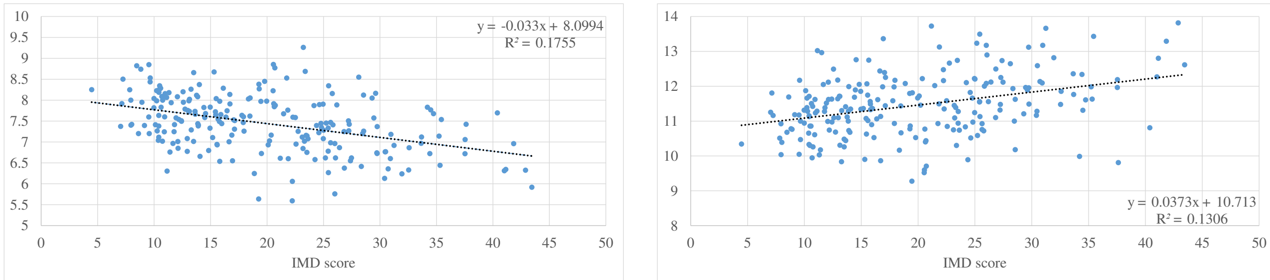
Our results regarding the predictive performance of PMPS and linear regression are given in Figure 3. The x-axis is the number of training participants, $|P^b|$, reflecting the amount of training data available to the model. The y-axis is the prediction root mean squared error (RMSE), averaged across the trials, with cross-validation. The figure shows that PMPS achieves a superior prediction quality over a linear regression method. The quality gap is especially big for treatments with less training data (fewer training participants). Figure 3 shows the results for PMPS model with $K = 6$, though other K values showed similar results and outperformed linear regression.

4.3 Latent Traits

The high correlation between the issues suggests that the response data can be compressed. For example, consider the UK dataset, where the full data for each user consists of an 18-dimensional vectors, representing that person's importance rating for each of the issues. Rather than representing the entries of the dataset as 18-dimensional vectors, with one number per issue / question, we can project them into a lower dimension using either PMPS or traditional linear techniques such as PCA. This results in "compressing" the opinions dataset, at the cost of introducing some noise to the the entries.

We note that as opposed to the predictive performance experiment in Section 4.2, the goal of using such a dimensionality reduction is a lossy compression of the data, rather then predicting some entries or responses of a user based on their responses to other questions. In particular, in order to find a person's representation as a low dimensional vector, we process *all* of their responses.

For both PMPS and PCA the projection into a low dimensional vector is revertible, i.e. we can easily interpret the low-dimensional representation back into the 18-dimensional space. For PMPS this is done by plugging the learned weights and low-dimensional vectors in the model



(a) The average rank of "Poverty" as a function of the IMD score. (b) The average rank of "The EU" as a function of the IMD score.

Figure 5: Correlation between the IMD score of LA and the average rank the issues in the LAs.

and inferring, rather than observing, the rating r of each question—this key operation is the inner product.

As the data is not perfectly reconstructed from this compression, a good measure of the noise resulting from the compression is the RMSE of the reconstructed 18-dimensional point compared to the original.

Figure 4 reports the compression RMSEs for both PCA and PMPS. Two conclusions can be drawn from the figure:

- Both PCA and PMPS can be used to characterize a *political spectrum*, which allows a tradeoff between the conciseness of peoples' description and the error in predicting their responses using this compressed description. The tradeoff between the compression and the error is very similar for these methods.
- Applying PCA on the data does not give a clear indication on the *true dimensionality* of the data: even for relatively high dimensions, the RMSE is not negligible.

As we discussed in section 3.1, one significant advantage of a Bayesian method such as PMPS over non-Bayesian methods is that they can trade-off some of the accuracy of compression for a greater simplicity of the representation, and that they allow for a rigorous method for choosing the "correct" dimensionality of the spectrum used.

4.3.1 Socio-Demographic Factors

The latent traits of a user, captured by their representation in the low dimensional space, may not correspond to an objective and observable measure of that individual. However, in the case of our political data, some correlations between ratings of different items can be explained as a result of socio-demographic influence.

Consider, for example, the participants in the UK dataset. It stands to reason that poor areas would be more concerned about issues such as poverty. To examine the correlations between rating and socio-demographic influence in the UK, we use the English Index of Multiple Deprivation (IMD) [12]. The IMD is a score that is given to every local authority (LA) in the UK. This score is based on employment, income, extent and concentration of the local authority. It ranges between 0 and 50: the higher the score, the more deprived the local authority.

There is a high correlation between the average rank of an issue in a specific LA and the IMD score of the LA. As, for example, it is shown in Figure 5, the average rank of "Poverty" (5a) is generally lower (more important) in LA's with high IMD, than in LA's with low IMD; and the average rank of "The EU" (5b) is generally higher in LA's with low IMD than in LA's with high IMD. Thus, LA's that ranked "Poverty" in a relatively high location are likely to rank "The EU" in a relatively low location and vice versa. Therefore, observing the IMD score of the user's LA could give us information about the ranks of the issues.

This analysis illustrates that it may be possible to capture a lot of the variability in People's responses to political questions by considering some of their objective traits, such as poverty.

4.4 Model Dimensionality

Our main goal is determining the true dimensionality of the political spectrum required to accurately represent peoples' stances regarding a wide range of political issues. To this end, we applied PMPS learning to our UK dataset and USA dataset. In both cases we examined the model evidence of every dimension so as to reveal the dimensionality of the datasets.

4.4.1 Dimensionality in the UK Dataset

When analyzing the UK dataset, we applied the PMPS model to the *entire UK data*, as well as the data from *specific UK local authorities*. As there is a high correlation between the IMD score of LA and the average rank of an issue, one of the latent dimension, in the PMPS model might be related to the IMD score, and the dimension of the data

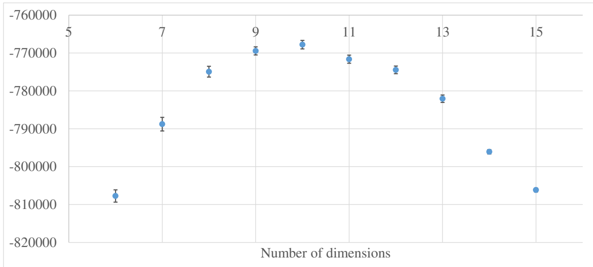


Figure 6: Log model evidence as a function of the dimension, on the data from UK.

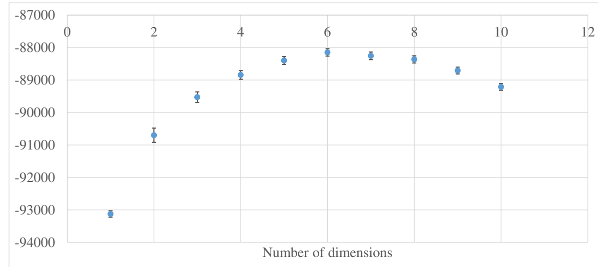


Figure 7: Log model evidence as a function of the dimension, on the data from USA.

is likely to be lower than 18.[7] The log of the model evidence as a function of the dimension can be found in Figure 6. The model in which the latent dimension was 10 resulted in the highest model evidence, and therefore the data suggests that the dimension is 10. We applied the same technique for every LA. In this experiment the dimension ranged from 7 (Metropolitan Borough of Wigan) to 11 (Rossendale District), with average of 9.717 and standard deviation of 0.678. Thus, the dimension of the data remains relatively homogeneous across local authorities.

4.4.2 Dimensionality in the USA Dataset

In the 114th United States Congress, out of 435 seats 431 are occupied by members of the Democrat and Republican parties. That is, the vast majority of the people in the US are represented by the two parties. For comparison, in the Parliament of the United Kingdom, there are 11 parties.

Historically, the Democratic party supports gun control laws, keeps elective abortions legal, and tends to favor equal rights for gay and lesbian couples. In contrast, the Republican Party opposes gun control laws, and the Republican party's agenda states that abortions should not be legal and marriage should be between a man and a woman. Hence, it appears that a single left-right axis could describe the two parties.

However, it is not clear why, for example, a person who supports abortion is likely to opposes state involvement with religious institutions. Therefore, we attempted to investigate whether the political dimension in the US is indeed lower than in the UK and, whether two parties can truly represent the American people. Alternatively, it could be the case that the election system in the USA results in a two party system, even though more parties are actually needed to truly represent the electors.

In order to find the dimensionality of the data, we applied PMPS learning. The log of the model evidence as a function of the dimension can be found in Figure 7. The model in which the latent dimension was 6 resulted in the highest model evidence, and therefore data suggested that the dimension is 6, while the same technique suggested that the dimension within UK is 10. In addition, we applying PMPS learning on the data from different states. Unlike the data from UK, the results were less homogeneous across states, and range from 4 in Texas and California, to 6 in Michigan, New York and Florida.

5 Related Work

While the traditional political spectrum is a simple left-right axis, it has been doubted whether ordinary citizens actually use the specific ideology associated with this representation (see, e.g., [8]). Many Political scientists have suggested more complex, multidimensional representations of the political spectrum. One of the earliest work is by Ferguson [10], who suggested that peoples' positions with regard to 10 broad topics were influenced by 3 broad underlying dimensions: Religionism (with issues such as evolution, birth control and God), Humanitarianism (with issues such as war, capital punishment and treatment of criminals) and Nationalism (censorship, law patriotism and communisim).

Christie, and Meltzer [7] suggested a four dimensions diagram: fabianism to radicalism, fascism to anarchism, conservatism to social democracy and capitalist individualism to state communism. A similar research has been done by Eysenck [9] in the United Kingdom and in Germany. The research identified two independent principles: *Radicalism* (R) and *Tender-Mindedess* (T). In both countries, all but one attitude were found to have the same coordinates on the R-T Cartesian coordinate system. In Sweden, Husen showed a similar pattern [15].

Our method for determining the dimensionality of a political spectrum given a dataset is a Bayesian one, relying on a dimensionality reduction Probabilistic Graphical Model. Graphical models [27] have been widely studied

[7] Obviously, one needs to know the ranks of 17 issues in order to give an accurate prediction on the rank of the remaining issues. However, the PMPS model is not aware to the fact that for every user every rank appears only once.

in the context of AI. For example, Porteous et al. [28] use graphical model for Bayesian probabilistic matrix factorization. Schmidt et al. [30] dealt with learning the structure of undirected graphical using L1 regression. Bachrach et al. [5] presented a graphical model for inferring the correct answers, difficulty levels of questions and ability levels of participants in multiple problem domains. A line of work has considered Bayesian methods and matrix factorization techniques for collaborative filtering based recommender systems [14, 13, 16, 32], which also capture peoples' ratings of various items. Our model is similar to the Matchbox recommendation system [33], in which users and items are mapped into a low-dimensional 'trait space'.

We used a rigorous and theoretically justified method for dimensionality selection. Alternative model selection techniques such as AIC [3] and BIC [31] have been previously for other domains. These technique also trade-off bias and variance, using the likelihood function, number of parameters of the model and the number of sampled data.

6 Conclusions

We proposed an approach for choosing the optimal dimensionality of the political spectrum, based on a dataset of responses of participants regarding political stands. Our method uses a probabilistic graphical model for dimensionality reduction, allowing us to express the political spectrum dimensionality selection problem as a Bayesian model selection problem, which we solve by choosing the dimensionality of the model with maximal evidence.

We applied the model on two types of datasets. The UK dataset contains participants ranking regarding many political issues, whereas in the US dataset participants rate their degree of agreement with many sentences representing a political stand. Our model indicates that for both datasets, there are correlations in the data regarding seemingly unrelated political issues, allowing for a more concise representation of peoples' political stand than the naive encoding of their responses to all questions. Further, our analysis of the UK dataset indicates that socio-demographic factors correlate with political opinions. This allows predicting political stands based on such socio-demographic factors.

Despite these correlations between responses to different questions (or between socio-demographic factors and these responses), our model indicates that a "left-right" political spectrum, or even a two dimensional spectrum, are far too simplistic, and insufficient to represent peoples' political opinion. Our model's choice for the optimal dimensionality is 10 dimensions for the UK dataset, and 6 for US dataset. Interestingly, the optimal dimensionality for the political spectrum differs across states in the USA.

Many questions are left open for future research. First, the political dimensions found by our model are the result of the feature extraction during the dimensionality reduction. Could these dimensions be interpreted in a human understandable form? Second, would alternative Bayesian models for dimensionality reduction achieve a lower error, and perhaps result in a different choice of dimensionality for the political spectrum, or is this dimensionality an inherent property of the data? Finally, our results indicated correlations between socio-demographic features of participants and their political opinions. To what degree of accuracy is it possible to predict demographic traits of people based solely on their political stand?

References

[1] 38 Degrees. 38 degrees — people. power. change., 2014.

[2] D. Agarwal and B.-C. Chen. flda: matrix factorization through latent dirichlet allocation. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 91–100. ACM, 2010.

[3] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer, 1998.

[4] Y. Bachrach. Human judgments in hiring decisions based on online social network profiles. In *DSAA*, pages 1–10. IEEE, 2015.

[5] Y. Bachrach, T. Graepel, T. Minka, and J. Guiver. How to grade a test without knowing the answers—a bayesian graphical model for adaptive crowdsourcing and aptitude testing. *ICML*, 2012.

[6] Y. Bachrach, O. Lev, Y. Lewenberg, and Y. Zick. Tracking performance and forming study groups for prep courses using probabilistic graphical models. In *AAMAS*, 2016.

[7] S. Christie and A. Meltzer. *The floodgates of anarchy*. PM Press, 1970.

[8] P. E. Converse. The nature of belief systems in mass publics (1964). *Critical Review*, 18(1-3):1–74, 1964.

[9] H. J. Eysenck. *Sense and nonsense in psychology*. Penguin Books, 1963.

[10] L. W. Ferguson. The stability of the primary social attitudes: I. religionism and humanitarianism. *The Journal of Psychology*, 12(2):283–288, 1941.

[11] M. I. Gorinova, Y. Lewenberg, Y. Bachrach, A. Kalaitzis, M. Fagan, D. Carignan, and N. Gautam. Predicting gaming related properties from twitter accounts. In *AAAI*, 2016.

[12] GOV.UK. English indices of deprivation 2010: local authority summaries, 2010.

[13] T. Hofmann. Latent semantic models for collaborative filtering. *ACM Trans. Inf. Syst.*, 22(1):89–115, Jan. 2004.

[14] T. Hofmann and J. Puzicha. Latent class models for collaborative filtering. In *IJCAI*, volume 99, pages 688–693, 1999.

[15] T. Husen. Matningar av intressen och attityder. *Personalen och Foretaget*, pages 80–87, 1951.

[16] R. Jin, L. Si, and C. Zhai. Preference-based graphic models for collaborative filtering. In *UAI*, pages 329–336, 2002.

[17] J. T. Jost, C. M. Federico, and J. L. Napier. Political ideology: Its structure, functions, and elective affinities. *Annual review of psychology*, 60:307–337, 2009.

[18] R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

[19] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[20] M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell, and T. Graepel. Manifestations of user personality in website choice and behaviour on online social networks. *Machine learning*, 95(3):357–380, 2014.

[21] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.

[22] Y. Lewenberg, Y. Bachrach, and S. Volkova. Using emotions to predict user interest areas in online social networks. In *DSAA*, pages 1–10. IEEE, 2015.

[23] D. J. MacKay. *Information theory, inference, and learning algorithms*, volume 7. Citeseer, 2003.

[24] W. S. Maddox and S. A. Lilie. *Beyond liberal and conservative: Reassessing the political spectrum*. Cato Institute, 1984.

[25] T. Minka, J. Winn, J. Guiver, S. Webster, Y. Zaykov, B. Yangel, A. Spengler, and J. Bronskill. Infer.NET 2.6, 2014.

[26] T. P. Minka. Expectation propagation for approximate bayesian inference. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 362–369. Morgan Kaufmann Publishers Inc., 2001.

[27] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[28] I. Porteous, A. U. Asuncion, and M. Welling. Bayesian matrix factorization with side information and dirichlet process mixtures. In *AAAI*, 2010.

[29] D. Preoţiuc-Pietro, S. Volkova, V. Lampos, Y. Bachrach, and N. Aletras. Studying user income through language, behaviour and affect in social media. *PloS one*, 10(9):e0138717, 2015.

[30] M. Schmidt, A. Niculescu-Mizil, K. Murphy, et al. Learning graphical model structure using l1-regularization paths. In *AAAI*, volume 7, pages 1278–1283, 2007.

[31] G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[32] L. Si and R. Jin. Flexible mixture model for collaborative filtering. In *ICML*, volume 3, pages 704–711, 2003.

[33] D. H. Stern, R. Herbrich, and T. Graepel. Matchbox: large scale online bayesian recommendations. In *WWW*, pages 111–120. ACM, 2009.

[34] A. M. Turk. Amazon mechanical turk, 2014.

[35] S. Volkova and Y. Bachrach. On predicting sociodemographic traits and emotions from communications in social networks and their implications to online self-disclosure. *Cyberpsychology, Behavior, and Social Networking*, 18(12):726–736, 2015.

[36] S. Volkova, Y. Bachrach, M. Armstrong, and V. Sharma. Inferring latent user properties from texts published in social media. In *AAAI*, pages 4296–4297, 2015.

[37] S. Volkova, G. Coppersmith, and B. Van Durme. Inferring user political preferences from streaming communications. In *ACL (1)*, pages 186–196, 2014.

[38] K. Von Beyme. *Political parties in Western democracies*. Gower Publishing Company, Limited, 1985.

[39] A. Ware. *Political parties and party systems*, volume 9. Oxford University Press Oxford, 1996.

[40] E. Wit, E. v. d. Heuvel, and J.-W. Romeijn. all models are wrong...: an introduction to model uncertainty. *Statistica Neerlandica*, 66(3):217–236, 2012.