The Information Society: An International Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utis20

Expanding the Possibilities of Deliberation: The Use of Data Mining for Strengthening Democracy with an Application to Education Reform

Juan D. Velásquez [a] & Pablo González [b]

[a] Department of Industrial Engineering, School of Engineering and Science , University of Chile , Santiago, Región Metropolitana, Chile

[b] Center for Applied Economics, Department of Industrial Engineering, School of Engineering and Science and Center for Advanced Research in Education , University of Chile , Santiago, Región Metropolitana, Chile

Published online: 19 Jan 2010.

To cite this article: Juan D. Velásquez & Pablo González (2010) Expanding the Possibilities of Deliberation: The Use of Data Mining for Strengthening Democracy with an Application to Education Reform, The Information Society: An International Journal, 26:1, 1-16, DOI: 10.1080/01972240903423329

To link to this article: http://dx.doi.org/10.1080/01972240903423329

The Information Society, 26: 1–16, 2010

ISSN: 0197-2243 print / 1087-6537 online
DOI: 10.1080/01972240903423329

Routledge
Taylor & Francis Group

Expanding the Possibilities of Deliberation: The Use of Data Mining for Strengthening Democracy with an Application to Education Reform

Juan D. Velásquez

Department of Industrial Engineering, School of Engineering and Science, University of Chile, Santiago, Región Metropolitana, Chile

Pablo González

Center for Applied Economics, Department of Industrial Engineering, School of Engineering and Science and Center for Advanced Research in Education, University of Chile, Santiago, Región Metropolitana, Chile

Deliberation is important for strengthening democracies and enhancing the legitimacy of public policy. However, deliberation has been constrained by limits of time, space, and human capacities for listening and processing information. In this article, the authors discuss a new technology-based tool and show how it can help to partially remove these constraints. Although the Internet already provides the means to deliberate without the need to meet at the same place and time, its conjunction with data mining solves the "large numbers deliberation dilemma" that arises when large amounts of data have to be processed. The author's proposal adapts particular data-mining techniques, which simulate the learning process of a human brain with almost infinite relational capacities. The methodology was applied in a real-world case of Chilean education reform and demonstrated its potential effectiveness.

Keywords data mining, deliberation, education reform, participation, technological tools

Democratic nations are facing what has been named the *crisis of the state*, the *democratic deficit*, or the *crisis of representative democracy*. The symptoms include reduced participation in electoral voting, declining affiliation with unions and political parties, and the emergence of new forms of collective action (e.g., civic unrest, boycotts, and protests). Citizen confidence in representative democracy has been eroded. In this context, national, regional, and local governments as well as citizen organizations are experiencing different mechanisms to increase people's involvement in public decision making.

As expected, new social technologies have been developed to facilitate public participation (Dahl 1999). A key issue in participation exercises is deliberation. Although deliberation is not unanimously valued in the history of political philosophy,[1] nowadays most authors consider it the cornerstone of democracy. Examples of new deliberative participation processes include consensus conferences, citizen juries, permanent citizen panels, deliberative opinion polls, formal public discussion procedures, scenario workshops, voting conferences, and so on. All these procedures are expected to improve the functioning of democracies and the legitimacy of decisions and to provide a deeper definition of the *general interest*. However, all these new social techniques of participation involve a limit in the number of persons taking part in the deliberative exercise, which implies that citizens are "represented" by other members of the public (using different forms of selection or election of these representatives).

The apparent tension between the pair deliberation-representation, on the one hand, and direct democracy, on the other, is a long-lasting issue in political sciences. Two aspects of deliberation seem to be better achieved through

Received 28 August 2008; accepted 21 May 2009.

The authors thank the Millennium Scientific Institute on Complex Engineering Systems and the project CIE-05 PBCT-Conicyt, which have partially funded this work.

Address correspondence to Juan D. Velásquez, Department of Industrial Engineering, School of Engineering and Science, University of Chile, Av. República 701, Office 301, Santiago, Región Metropolitana, 837-0720, Chile. E-mail: jvelasqu@dii.uchile.cl

"representation" in a way that is definitely opposed to direct democracy: careful examination of issues and listening to others' perspectives (Roberts, 2004). The first requires, ideally, gathering all the relevant information on an issue and establishing the relationships and causalities between the different elements involved and the consequences and trade-offs associated with different policies. The second requires small numbers almost by definition, as, historically, it has not been possible to listen to more than one individual at a time. In fact, even in small representative bodies such as parliaments, for each particular matter only a few individuals speak and the vast majority listens.

The limitations of space, time, and human capacities to process information and to communicate with each other are so obvious that nobody has so far, to our knowledge, attempted to overcome them simultaneously. How can large numbers of people deliberate while retaining the ideal of deliberation that is best enacted in small groups (Cleveland 1975; Roberts 2004)? Let us call this problem the *large-numbers deliberation dilemma*.

Since the 1970s, new social technologies have been designed and applied to solve the large-numbers deliberation dilemma, enabling ever more citizens to be directly involved in large-group problem solving and decision making (Rautenfeld 2005; Roberts 2004; Fishkin 1991). Citizen collaborations can now accommodate thousands of people at a time. However, so far, these procedures are constrained by time and space, as first people must meet in both dimensions, and second, people are limited by human processing and communication capacities. Few initiatives have been launched to use information and communication technologies (ICT) to combat the constraints of time, space, and human capacities simultaneously. For instance, some applications of ICT, such as Internet voting or teledemocracy, appear to have restricted the possibility of deliberation to a few individuals (Poster 1997). Vedel (2006) in a review of electronic democracy concludes: "The medium certainly embodies unprecedented potential but the transformation of a utopian 'strong democracy' into practical systems remains a virtual vision waiting to materialize" (Vedel 2006, 234).

The key problem is not the confluence of people simultaneously in the time and the space, as the Internet can sort out that problem, but rather that we have not used the potential mechanisms for processing massive deliberations and to obtain certain "outcomes" useful for policies. So far, citizens speaking at the same time produce noise, not information, and millions exchanging their opinions on public issues will only discredit the possibility of participation as nobody will be able to effectively listen. Solving the large-numbers deliberation dilemma requires moving from the "noise" of crowds to the "voice" of the people.

This article proposes a mechanism for extracting significant information from a deliberation process, which could involve many individuals and an ever greater number of relevant opinions. It is our contention that this mechanism expands the possibilities for deliberation by using computational capacities to process information, enlarging human capabilities. In large participation processes, one of the main problems that require solution is the collection and processing of substantial and varied data, which makes it difficult to find relevant information. The solution proposed in this article is an application of data mining (Berry and Linoff 1997).[2] Data-mining algorithms use data in traditional formats as inputs—integers, floats, strings, and characters. When the data come in other formats, the algorithms should be adapted, and in some cases this might require the creation of new data-processing techniques. This can give rise to a new branch in the data-mining taxonomy, directly related to the data format to be processed, as in our case, where the data are deliberations in a textual form and the algorithms being applied belong to a text mining subset (Delgado 1999).

Among the many different data mining techniques, we use a clustering algorithm (Fung 2001) to automatically analyze opinions by grouping them into clusters with common textual content that represent the concerns of a large number of individuals (Ríos et al. 2006). The output of this process is information organized into clusters that can be analyzed by an expert in the phenomena under study (Theodorisis 1999) to identify new knowledge and, in particular, consensus around a given theme (Velásquez and Palade 2007).

In theory, the tool can be applied to a large participation experiment, in which millions of citizens might deliberate and give their opinion in different relevant political issues. Aside from the data mining tool and the support of expert opinion, this large, participatory exercise requires a careful preparation of the information that will be distributed to participants and of the space where deliberation will actually occur (whether physical, virtual, or both), as has been done in previous experiences such as Deliberation Day or Deliberative Polls (Ackerman and Fishkin 2002; Fishkin et al. 2004). In this sense, our methodology is not alternative to actual deliberative practices but extends their possibilities by removing constraints of time, space, and human information processing capabilities.

It is important to stress that the method is not intended to give feedback to the deliberation process: It requires the entire deliberation process as an input.[3] Therefore, it is mostly useful for policymakers and analysts, as it allows the extraction of patterns in the data that might not be represented in the outputs or conclusions of the deliberation process or it might provide another mechanism to obtain useful information for policy purposes from a deliberation process not intended to arrive at conclusions, as is the

case for the conversations that occur in the Internet every day.

Aside from presenting the technological tool, we describe its first application to the real world, in a country with a long-standing tradition of centralization, vertical decision making, and absence of citizen participation (Veliz 1980; UNDP 2002). The application was used to process the Citizen Dialogs, a deliberation process conducted in Chile in 2004 that integrated advances in social technologies, where thousands of opinions were expressed at physical meetings and via the Internet. This was the first application of this technique, facilitated by ICT and developed for private business, to the public sphere.

The problem was recognized long ago in the political literature. For instance, an old Chinese tale described the preparation of a young prince before assuming government duties. The master trainer sent the prince to live in the forest for one year and after that period, to return and tell him what he had listened to. After one year, the prince returned to the wise advisor and told him: "Master, I have learned a lot about the forest, I have listened to many sounds, thousands, for instance the owl calling, the fox hunting . . . "

The master suddenly stopped him: "It is not enough," he said, "those sounds can be listened to by anyone, but you don't need to listen to them all, only the important ones. You will return to the forest for one more year, and, this time, pay attention to what are you listening to."

Again in the forest, at the end of his journey, the prince stopped near a river, sat down in the grass, and in lotus position, he meditated. After a few minutes of tranquility, he heard imperceptible sounds from his heart, from the depth of the forest, from the life around him. The young prince returned to the master and told him about the experience in the forest. The master looked at the prince, very pleased, and said, "Congratulations, young prince, you are ready now, because you have understood the most important skill for an emperor, to listen to what is not always evident but is important. This is the essence of the voice of your people. You must learn to listen to the people's voice, which is often imperceptible, but when this voice does become perceptible, it is too late to fix things."

In the Citizen Dialogs, we developed the technological equivalent to the prince's ear.

The voice of the people is inside each dialog, but the dialogs are made up of hundreds of people giving their opinions; each is often imperceptible or unclear when using common methods of listening or interpretation and so "lost in the forest."

Having explained its potential, it is important to note that the "good use" of data mining applied to the domain of deliberation is not straightforward. It might expand human freedoms in the social and political arena or it can be used for control and to gather information on our private lives. For this potentially powerful instrument to contribute to democracy and return power to the people, the problem is no longer the technology but how that power will be used.

This article is structured as follows. First, we provide an overview of the technology. Then, we discuss the adaptation of the technology used for mining the citizen dialogs and extracting consensus opinions and its first application to the real world. Thereafter, we examine the advances made possible by our approach and how it might contribute to strengthening the legitimacy of modern democracies. Lastly, we offer our conclusions.

DEVELOPING A TECHNOLOGICAL TOOL

Data Mining Theory

Data mining is best defined by the techniques and algorithms it uses to analyze "(often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner" (Fayyad et al. 1996, 6). These techniques are a subset of a broader type known as *knowledge discovery in databases* (KDD), defined as "the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad et al. 1996, 4). The expert in the material under study is responsible for deciding whether the extracted patterns are important or not. Furthermore, her opinion is crucial for improving the entire KDD process.

Data mining uses algorithms, methods, and techniques developed for extracting significant patterns from data sets. The identification of these patterns becomes the basis for creating new knowledge about the solution for a given problem (Velásquez and Palade 2007). Data-mining techniques can be classified mainly as statistical, genetic algorithms, association rules, sequence patterns, decision trees, classification, clustering, and artificial neural networks (ANN) (Goebel and Le Gruenwald 1999). In this study, we applied a specific ANN, which by construction and operation is considered a clustering algorithm, namely, the self-organizing feature map (SOFM), also known as the Kohonen neural network (Kohonen 1987).

Clustering is a technique commonly applied to analyze human behavior. Human beings, from the beginning of time, have formed communities with common characteristics. The purpose of clustering algorithms is to discover intrinsic patterns and from them identify common preferences, similar behaviors, distinct groups, etc., understanding their actual behavior or even predicting their future activities (Velásquez and Palade 2008).

ANN have a special place in data-mining research as they have a powerful capacity to extract hidden or unknown patterns from an apparently unrelated set of data

and functional approximations (Tickle et al. 1998). Originally developed by McCulloch and Pitts (1943), ANN is a model for simulating the human brain neural network with a model that describes how the neurons might work, using a simple neural network based on circuits.

Processing Texts Using an ANN Algorithm

Let us summarize the technique, to allow the nontechnical reader to skip the sections *From Dialogs to Vectors*, *Self-Organizing Feature Map*, and *Reverse Cluster Analysis*. SOFM receives as inputs the opinions expressed in the dialogs in a format that allows processing in a computer. Then it performs comparisons between all opinions to group them by content. The output of this comparison is a set of clusters that have two key characteristics: Opinions in the same cluster have the same or very similar content, and opinions in different clusters are completely dissimilar—they have no relationship. After the clusters have been formed, the centroid of each cluster is analyzed and used to name the cluster. Each *centroid* represents the most representative opinion in each cluster, the "consensus" opinion. In other words, it is the opinion that is closest in content to all other opinions in the cluster. Afterward, the remaining opinions in each cluster, which are not centroids, are used to complement the information embodied in the consensus opinion. This process provides important information from each cluster. The information represents the essence of the opinions expressed during the deliberation process.

If extracting consensus opinions from a small deliberative group can be a difficult task, which might take a long time, it is impossible for human beings to process thousands of opinions. This is where the use of technology is required (of course, it is not the only aspect where the application of technology might contribute to improving the deliberative processes, as we discuss in the conclusion). We use a data-mining algorithm to process a large number of opinions, simulating, in a very simple way, the learning process in the human brain during deliberation and getting significant patterns to extract consensus opinions.

Although several models for representing the biological neurons have been produced over the last decade, only SOFM has come close to simulating the biological learning process. The SOFM is an unsupervised training algorithm that represents the result of a vector quantification process that uses a set of high-dimensional input vectors and maps them in an ordered fashion to a two-dimensional space. The process simulates what happens in the human brain cortex as it learns. Several neurophysiologic experiments have shown evidence about the way that the cortex self-organizes during a learning process. In essence, the experiments show that the neurons that respond to similar features of sensory inputs are located near one another (Freeman and Skapura 1991). This is the basic training algorithm that SOFM performs, as explained later.

From Dialogs to Vectors

Before applying any data-mining algorithm, one must prepare the data to be processed with a cleaning process. The application of filters improves data quality and reduces noise. Data preparation finishes with transformation of the inputs into feature vectors (i.e., data structures representing the intrinsic characteristics of the data to be analyzed).

The creation of the feature vector is a vital step in the knowledge extraction process from a large collection of data. If the vector does not reflect a real consistency with the intrinsic characteristics of the phenomenon under study, the results from applying a pattern extraction tool, such as SOFM, can be uncertain, with a high quantity of noise, unusable for the identification of new knowledge. In other words: "garbage in, garbage out."

The feature vector is the minimum data repository to which the pattern extraction tool is applied. In the case of SOFM, a database (generally as a stream with these vectors) is created for training proposes.

Data-mining algorithms receive the feature vectors as numerical representation, which allows the application of metrics for comparing, correlating, projecting, and so on. In the case of SOFM, because a similarity measure is used for comparing and grouping vectors with similar characteristics, it is quite essential that the feature vector's components are numeric values. It follows that it is necessary to set out the texts as numeric representations.

Then, to work with numeric feature vectors, the citizen opinions collected during the deliberation processes (citizen dialogs) must be transformed into a digital document. As many of the opinions were collected electronically, for example, in a Web-based electronic forum, only those opinions written by hand or expressed by voice needed to be digitized. Next, each document is transformed into a numeric representation (Aas and Eikvil 1999). The *vector space model* (VSM) is applied by assigning to each word a numeric weight that represents its relative importance in the document and the entire set of documents of the data source. In this way, each citizen's opinion becomes a numeric vector, creating a set of vectors that are used as the input for the SOFM.

The method works as follows. Let Q be the number of documents to be processed (figure 1, part A). Let R be the number of different words in the entire set of documents. This parameter could be huge, demanding a lot of computer resources for processing the documents. However, it is important to remember that the words in a document have different relevance and characteristics. From the point of view of meaning, articles, prepositions, and

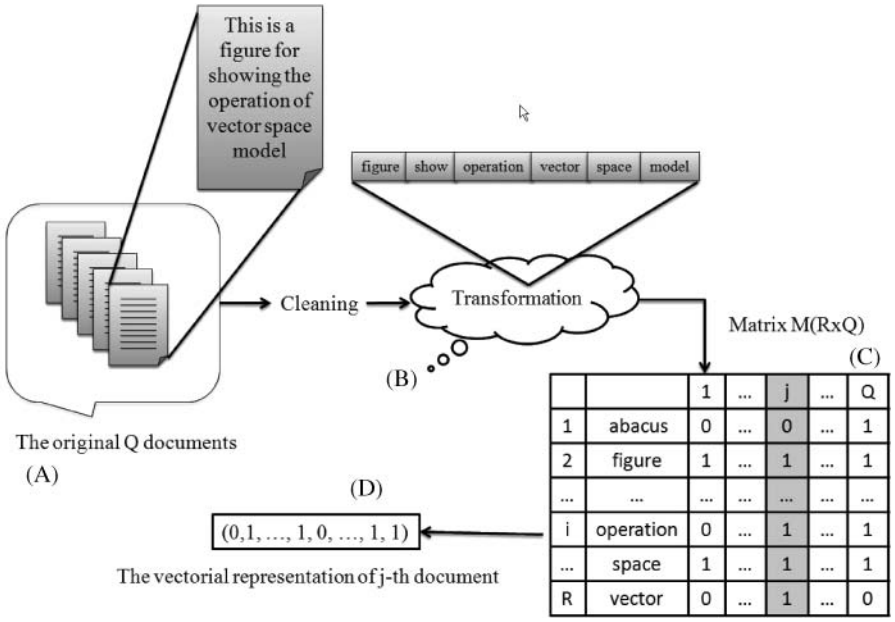

FIG. 1. An example of the vector space model operation.

conjunctions are considered stop words, and should be cleaned. Some words have the same meaning (synonymous), and this allows a reduction of the word set. Finally, several words can have the same root meaning, for instance "write," "written," and "wrote." In that case, a stemming process is applied, reducing again the entire set of words considered in the VSM. The described preprocessing task is performed in the cleaning stage in figure 1.

For each document in the original set, we have a compact representation by using its essential words (figure 1, part B). In the transformation stage, each document is transformed into a vector with numeric values as follow; let **M** be the matrix with dimensions $R \times Q$ (figure 1, part C). A simple representation of the j-th document in **M** is made by associating a weight "1" in the position (i, j) if the i-th word appears in the j-th document, and "0" otherwise. With this process, the j-th document is represented as a vector of "0" and "1" (figure 1, part D).

A more refined expression in the VSM relates the word in a particular document and the entire set of documents. The number of occurrences of the i-th word in the j-th documents and the total number of times the i-th word appears in the entire set of documents are combined in a formula that expresses the relative weight of the word in the whole set. There are several versions that show this relationship. A general version is known astext frequency $\times$ inverse document frequency (TF $\times$ IDF), because it combines the word occurrence in a document (TF) with the inverse word frequency in the entire set of documents (IDF) (for more explanation, see Aas and Eikvil 1999).

When the cleaning and preprocessing tasks are finished, the VSM is applied to the documents, setting out their vector representation. Now a simple operation such as vector comparison using a similarity measure can be performed.

The final result of applying the VSM is the creation of the matrix **M** where each column is the vector representation of a particular document, that is, the column j is the vector representation of the j-th document in the entire set considered for applying the VSM. By using a similarity measure to compare these vectors, if two documents are totally identical, then the similarity between them is "1." Otherwise it is a value between "0" and "1," being "0" in the case of two totally different documents.

The VSM is a simple method to represent documents as vectors with numeric values. However, some semantic problems remain. As VSM is a term representation it is possible that two different documents containing similar terms but with different style or meaning are considered to be the same; in some languages, like Spanish, this is highly probable. To avoid these issues, semantic models have to be applied as part of preprocessing. This makes the VSM representation more complex, with the result that the pattern extraction process can become inefficient (Ríos et al. 2006).

Self-Organizing Feature Map

The SOFM defines a mapping from the input n-dimensional data space onto a regular two-dimensional array of nodes or neurons. Each neuron is an n-dimensional vector, with the components as the coupling factors

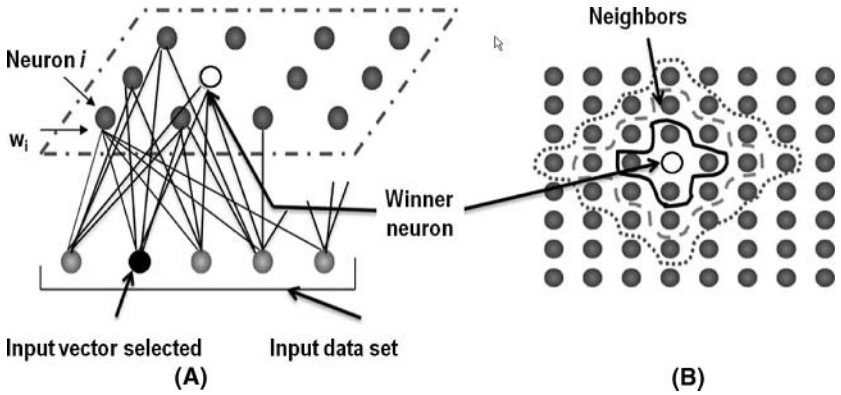

FIG. 2. SOFM training algorithm.

weights. By definition all the neurons receive the same input dimension, also called feature, at a given instant. In the SOFM, the process is by competitive learning—the set of input vectors are presented to the network, which uses a metric to determine the most similar neuron (i.e., the center of excitation, the winner neuron). The winner neuron is defined as the best matching with input vector in the whole network; it follows that the input vector is equivalent to the training vector.

The initial values of the coupling factors among the neurons of the network are randomly set. Next the neighbor nodes of the winner neuron are activated (updated) to "learn" the same sample, by using a weighted updating rule, closing their content to the current training vector, through the following process (Lin et al. 1991):

- Select an input vector from the whole input data set (figure 2, part A).
- Find in the entire set of neurons, the node whose weights are closer to the input vector. This node is called winner neuron (figure 2, part A).
- Update the winner neuron's weights making it closer to the input vector (figure 2, part B).
- In the same way, the weights of the winner neuron's neighbors are updated (figure 2, part B).

Figure 2, part A shows the input data set as a group of feature vectors, whose content represent the characteristics of the phenomenon under study. A numeric value is set for each variable in all the vector components. The feature vector definition is crucial to the clustering process: On this depend cluster quality and the convergence process. For instance, if the vector contains irrelevant values, the results may contain high levels of noise.

Another important issue is the measure used to compare neurons and input vectors in the input data set. It can be a similarity or distortion measure and must reflect what is relevant to the clustering exploration process. For instance, if we are interested in analyzing the user browsing behavior in a Web site, then the similarity metric or measure should consider a user's motivation when visiting a Web page at a given site. This method could involve the textual content of different pages, the sequence of pages visited, the time spent in each page, and so on, with the objective of extracting significant patterns about user browsing behavior at and in the site. Sometimes this measure must be tailored to the material, for it is clear that in many cases a simple similarity measure like Euclidean distance (comparing two vectors by the distance that projects the straight lines between them) could be enough.

Composition is another important component of the similarity measure. Because the SOFM training process uses considerable computer resources, a complex similarity measure could be counterproductive, increasing the computer processing time by several orders of magnitude.

The clustering process allows identifying patterns about the phenomenon under study through the analysis of cluster's composition. However, some clusters do not contain patterns with relevant information for analyzing the problem, so accept-reject criteria have to be developed to filter the extracted clusters. At this point it is advisable to be able to rely on the assistance of an expert in the phenomenon under study to help provide informed opinion about the filtering process. Although an expert might be considered subjective, empirical experience has shown that it is a good alternative to filter the clusters without relevant information.

Finally, the SOFM's output are the result of the neuron training, or more specifically, the training-set mapped in the vectors that represent each neuron. As the winner frequency is maintained for each neuron, it is possible to make a graph using neurons and their respective winner frequencies, which allows a visualization of cluster formation.

Reverse Cluster Analysis

Finally, if, after applying SOFM, there are groups of neurons with similar characteristics, then some of them will

be the best winners, showing a clear cluster definition, each with its respective centroid (a pattern extracted in the training process).

The centroid and its neighbor neurons (see figure 2, part B) form the structure that contains the results of document comparisons with vector representation; that is, the result is a set of vectors with similar characteristics and represents a group of texts whose meaning is close. However, it is impossible to obtain the original document's text from vector content only. The problem now is how to reconstruct a document from an expression with numeric values. One possibility is to maintain a data structure, for instance, a matrix or a simple linked list, with information about the input data vectors, winner neurons and neighbors. However, by construction, in SOFM both winner neuron and neighbors change their weights during the training process. Then we must identify which texts in the original data set are related with the winner neuron's neighbors in the SOFM for updating the information in the data structure. This is not a trivial step at all and consumes considerable computer resources.

An alternative method is to identify which original texts are related to the centroid and neighbors in each extracted cluster. In fact, with the application of the VSM previously discussed, each document is represented by a vector of numeric values and stored in a matrix's column, thereby preserving the column-document relation. Then it is possible to perform a reverse process as follows.

For each vector in a cluster, using the same similarity measure in the SOFM training process, the matrix is automatically reviewed column by column until the closest column to the cluster's vector is found; the next step is to find the original document. This process is also called reverse cluster analysis (RCA) (Ríos et al. 2006).

By using RCA, a reject-accept criterion can be applied to the extracted clusters. In fact, because for each cluster we still have the original nearest dialog, if the group of dialogs related to a cluster doesn't share a common theme, then the cluster is rejected. Otherwise, the cluster is accepted for review by the outside expert, who will determine the cluster's potential contribution for understanding the phenomenon under examination.

Finally, for each cluster, the SOFM's end user can read the set of original documents that better represent the vectors in each cluster and prepare a summary with this information.

APPLICATION TO EDUCATION REFORM

Data Sources

The earliest applications of this technology to public policy were called Citizen Dialogs, conducted in Chile in 2004. These Citizen Dialogs were essentially a consultation process, attempting to obtain input from the different stakeholders on their views about what were the key problems for improving the quality of primary and secondary education. The dialogs were structured in three building blocks. The first building block consisted of traditional focus groups and interviews, which helped to define the exercise's objectives with greater clarity and to identify the main issues that concern Chile's educational quality. In total, 320 individuals participated in these small workshops in five cities[4] and 31 interviews were conducted. The results of these focus groups and interviews were used as inputs for the other two building blocks of the process.

The second building block was the "physical" citizen dialogs. These were conducted in five of Chile's most populated cities[5] and were attended by a total of 1,165 participants: 185 students, 187 parents, 276 representatives of parent organizations, 96 school principals, 296 teachers, 37 experts, 10 representatives of the teacher union, 54 representatives of the regional government, 16 representatives of municipalities, 3 owners of privately subsidized schools, and 5 entrepreneurs related to education. All participants were randomly selected from databases of the group they represent,[6] but if a selected participant did not attend it was not possible to replace him or her.

All deliberations took place on a Saturday (a day when most of the population does not work) and in different consecutive months. The details of the organization of the dialogs fulfilled the following key characteristics as discussed by Callon et al. (2001): balanced representation, shared rules of the game, equal access to speech, and transparency described. A balanced representation of key stakeholders and the use of small groups, first according to self-selected subjects and later in combinations of the different subjects, gave each participant the possibility of expressing her opinions (several times during the day) and listening to other perspectives. The physical dialogs were fully recorded and typed up for the application of text mining tools.

The third building block was an e-forum, consisting of an interactive web page maintained during six months during the same time span as the "physical dialogs." Participants were able to upload or send their opinions and maintain dialogs with others, in different chat rooms, organized by interests as in the first part of the physical dialogs. This is the form of participation that allows a greater number of participants, but was used well below its potential as only 2,000 individuals participated.

Data-mining techniques were applied to the second and third building blocks, where major deliberation processes occurred. Note that the second building block might arrive, by design, at summary conclusions and proposals, but this was not possible for the third block. It is important to stress

that data mining allows one to extract centroid opinions from the process, not giving more weight to summary conclusions than to any other opinion pronounced.[7] In the SOFM operation, each citizen opinion is considered as a single document. For example, the following is an opinion expressed by the parent of a student in Antofagasta: "I choose this topic because I'm interested in the quality of the people that work in a kindergarten; I mean I'm worried about the psychological test, because I believe that it is important before somebody is hired." Each opinion is labeled with a name that summarizes its region of origin, the stakeholder's role (parent, guardian, teacher, etc.), and an opinion number. For instance "II_Teacher_12" tells us that the city is Antofagasta, the participant is a teacher, and the opinion number is 12. Next the vector space model was applied to the entire set of opinions, transforming it in feature vectors with numerical values as subsequently described.

Data Cleaning and Preprocessing

The first step was to standardize the characters in each dialog. For example, we had to rewrite the entire set of letters in lower case, the accents that are used in Spanish were removed (e.g. ,we replaced "í" by "i"), and special letters like "ñ" were changed to "n."

The second step consisted of removing stop words, including articles, prepositions, and conjunctions, from the dialog. In the third step, we applied a table of synonyms, changing some words to a corresponding synonym to reduce the total number of words. An example is the reduction of the words "boy," "boys," "kids," and the like to one word, "child." The same method was applied to compounded words.

Finally, we applied a stemming process, reducing a word to its root; for instance, the words "write," "wrote," and "writing" were transformed into "write." From the initial 414,480 words, the cleaning and preprocessing tasks yielded 9,850 terms that represent the meaning of the entire set of words, and that support the VSM for transforming text to numeric values efficiently. Following the process discussed earlier, the 3,476 citizen opinions were transformed in vectors, each of them with 9,850 components of length.

Applying SOFM and Reverse Clustering Analysis

The SOFM was implemented using the processing capabilities of the Perl computing language in a Computer Xenon III, with two CPUs (1 GHz; Dell, Round Rock, TX). The SOFM training process took approximately four hours. The results are displayed in graphics showing each neuron and its winner frequency. An example is presented in figure 3, which represents the Antofagasta dialogs. A cluster is obtained from the neuron winner with the greatest winner frequency during the mining process. In figure 3, it is possible to distinguish five main local maxima. Next, it is necessary to identify which of these five are adequate for extracting significant patterns about the phenomenon under study.

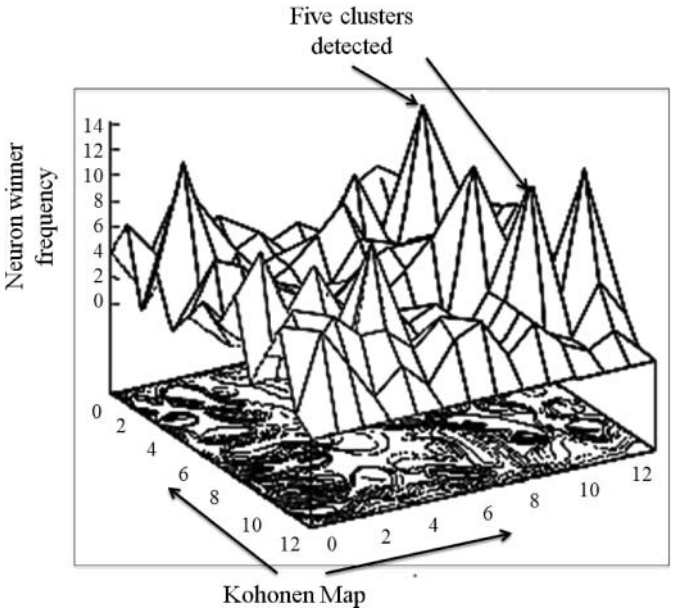

FIG. 3. Cluster for Antofagasta region.

The respective original dialogs were identified by applying RCA to each cluster. Then three clusters were accepted by applying the reject-accept criterion explained earlier.

As an example of RCA, cluster 1 in figure 3 contains eight vectors whose nearest original opinions were: II_Apod
erados24.txt (centroid), II_Director9.txt, II_ CentroApod erados12.txt, II_CentroApoderados56.txt, II_ CentroApod erados80.txt, II_CentroApoderados85.txt, II_ CentroApod erados92.txt, and II_Consejero1.txt.

The opinion centroid contains the following text: "There is another thing about the child, the child's rights. Sometimes, we say anything to the student, maybe speaking in a strong way, and the student say "you are harming my rights," but where are the child's duties?, it seems as if we need some child obligations, for instance that they should obey their parents."

Analyzing the other opinions belonging to cluster 1, the summary or consensus opinion is: "They discussed the problem generated by the lack of clarity about students' rights and duties and the role of other actors. For example, they perceive that the United Nations convention on children rights is contradictory with the enforcement of discipline."[8]

The process was applied to each region separately and for the entire data set (country). Table 1 displays the

TABLE 1
Cluster solution for the entire country: physical participation

Cluster	Number of opinions	Main theme
1	85	Integrating education stakeholders
2	82	Ineffective education infrastructure investments
3	73	An incomplete education project
4	71	Unclear policies for improving education
5	67	Adding new professional disciplines and educational content to school curricula

Source: Authors' elaboration on the basis of the Citizen Dialogs.

clusters found for the country case, after applying the selection criterion. Because the amount of opinions per cluster is large, and there is a possibility that they all share a common theme, the reject-accept criterion must be relaxed. In our case, if 70 percent of the opinions in a cluster were related with a common topic, the cluster was accepted. The remaining 30 percent of opinions generally provided adequate details for the outside expert to understand the situation expressed in each cluster. Explicit revision of the original opinions of each cluster gives us an idea of which educational topics are the most important for the Chilean population.

In table 1 the cluster's content is interpreted as follows:

- Cluster 1: It is suggested that there is a concern for the integration of all the stakeholders in the educational process, for example, the principals, board members, teachers, students, and so on. It is necessary to establish a clear definition of educational roles to benefit and support the wide range of personnel in an educational establishment. Many critics argue that there is no clear idea of what one should do to improve the education of students, which suffered from lack of funding.
- Cluster 2: It has been acknowledged that the Ministry of Education has contributed funds to the improvement of the infrastructure, educational materials and resources, and development-training programs. However, it has become evident that the funding has not been used effectively. For example, there are colleges that have invested money in audiovisual technologies but have not adequately trained staff to use them effectively.
- Cluster 3: There are suggestions of a lack of clarity about aims and methods for carrying out an integrated evaluation of the quality of education. The policy is created and approved by the government but requires many modifications to be implemented within the realities of the educational context.
- Cluster 4: There is a lack of general and specific information on what one understands about a policy that strives to improve the quality of the education. Often, when new legislation is passed, those in charge of educational establishments, such as board members and principals, can often adjust or manipulate the terms and conditions of the new policy, thus leading to confusion as to where this policy will lead. This lack of commitment on the part of those in charge of educational establishments can have serious repercussions for the education of the student.
- Cluster 5: One can determine the extent of the necessity of change in the current educational establishments, incorporating professionals from other faculties, for example, psychology, with the purpose of organizing a new work plan and new working materials and resources in the face of a more competitive world in which one requires major leadership and effective team work.

Additionally, the e-forum opinions, published on the Web site, were analyzed. There we found seven clusters as shown in table 2.

The pattern extracted from table 1 is comparable to table 2, illustrating the effectiveness of the Web-based systems (the e-forum in this case) for collecting information. In both cases, the main discussion themes appear to be validated by the e-forum results.

One of the activities performed during the Citizen Dialogs was the implementation of focus-group sessions. In each of these sessions, the participants were free to discuss the quality of education in different ways. This experience allows us to compare the results from implemented data-mining tool with those from a traditional method—focus groups—for getting opinions about a given theme.

The discussions in the focus groups can be summarized in nine main themes:

- Continuing education project for teachers and how they use this new knowledge in the classroom (adding new professional disciplines and educational content to school curricula).
- Management and administration of the educational establishments.
- Educational equity.

TABLE 2
Cluster from the e-forum

	Opinions	Opinions male	Opinions female	Opinions undefined	Average age	Coming from metropolitan region	Main theme
C 1:	44	28	15	1	43.4	29.5%	Integrating education stakeholders
C 2:	27	22	5	0	46.9	63.0%	An incomplete education project
C 3:	27	15	6	6	42.4	51.9%	Ineffective education infrastructure investments
C 4:	20	13	5	2	42.8	35.0%	Adding new professional disciplines and educational content to school curricula
C 5:	20	5	12	3	40.2	30.0%	Unclear policies for improving education
C 6:	20	7	11	2	37.7	50.0%	Dissatisfaction with centralized and nonparticipative decision making
C 7:	18	7	10	1	44.3	50.0%	Discipline and children rights and duties

Source: Own elaboration on the basis of the Citizen Dialogs.

- What the students must learn? What the teachers must teach? (*cluster centroid:* An incomplete and unclear educational project).
- Education quality: What is it? It is not clear what the government policy about it is (*cluster centroid:* Unclear policies for improving education).
- Complete scholastic day[9] and infrastructure (*cluster centroid:* Ineffective education infrastructure investments).
- Student conduct in the educational establishments (*cluster centroid:* Students' rights and duties).
- Number of students per educational establishments (cluster centroid: Ineffective education infrastructure investments).
- Method for evaluating educational establishments and students.

Some of these themes appear also in the cluster analysis for the e-forum and country cases. The remaining themes also appear in the cluster analysis for regions (see figure 3 cluster revision). Then, by using both methods, the focus group and the data mining tool, it is possible to extract some similar results. However, comparing both methods, the focus group is very limited in number of participants, the participant's opinion can be influenced during the process, and the main themes identified are fewer than those generated with the help of the proposed tool.

Vox Populi Vox Dei

As indicated, the application of data mining to the dialogs produced an average of 4 clusters per city and 5 national clusters. The e-forum produced 7 additional clusters. These 32 clusters are now briefly summarized, as the objective here is to use the information for illustrative purposes and not to analyze the results from the educational policy perspective. It is important to note that the objective of the exercise is not to regroup clusters, as each cluster provides important information for policymakers and is done here only for clarity of exposition and to remain focused on the main argument.

A demand for more information and orientation appears in eight clusters. Chile has a decentralized market-oriented educational system in which the role of the different actors is neither unambiguously defined nor clearly understood. Not surprisingly, clusters of opinions are found around the need to clarify the role of the different actors and to improve accountability.

Ten other clusters concentrated on participation and the educational role of families. Chile has a long tradition of authoritarianism and vertical hierarchical relationships, especially in the educational system. This organization is being challenged by demands from families for a more active role in their children's education and for more information and communication from schools and the government, including media campaigns for parental education.

Since 1990, the Ministry of Education has played an active role providing different inputs to schools and deciding, instead of local school administrators, which inputs are the most convenient for them. Four clusters show dissatisfaction with centralized and nonparticipative decision making by the Ministry of Education, which provides similar inputs to all schools, without proper consideration of the real needs of local communities.

Ten opinion clusters expressed concerns regarding teacher capacities and workload. Three out of seven e-forum clusters addressed this issue. Two clusters expressed concerns regarding discipline and children rights and duties. Finally, one cluster pointed to the need for preparing adolescents for options other than university

education, as most of the population would not follow that path.

Public Polities in Action

The original objectives of the Citizen Dialogs were to obtain inputs for "a better communication of educational policies, to improve educational policies and to increase their legitimacy." It is interesting to evaluate which objectives were achieved and outcomes obtained as direct and partial results of the dialogs. Nevertheless, this was not an easy task, as the team in charge of the design and implementation of the dialogs was disbanded and did not continue working on the issue after the change of government in March 2006. There was no evaluation or follow up of the dialogs, and participation and accountability issues were left unassigned within the Ministry of Education. With the exception of environmental issues (where consultation to citizens was established by law), no other large participation exercises involving deliberation have been carried out so far in Chile.

To evaluate the effect of the Citizen Dialogs, we carried interviews with the former minister and undersecretary of education 16 months after the process occurred to discuss their views about its impact.[10] Lower ranking officials were also consulted.

Both the minister and undersecretary of education held similar views—they valued the process and claimed it had influenced their own perceptions and decisions, but not necessarily the Ministry of Education's rank and file. It was surprising how well they remembered the key issues posed by the dialogs. The most important influence was mainly in the definition of policy priorities and communication strategy. The minister of education also realized the importance of a friendly and interactive Ministry of Education Web page, which later was awarded a prize for the best Web page in the Chilean public administration. The minister of education established personal communication with teachers. In his discourse, the dialogs appeared as an essential element to redirect policy management and he regretted that no administrative division within the Ministry of Education had become responsible for the issue and projected it into the future.

Apart from communications, the minister of education considered the Citizen Dialogs highly influential in his decision to establish and appoint four national commissions of experts and stakeholders for four key policy areas—children with special needs, civic education, sexual education, and national system for the measurement of educational quality. So too the minister of education indicated that some legislators[11] had been influenced by the Citizen Dialogs, especially in the creation of school councils, which gave school communities a voice in the administration of the schools and the right to request specific information about the situation of their school.

The bureaucratic structure, however, was not affected and seems to remain impermeable to this type of participatory initiatives. There was no continuity in the relationship with stakeholders, except for the forum of Education for All, a panel established by the United Nations Educational, Scientific and Cultural Oragnization in different countries, independent of the Citizen Dialogs experience. The original ambition of creating a permanent people panel was aborted because of the failure to assign responsibility and resources. Participation continues to be acknowledged as a crosscutting issue by the Ministry of Education, but with no unit accountable for its continuity, in spite of the pronouncements by top authorities of the importance of integrating citizens into public policy. Only a few enthusiastic regional offices of the Ministry of Education undertook a systematic policy of organizing stakeholders, especially students in secondary education.

All the individuals invited to participate in the exercise were satisfied. The fact that top educational authorities were present during the whole process was important, as they felt they had been heard. All received a copy of the Citizen Dialogs' conclusions. However, lower ranking officials in the Ministry of Education were more critical about the experience, as they did not perceive the need for participation and continued to prefer traditional top-down decision making. By the end of 2006, less than a year after the interviews were conducted, policy directions had changed substantially toward strengthening the regulation of the market system and development of a better governance structure (González 2008).

We consider that the Citizen Dialogs' first objective, communications, was a success. Ex-post it seems that communication was the key objective for top officials, as they had firsthand opinions and exchange with stakeholders representing different interests. Among the Ministry of Education administrative units, the communications staff was particularly competent and well-prepared. They also took full advantage of the abundance of materials provided by the Citizen Dialogs and its subsequent analysis through data mining.

With regard to the second objective, improving educational policies, it is not possible to disentangle the effect of the Citizen Dialogs from influences from other sources. In any case, although the exercise produced information valuable for policy design, its use depends on the public agency in charge of such policy. It is impossible to trace back policy change to the Citizen Dialogs, as many other events influenced policy design and implementation. However, many initiatives are currently being developed or discussed in the Congress to strengthen decentralization, foster participation at the local level, improve accountability, and clarify the governance structure

of the system (see laster description). Discussions are taking place on how to improve the teaching profession beyond what has been attempted since the return of democracy. However, it is not possible to isolate the effect of the Citizen Dialogs from other influences on these achievements, such as secondary students' unrest in the winter of 2006 or the presidential commission report delivered in December that same year. A more direct influence might be traced, for instance, when the former Ministry of Education recognized the inspiration of the Citizen Dialogs for the creation of the school councils, enacted before the change of government in 2006.

In any case, several other concerns formulated in the Citizen Dialogs are actually being considered by policymakers. For instance, three pieces of legislation dispatched to Congress by the Bachelet government address several issues raised in the first eight clusters previously described. One of them is intended to amend the constitution. All three were expected to be approved, together with a law strengthening public education, during the first semester of 2009. Aside from the school councils introduced shortly after the Citizen Dialogs, an early education program named "Chile Crece Contigo" (Chile Grows with You) considers parental education to support their children's educational and life experience. A National Evaluation of Educational Quality, the SIMCE, is also providing detailed information about each school performance as compared to others. Finally, the "Subvención Preferencial" enacted in 2007 created a means-tested voucher directly transferring roughly 50 percent more resources to schools for children belonging to the most vulnerable sectors, giving schools more freedom to decide which inputs and investment to undertake. In the case of the worst-performing schools, this must be decided in the framework of a development plan produced with expert external assistance. The Ministry of Education no longer decides on the specific inputs but monitors the development plans and certifies the external agencies.

The effect on policy legitimacy is also difficult to trace back. The lack of follow-up, the limited number of participants compared with the total population—despite being the largest participation exercise ever conducted in Chile, with the potential to have included 1,000 times more participants (through the Internet)—the limited media coverage, and the ignorance about its influence on policy decisions vitiated against policy legitimacy (except for participants themselves).

The absence of effects on legitimacy might be confirmed by the vigorous high-school-student protest of 2006, which questioned the legitimacy of key aspects of the educational system. Several questions posed by the student movement had antecedents on the dialogs, especially the critique of the decentralized market-oriented system that has failed to produce equal opportunities for all. Also, it is impossible to determine to what extent the explicit policy of organizing student centers in secondary schools (a consequence of the enthusiasm of some regional Ministry of Education offices for the Citizen Dialogs) was responsible for the strength of this movement, or what resulted from the fact that participation channels were not kept open after 2005 and that the strong demand for greater voice clashed with Chile's authoritarian culture (Veliz 1980; UNDP 2002; Heine 2002). However, well before these events, in our interview, Minister Sergio Bitar anticipated the need to maintain participation mechanisms open in the sector, to improve policy legitimacy, and to prevent a backlash like the one observed a few months later. In fact, the current government response to the already-mentioned public unrest was the creation of a National Commission for Education (primary and secondary) and later one for higher education. Had Bitar been appointed as minister of interior or as secretary general of the president, as expected in early 2006, a national dialog on social protection would have occurred and the promise of participatory government would have been better realized.

SUMMING UP: IT WORKS

So far we have presented a technological tool capable of processing large amounts of opinions and classify them into clusters. Further, we have applied it to a large deliberation process, showing its potential to extract shared concerns, opinions, and desired solutions. The exposition in the last section has grouped these clusters for the sake of saving space and concentrating attention on the main issue, but each cluster provides interesting information about opinions shared between human beings. This section answers two questions: why it was necessary to develop a specific tool for our purpose, and why we consider this route to be a fruitful avenue for strengthening modern democracies.

To answer the first question, it is interesting to consider an early application of data mining to political sciences: the work by Laver et al. (2003). We have already mentioned that the tool used in that work was adequate for the purpose at the time it was designed. However, their procedure was not adequate for handling a large deliberation process (and this is not a criticism of their work, as it was not designed for that purpose). It is interesting, however, to briefly explain the limitations of their algorithm, as most of them apply or are amplified in the case of standard commercial software.

First, Laver et al. (2003) require the selection of a small sample of texts whose content is known a priori, extract from them the relevant words to use for comparison, and afterward search for these words in a larger unknown set of documents. Second, the comparison proceeds by words, not by meaning. Third, the procedure is

"supervised"—in other words, it restricts the possible results to those predefined by the small set of documents known a priori. In contrast, our procedure compares all the documents with each other without imposing the possible results and compares by semantic meaning instead of exact words, which are the two key requirements to process the information of a large open-deliberation experiment.

Of course, we do not pretend to have obtained a definitive solution, especially considering the rapid rate of new knowledge accumulation in the field of data mining. However, we have developed a solution that works for a new area of applied research: the use of data mining for processing large deliberation experiences. To our knowledge, this proposal is the closest we have come so far in achieving the dream of deliberative direct democracy, in the sense that it is now possible to process a deliberation with several thousands (even millions) of participants. This dream has its detractors, however. Let us consider some of the potential objections and answer our second question.

First, it might be argued that 30 representative persons can arrive, after deliberation, at the same conclusions as 1,000 or 1 million people. In our view, a key difference between both situations is the feeling of participation and exercising citizenship that can be obtained from the latter and is impossible to reach with the former. In fact, one key reason for the popularity of participation in Western societies is voters' disenchantment with democracy (Font 2003; Dahl 1999). Our proposal expands the possibilities for "enchantment" by moving upward the number of people who can participate and deliberate in any particular issue.

A second reason why "large numbers" deliberation might be superior is that by increasing the number of persons participating, the potential for good ideas considerably increases. This might be more relevant in certain issues, for instance, where uncertainty, novelty, and the scope for innovation might be high. Our data-mining approach has the limitation that it can only detect ideas or issues that attract the attention of several individuals. However, it is likely that good ideas will capture the interest of the people participating in the space where the idea is generated, and if they do, the ideas will reach decision makers. The possibility that good ideas reach decision makers is actually enlarged if top government authorities participate in the physical events, as is the case in the application presented later in this article.

Another objection to large participation exercises might be its costs, as it consumes the time of the people involved in the experiences (both in the production and the participation side).[12] Even from a liberal point of view, if the participation in the exercise is voluntary, the theorem of revealed preferences ensures that everyone participating is at least as well off as those not exercising their right to participate.[13] In the experience reviewed in this article, the production costs were relatively low, as schools normally closed on weekends were used for the physical dialogs and most human resources involved in the production of the events were volunteer university students.

Finally, consider the elitist argument (Manin 1997) that elected representatives or a group of experts might arrive at better solutions than uninformed common citizens. It is obvious that in the design of the deliberation process the lack or incompleteness of information possessed by participants must be addressed, and this might be more or less complicated, depending on the complexity of the issue. However, ordinary citizens involved in a public issue—and therefore with some knowledge of it—might provide important new information to the policy forum, especially in issues with high uncertainty, low "specificity" (the production functions are unknown), and where relationships between people are important. Beyond scientific evidence, of which there is little in the case of quality of education in Chile,[14] knowledge about the problem is scattered in many small pieces diffused across many actors. In addition, participation and deliberation might bring to the political forum issues that elites have been reluctant to address. For instance, elites in Chile have been traditionally opposed to push decentralization further or to foster participation (UNDP 2005). The social deliberation process that started with the Citizen Dialogs has put some of these issues on top of the educational policy agenda. Finally, note that large participation exercises might be complementary to, not a substitute for, elite representation. In any case, the conclusions and results of the participatory exercise were not mandatory for the government, and they cannot be, as the conclusions and recommendations of the different spaces of deliberation will certainly differ. They were used for policy formulation after educational experts processed the information obtained and compared it with available international scientific evidence. However, the opportunity provided by voice mechanisms (whether Citizen Dialogs, secondary student movement, or presidential commission) put in place a powerful social force for educational change that the political elite could no longer control at will.

Besides, the technological tool we have developed might be used for other purposes as well. For instance, it can process what Habermas (1996, 308) called "informal opinion formation" and expand their potential to improve contextual information for public decisions. It is also complementary to actual democratic institutions. For instance, it might be used to multiply the number of representatives in "promising representation," moving the deliberation process (using representation) closer to direct democracy (Mansbridge 2003). Or representatives might use the technique to better represent their

constituencies. The possibilities of application are almost unbound.

CONCLUSION

A major problem of modern democracies is the legitimacy of public decision making. Participation and deliberation are perceived as key for enhancing this legitimacy. However, deliberation is constrained by limits of time, space, and human capacities for listening and processing information. In this article, we describe a technological tool that can help remove these constraints. Although the Internet already provides the means to deliberate without the need to meet at the same place and time, its conjunction with data mining can be used to solve what we have labeled the "large numbers deliberation dilemma." Traditional data-mining algorithms provide the capacity to extract significant patterns from huge quantities of data. We describe a new data-mining-based technological tool, to extract consensus opinions from a large number of "Citizen Dialogs." The data mining tool is based on SOFM, in a simple version for processing text. Future work must include more sophisticated text processing methods such as natural language processing, which allows the detection of new themes and the comparison of texts by using more complex semantic models.

> Compared with a traditional method—the focus group—for extracting opinions about a specific topic, the technological tool developed can be applied in forums with a huge number of participants, being only limited by the available computer resources, whose capabilities increase and prices decline with time. Also, the data-mining tools allow for identification of themes that do not emerge out of focus groups. In that sense, both methods can be a good complement to each other. Maybe the most complicated part is the transformation of a verbal opinion in a text. This step is necessary for preparing the data to use as input of the data mining tool. Nowadays this challenge also can be tackled by using another technological tool, for instance, the e-forum Web site as method for collecting the citizen opinions directly in text format.

The application was successful for extracting new information. It gave a clear map of the opinions of those participating in the exercise, which constituted a balanced representative sample of stakeholders. It remains for further research to see whether the information obtained is somehow different from that of conventional sources. For the moment, we can confirm the feasibility of processing the deliberations of a large number of individuals. Correctly used, this could strengthen the legitimacy of modern democracies. While the data-mining tool is potentially very powerful, realization of its full potential requires that it be implemented in suitable settings, for instance, with a strong government commitment to participation and with adequate follow-up strategies. Barber (1998/1999) put the argument in the following manner: "If democracy is to benefit from technology then, we must start not with technology but with politics." This article discusses an example of a very promising technology, but its success requires commitment and effort. It is up to human beings and especially politicians to use this technology effectively, to strength our democracies through the direct deliberation of millions of empowered citizens.

NOTES

1. For instance, a Rousseauian model of direct democracy requires overcoming disagreements to achieve the general good, and this is only feasible in isolation, as in that state each individual, free from passions and emotions, would get in touch with the "general good" as an outcome of his or her own reasoning.

2. An early application to political sciences is Laver et al. (2003), who used a basic algorithm to extract patterns from political discourses. We explain in the second last section why our methodology is superior and the other is not suitable for our purpose.

3. However, the organization of the Citizen Dialogs was intended to provide a feedback process, as is explained later in the article.

4. Puerto Montt, Viña del Mar, Valparaíso, Antofagasta and Santiago.

5. Same as for the first stage except for the replacement of Viña del Mar by Concepción.

6. Chile has extremely good databases for individuals as a consequence of a unique ID number used for all civic and even private purposes, including school enrolment and labor contracts.

7. Of course, if conclusions are generated for the second building process of the Citizen Dialogs, they might be contrasted with the clusters opinions obtained by applying data mining, to check for consistency and differences of both summary mechanisms.

8. This is consistent with the authoritarian and centralist traditions of the country. The school system is slowly incorporating democratic conflict-resolution methodologies.

9. Study day lasts from 8:00 a.m. until 5:00 p.m.

10. At the time the interviews took place, the former undersecretary, Maria Ariadna Holnkron, was Minister of Education and the former Minister of Education, Sergio Bitar, was leaving his duties as coordinator of the presidential campaign of the newly elected president.

11. In Chile, there is no Common Law and therefore new issues, such as the protection of the right to education, need to be incorporated through new legislation.

12. Representation liberates citizens from their civic obligations, that is, having to discuss public affairs, thus allowing them to specialize or dedicate themselves to other issues (Macey 1994).

13. The requisite representative sampling of participants ensures that each citizen deciding not to participate will be replaced by a similar citizen, at least sharing similar interests regarding the issues being discussed.

14. And the available scientific evidence was integrated in the deliberation exercise through the participation of experts and the provision of information at the beginning of the process.

REFERENCES

Aas, K., and L. Eikvil. 1999. Text categorisation: A survey. Technical report, Norwegian Computing Center, http://www.nr.no/files/samba/bamg/tm_survey.ps (accessed February 1, 2009).

Ackerman, B., and J. S. Fishkin. 2002. Deliberation day. *Journal of Political Philosophy* 10(2):129–52.

Adkeniz, Y. 2000. Policing the Internet: Concerns for cyber-rights. In *Reinvigorating democracy? British politics and the Internet*, ed. R. Gibson and S. Ward, 169–88. Aldershot, UK: Ashgate.

Alvarez, R. M., and J. Nagler. 2001. The likely consequences of Internet voting for political representation. *Loyola of Los Angeles Law Review* 34(3):1115–1152.

Barber, B. R. 1998/1999. Three scenarios for the future of technology and strong democracy. *Political Sciences Quarterly* 113(4):573–189.

Becker, T. 2001. Rating the impact of new technologies on democracy. *Communications of the ACM* 44(1): 39–43.

Berry, M. J. A., and G. Linoff. 1997. *Data mining techniques*. New York: Wiley.

Bimber, B. 1998. The Internet and political mobilization. *Social Science Computer Review* 16(4):391–401.

Callon, M., P. Lascoumes, and Y. Barthe. 2001. *Agir dans un monde incertain: essai sur la démocratie technique*. Paris: Le Seuil, La couleur des idées.

Cleveland, H. 1975. How do you get every body in on the act and still get some action? *Public Management* 57:3–6.

Coleman, S. 2001. *Democracy online: What do we want from MP's Web sites?* London: Hansard Society.

Dahl, R. A. 1999. *On democracy*. New Haven, CT: Yale University Press.

Delgado, M., M. J. Martín-Bautista, D. Sánchez, and M. A. Vila. 1999. Mining text data: Special features and patterns. *Lecture Notes in Artificial Intelligence* 2447:140–153.

Duhart, J. J. 2002. *Gestion publique et participation citoyenne: Leçons de l'expérience internationale et stratégies pour le Chili*. Une analyse des dispositifs innovants de consultation et de dialogue public, Memoire de recherché, Master en Administration Publique, ENA.

Elmer, G. 1997. Spaces of surveillance: Indexicality and solicitation on the Internet. *Critical Studies in Mass Communication* 14(2):182–191.

Etzioni, A., and O. Etzioni. 1999. Face-to-face and computer-mediated communities: A comparative analysis. *Information Society* 15:241–248.

Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth. 1996. From data mining to knowledge discovery: An overview. *Ai Magazine* 17:37–54.

Fishkin, J. S. 1991. *Democracy and deliberation: New directions for democratic reform*. New Haven, CT: Yale University Press.

Fishkin, J. S., S. A. Rosell, D. Shepherd, and T. Amsler. 2004. Choice dialogues and deliberative polls: Two approaches to deliberative democracy. *National Civic Review* 93(4):55–63.

Font, J. 2003. *Public participation and local governance*. Barcelona: ICPS.

Freeman, J. A., and D. M. Skapura. 1991. *Neural networks: Algorithms, applications, and programming techniques*. New York: Addison-Wesley.

Fung, G. 2001. *A comprehensive overview of basic clustering algorithms, Technical report*. http://www.cs.wisc.edu/∼gfung/clustering.pdf (accessed October 1, 2009).

Galston, W. 2003. The impact of Internet on civic life: An early assessment. In *Governance.com: Democracy and the information age*, eds. E. C. Kamarck and J. S. Nye, 40–58. Washington, DC: Brookings Institution.

Gibson, R. K. 2001. Elections online: Assessing Internet voting in light of the American Democratic Primary. *Political Science Quarterly* 114(4):561–583.

Gibson, R. K., W. Lusoli, and S. Ward. 2005. Online participation in the UK: Testing a 'contextualised' model of Internet effects. *British Journal of Politics and International Relations* 7(4):561–583.

Goebel, M., and L. Gruenwald. 1999. A survey of data mining and knowledge discovery software tools. *ACM SIGKDD Explorations Newsletter* 1(1):20–33.

González, P. 2008. Governance, management and financing of educational equity-focused policies in Chile. Background paper for Education for All Global Monitoring Report 2009. Paris: UNESCO.

Graham, S., and S. Marvin. 1996. *Telecommunications and the city: Electronic spaces, urban places*. London: Routledge.

Habermas, J. 1996. *Between facts and norms*. Cambridge, MA: MIT Press.

Hampton, K. 2003. Grieving for a lost network: Collective action in a wired suburb. *Information Society* 19(5):417–428.

Hand, H. M., and P. Smyth. 2001. *Principles of data mining*. Cambridge, MA: MIT Press.

Heine, J. 2002. PNUD. Desarrollo Humano en Chile 2002. Nosotros los chilenos: un desafío cultural. *Perspectivas en política, economía y gestión (Santiago)* 6(1):165–173.

Hinton, G., and T. J. Sejnowski. 1999. *Unsupervised learning and map formation: Foundations of neural computation*. Cambridge, MA: MIT Press.

Hirschman, A. O. 1973. *Exit, voice and loyalty*. Cambridge, MA: Harvard University Press.

Katz, J. E., R. E. Rice, and P. Aspden. 2001. The Internet, 1995–2000. Access, civic involvement, and social interaction. *American Behavioral Scientist* 45(3):405–419.

Kohonen, T. 1987. *Self-organization and associative memory*, 2nd ed. Berlin: Springer-Verlag.

Kraut, R., V. Lundmark, M. Patterson, S. Kiesler, T. Mukopadhyay, and W. Scherlis. 1998. Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist* 53(9):1017–1031.

Laver, M., K. Benoit, and J. Garry. 2003. Extracting policy positions from political texts using words as data. *American Political Science Review* 97(2):311–31.

Lessig, L. 1999. *Code and other laws of cyberspace*. New York: Basic.

Lin, X., D. Soergel, and G. Marchionini. 1991. A self-organizing semantic map for information retrieval. *Proceedings of the 14th International ACM/SIGIR Conference on Research and Development in Information Retrieval*, 262–69. New York: ACM Press.

Lipow, A., and P. Seyd. 1996. The politics of anti-partyism. *Parliamentary Affairs* 49(2):273–284.

Macey, J. R. 1994. Packaged preferences and the institutional transformation of interests. *University of Chicago Law Review* 61(4):1443–1478.

MaCulloch, W. S., and W. Pitts. 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5:113–133.

Manin, B. 1997. *The principles of representative government*. Cambridge, UK: Cambridge University Press.

Mansbridge, J. 2003. Rethinking representation. *American Political Sciences Review* 97(4):518–528.

Mossberger, K., C. J. Tolbert, and M. Stansbury. 2003. *Virtual inequality: Beyond the digital divide*. Washington, DC: Georgetown University Press.

Negroponte, N. 1995. *Being digital*. London: Coronet.

Nie, N. 2001. Sociability, interpersonal relations, and the Internet. *American Behavioral Scientist* 45(3):420–435.

Nie, N. H., and L. Erbring. 2000. I*nternet and society: A preliminary report*. Stanford Institute for the Quantitative Study of Society, http://www.stanford.edu/group/siqss/.

Norris, P. 2001. *Digital divide*. Cambridge: Cambridge University Press.

Poster, M. 1997. Cyberdemocracy: The Internet and the public sphere. In *Virtual politics: Identity and community in cyberspace*, ed. D. Holmes, 212–28. London: Sage.

Putnam, R. 2000. *Bowling alone*. New York: Simon & Schuster.

Rautenfeld, H. 2005. Thinking for thousands: Emerson's theory of political representation in the public sphere. *American Journal of Political Science* 49(1):184–197.

Ríos, S. A., J. D. Velásquez, H. Yasuda, and T. Aoki. 2006. Conceptual classification to improve a Web site content. *Lectures Notes in Computer Science* 4224:869–877.

Roberts, N. 2004. Public deliberation in an age of direct citizen participation. *American Review of Public Administration* 34(4):315–53.

Salton, G., A. Wong, and C. S. Yang. 1975. A vector space model for automatic indexing. *Communications of the ACM* 18(11): 613–20.

Schuefele, D. A., and M. Nisbet. 2002. Being a citizen online: New opportunities and dead ends. *The Harvard Journal of Press/Politics* 7(3):55–75.

Shah, D., N. Kwak, and R. L. Holbert. 2001. Connecting and disconnecting with civic life: Patterns of Internet use and the production of social capital. *Political Communications* 18:141–162.

Shah, D., M. Schmierbach, J. Hawkins, R. Espino, and J. Donavan. 2002. Non-recursive models of Internet use and community engagement: Questioning whether time spent online erodes social capital. *Journalism and Mass Communications Quarterly* 79(4):964–87.

Solop, F. 2001. Digital democracy comes of age: Internet voting and the 2000 Arizona Democratic primary election. *Political Science & Politics* 34(2):289–293.

Streck, J. 1999. Pulling the plug on electronic town meetings: Participatory democracy and the reality of usenet. In *The politics of cyberspace*, ed. C. Toulouse and T. Luke, 18–47. London: Routledge.

Sunstein, C. 2001. *Republic.com*. Princeton, NJ: Princeton University Press.

Theodoridis, S., and K. Koutroumbas. 1999. *Pattern recognition*. London: Academic Press.

Tickle, A. B., R. Andrews, M. Golea, and J. Diederich. 1998. The truth will come to light: Directions and challenges in extracting the knowledge embedded within trained artificial neural networks. *IEEE Transactions on Neural Networks* 9(6):1057–1068.

Toffler, A., and H. Toffler. 1995. *Creating a new civilization: The politics of the Third Wave*. Atlanta, GA: Turner.

Tolbert, C. J., and R. S. McNeal. 2003. Unraveling the effects of the internet on political participation? *Political Research Quarterly* 56(2):175–85.

UNDP. 2002. *Desarrollo Humano en Chile 2002. Nosotros los chilenos: un desafío cultural*. Santiago, Chile: UNDP.

UNDP. 2005. *Desarrollo Humano en Chile 2004. El poder: ¿Para qué y para quién?* Santiago, Chile: UNDP.

Vedel, T. 2006. The idea of electronic democracy: Origins, visions and questions. *Parliamentary Affairs* 59:226–235.

Velásquez, J. D., and V. Palade. 2008. *Adaptive web sites: A knowledge extraction web data approach*. Amsterdam, the Netherlands: IOS.

Velásquez, J. D., and V. Palade. 2007. A knowledge base for the maintenance of knowledge extracted from Web data. *Journal of Knowledge Based Systems* 20(3):238–248.

Velásquez, J. D., P. González, and J. J. Duhart. 2005. Citizen dialog for the quality of the education in Chile. In *Multidisciplinary program for the social dialog*, 141–186. University of Chile, Santiago, Chile.

Veliz, C. 1980. *The centralist tradition in Latin America*. Princeton, NJ: Princeton University Press.

Weber, L., A. Loumake, and J. Bergman. 2001. Who participates and why? An analysis of citizens on the Internet and mass public. *Social Science Computer Review* 21(1):26–42.

Wellman, B., A. Q. Haase, J. Witte, and K. Hampton. 2001. Does the Internet increase, decrease or supplement social capital? *American Behavioral Scientist* 45(3):436–455.

Wilhelm, A. G. 2000. *Democracy in the digital age: Challenges to political life in cyberspace*. New York: Routledge.

World Bank. 2004. *Information and communications for development: Global trends and policies*. Washington, DC: World Bank.