Camera-ready version

Civic CrowdAnalytics: Making sense of crowdsourced civic input with big data tools

Tanja Aitamurto
Stanford
United States
tanjaa@stanford.edu

Kaiping Chen
Stanford
United States
kpchen23@stanford.edu

Ahmed Cherif
UC Berkeley
United States
ahmed.cherif@berkeley.edu

Jorge Saldivar Galli
University of Trento
Italy
jorge.saldivargalli@unitn.it

Luis Santana
Fundación Democracia y Desarrollo
Chile
lsantana@fdd.cl

ABSTRACT
This paper examines the impact of crowdsourcing on a policymaking process by using a novel data analytics tool called Civic CrowdAnalytics, applying Natural Language Processing (NLP) methods such as concept extraction, word association and sentiment analysis. By drawing on data from a crowdsourced urban planning process in the City of Palo Alto in California, we examine the influence of civic input on the city's Comprehensive City Plan update. The findings show that the impact of citizens' voices depends on the volume and the tone of their demands. A higher demand with a stronger tone results in more policy changes. We also found an interesting and unexpected result: the city government in Palo Alto mirrors more or less the online crowd's voice while citizen representatives rather filter than mirror the crowd's will. While NLP methods show promise in making the analysis of the crowdsourced input more efficient, there are several issues. The accuracy rates should be improved. Furthermore, there is still considerable amount of human work in training the algorithm.

CCS Concepts
• Information systems~Crowdsourcing • Human-centered computing~Computer supported cooperative work • Information systems~Information extraction

Keywords
Crowdsourcing; civic engagement; knowledge discovery; participatory democracy, policymaking; democratic innovations;

AcademicMindtrek'16, October 17-18, 2016, Tampere, Finland
© 2016 ACM. ISBN 978-1-4503-4367-1/16/10⋯$15.00
DOI: http://dx.doi.org/10.1145/2994310.2994366

1. INTRODUCTION
Governments are increasingly inviting citizens to participate in policy-making online by deploying crowdsourcing as an open governance practice [3, 21, 23]. Crowdsourcing is a method for improving the policy with knowledge from the crowd and for engaging citizens [2]. Crowdsourced policymaking processes range from nation-wide law reforms to local governments using crowdsourcing in public policies that affect the residents of cities and towns. Crowdsourcing, when deployed in policymaking, is a democratic innovation [27] in that it aims to engage citizens in democratic processes even between elections.

When crowdsourcing is used in policymaking, the crowd is asked to submit their comments and ideas for the policy. For instance, if the policy deals with transportation in a city, the residents of the city can be asked to share their ideas about how to solve rush hour traffic issues or last-mile connection problems. As the crowd participates actively, submitting hundreds and thousands of comments to the policy, cities and national governments as crowdsourcers are facing a novel challenge: How to analyze and synthesize the crowd's input in an efficient manner? To date, in most crowdsourced policymaking processes the crowdsourced input is processed manually, if at all. The lack of useful analysis and synthesis tools for crowdsourced data hinders the use of the crowdsourced civic input, and can lead to stalling the policymaking process [1]. It can also prevent governments from using crowdsourcing, as they perceive processing the crowdsourced data unfeasible and too time-consuming. Furthermore, due to the lack of efficient analysis tools, the governments also face another challenge: How can they examine to what extent citizens' – or the crowd's – voices are reflected in the policies?

To address these challenges, we have designed Civic CrowdAnalytics, a web application, which analyzes civic data by using Natural Language Process methods and machine learning. Analyzing data from a crowdsourced policymaking process in the City of Palo Alto, California with Civic CrowdAnalytics, we address the question of: To what extent does crowdsourced citizens' input really matter in policymaking?

This paper is structured as follows: In the first section we review key concepts and theories. We then introduce the background of the crowdsourced process in Palo Alto, and we describe Civic CrowdAnalytics. In the third section, we introduce our data and methods. In the fourth section we present the findings. In the final section we present the discussion, conclusions, and the future research agenda, and we discuss the suitability of NLP methods in analyzing policy data, and their use as a part of a policymaking process.

2. THEORETICAL FRAMEWORK

2.1 Crowdsourcing for participatory democracy

Crowdsourcing is an open call for anyone to participate in an online task by sharing information, knowledge, or talent [6, 10, 15]. In crowdsourced policymaking, governments ask citizens to contribute to a policymaking process with their ideas, knowledge, and opinions. The crowd input is then synthesized and channeled into policy. The governments of Iceland and Finland have experimented with crowdsourced lawmaking [4, 19] and the Parliamentary body in Brazil have also deployed crowdsourcing for program reforms [26].

Public policy making follows a set of sequences: Problem identification and definition, data gathering, developing options and proposals, consultation, designing and drafting the policy, decisions, evaluation and implementation [8, 16, 22]. The crowd can be invited to participate in one or more sequences of the cycle. The role of crowd in crowdsourced policymaking is illustrated in Figure 1.

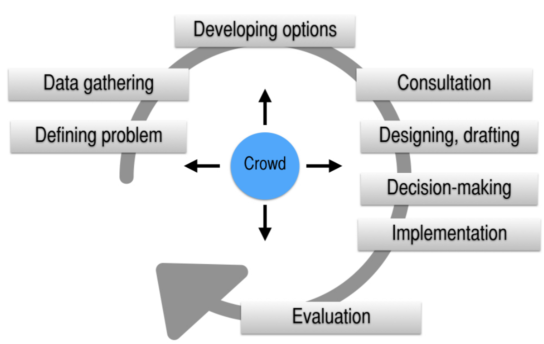

Figure 1. The role of crowdsourcing in policy cycle. The crowd can be part of one or more sequences in the cycle.

Crowdsourcing is a method for participatory democracy [24, 25] in that it provides citizens a possibility to participate in policymaking, and thus potentially influence the course of policy by providing additional information for decision-makers. Crowdsourcing is not a method for direct democracy [11], in which the citizens or residents of an area would directly decide about the policy. Crowdsourcing is typically applied in other stages in the policy cycle but not in decision-making and implementation stages. The elected representatives still decide about the policy, even if the crowd participated in earlier parts of the process. In public policies in local government, it is the City Council that decides about the policy. In legislative reforms, the elected representatives in the legislative system—the House of Representatives and the Senate in the United States, and the Parliaments in most European countries—decide whether the bill is accepted even if crowdsourcing was applied in earlier parts of a policy cycle. It is important to note that crowdsourcing is always based on self-selection, and therefore, the crowd's ideas or opinions are not a representative sample of the so-called public opinion. To detect the public opinion, a traditional opinion polling method with random sampling should be used instead of crowdsourcing. Crowdsourcing, instead, is typically used for searching for ideas, comments and solutions from the crowd as an additional data point in policymaking rather than detecting their opinion about a pre-determined selection of topics.

Despite its increasing use and its potential as a democratic innovation, crowdsourcing faces several challenges. One of the main challenges is managing the crowd's input and extracting epistemic value from it. The more actively the crowd participates, the larger the amount the data there is for the government to analyze and synthesize the crowd's input [1, 4, 5]. Human resources in governments are very limited, and processing the crowdsourced data manually takes a lot of time. Because of this synthesizing challenge, crowdsourced data has been left unanalyzed and unused in policymaking processes [1]. Therefore, there is a growing need for more efficient and automated methods for analyzing crowdsourced data, as that could amplify the possibility for the civic input to be properly taken into consideration in policymaking. To address the challenge of synthesizing and analyzing crowdsourced data, we introduce a novel tool called Civic CrowdAnalytics, which is designed for analyzing crowdsourced civic input.

The lack of appropriate data analytics tools for crowdsourced civic data leads to another related challenge: Despite the proliferating number of crowdsourced policymaking processes in local and national governments, it remains unclear what role of "the crowd" and its input plays in policymaking [12]. To address that question we need to examine the impact of crowdsourced civic input on policymaking processes. Therefore, in this paper, the focus of inquiry is on the following questions: What is the role of crowdsourced input in policymaking? To what extent does the citizens' involvement have an impact on the policy?

3. CASE PROFILE, DATA AND METHODS

3.1 Comprehensive City Plan update

In this paper we draw on data from a crowdsourced policymaking process in the City of Palo Alto, California. Palo Alto is a city in the Bay Area in Northern California with about 66,000 inhabitants. Since May 2015, the City of Palo Alto has been crowdsourcing feedback from its residents for its Comprehensive City Plan (CCP) update. CCP is a fundamental urban planning blueprint, covering transportation, housing, and growth management. CCP is a strategy for the city for next 15 years, and it was last updated in 2002. CCP is divided to several elements, which each represent a topic, such as transportation, growth management and housing. The elements contain programs and policy measures that the city will undertake.

As a part of the Comprehensive Plan update, the city established a Citizen Advisory Committee (CAC), consisting of 18 resident representatives of Palo Alto, to help in developing the policy and for evaluating the crowd's input and incorporate it into the policy. The residents were encouraged to apply for a position at CAC, and the city staff members chose the resident representatives to the Committee. The background of CAC members are urban designers, zoning experts and community organization leaders. CAC meets monthly, and it is divided into subcommittees focusing on specific topics, which also meet once a month. Each CAC meeting focuses on certain element in the policy. Thus, there are three key groups of players in the crowdsourced Comp Plan update in Palo Alto: the online crowd, the CAC members, and the City of Palo Alto: the city staff members involved in the Comp Plan update and the City Council Members, who will decide about the final policy in 2017 and discuss the policy in their meetings before that.

The "crowd" – the residents of Palo Alto – has been invited to participate in several sequences of the Comprehensive Plan update. In the first part of the online crowdsourcing, the city asked the residents to share their ideas for several topics, including transportation, growth management and housing, on an online platform called Open City Hall. People could participate online in the process after they register on the Open City Hall and have a verifiable email address. The crowd input was then analyzed, categorized and synthesized, and provided to the CAC members. In the next step, an earlier version of the Comprehensive Plan was published on a platform called Digital Commenter. The city invited the residents to comment on the draft on Digital Commenter, which allows the users to annotate the draft, leave their ideas and comments there and comment on other users' input. The residents can comment with a nickname, anonymously or a real name. The CAC members are not allowed to comment there because of the Brown Act. This crowdsourced input from the Digital Commenter was then analyzed, categorized and synthesized, and provided to the CAC members.

In the next stage of the policy cycle, the city published a preliminary draft of the first element, the Transportation element on Digital Commenter. The draft received 182 comments from the online participants. The draft was written by city staff members, incorporating the earlier input from CAC and the crowd. The city invited the residents to comment on the draft on Digital Commenter. Until all the elements have been revised, the Comp Plan update will continue in the following cycle: CAC meetings discussing one element at a time, drafting the policy element, publishing it online for residents' comments, analyzing and categorizing crowdsourced input, and discussing the policy element again. The goal of the process is to include civic input in the policy as widely as possible. The update is scheduled to be ready by 2017, and then the City Council makes a final decision about the policy.

3.2 Civic CrowdAnalytics: Making sense of crowdsourced data

While several solutions have been developed for crowdsourcing civic input [13, 17] there is a lack of efficient tools for analyzing the crowdsourced material. Some attempts have been made to use NLP technologies to analyze the quality of writing (structure, grammar) crowdsourced research proposals and to analyze crowdsourced ideas in companies' innovation challenges based on rhetorical structure of the text [7, 29].

These solutions show that it is possible to analyze unstructured data in an efficient manner; however, the solutions have two main issues: First, they are not designed for analyzing crowd's input in civic crowdsourcing projects, in which the threshold for participation is kept low, and thus the submitted data vary in format and content, and hence are very unstructured in nature.

Second, there is not a tool that is designed for public use of crowdsourced data, and specifically the open policymaking use cases in mind. Civic CrowdAnalytics is designed to address the analytics needs in crowdsourced policymaking processes. By using methods for knowledge discovery in data [14], Civic CrowdAnalytics aims to automate as large part of the data analysis and synthesis in a crowdsourced policymaking as possible.

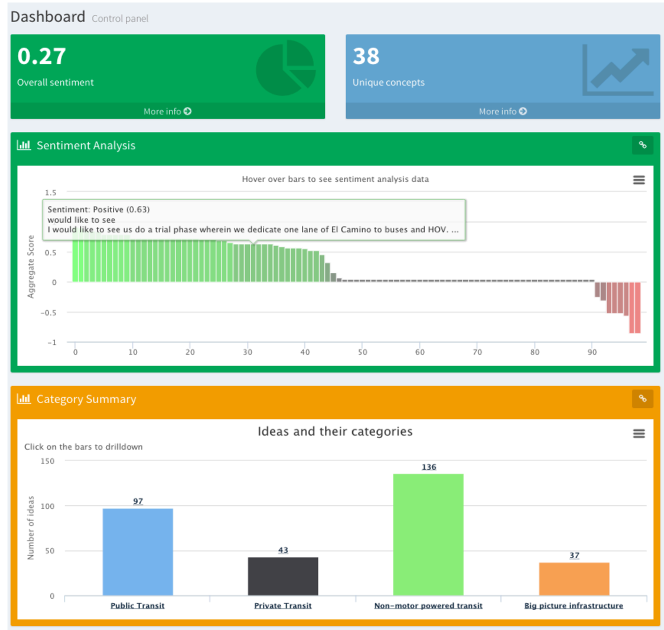

Figure 2. Dashboard for the analytics results in Civic CrowdAnalytics.

We designed Civic CrowdAnalytics in collaboration with the City of Palo Alto to make the data analysis and synthesis crowdsourced policymaking more efficient. The city staff members and policymakers have been facing the issue of overwhelming amount of citizen comments in the crowdsourced Comp Plan update. Civic CrowdAnalytics is a web application, which allows the user to submit data sets and analyze them in various ways. The application uses APIs at Hewlett-Packard Enterprise's big data tool Haven onDemand.

The dashboard in the application most recent analysis results, as Figure 2 shows. Navigating further from the dashboard, the user starts the data analysis by submitting the data in an advised format. The application then analyzes the data based on the parameters the user chooses: categorizing the data to main- and subcategories, finding the most common occurrences and analyzing sentiments. The results are shown in interactive visualizations, in which the user can examine the results by drilling down to more details in the results. For instance, in the Categorization feature results, the user can move deeper to the subcategories by clicking on the bar graph, as Figure 3 shows. The user can also choose the type of visual out put of the analysis, and export the results as csv or pdf –files. In the following, we describe each data analysis feature.

Categorization. The feature categorizes the data into main- and subcategories by using concept extraction. To train the algorithm, the user first codes a part of the data by labeling main categories and subcategories, and then lets the algorithm to categorize the rest of the data. For details about the algorithm, see https://dev.havenondemand.com/apis/extractconcepts#overview

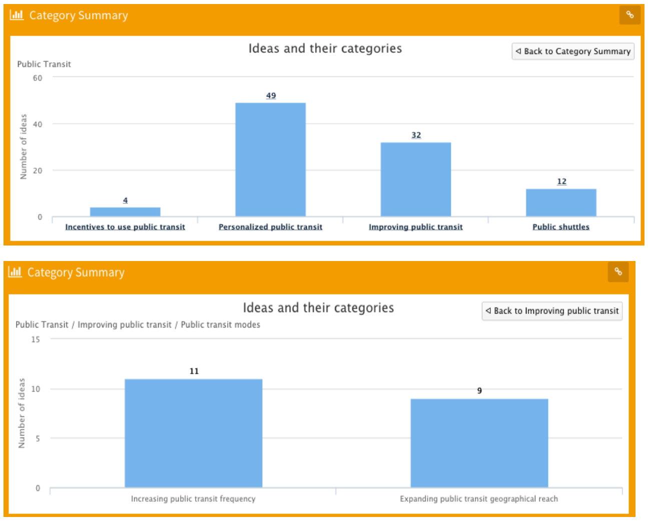

Figure 3. The categorization output, which allows the user to drill down in the categories.

Sentiment Analysis. The data is analyzed based on their positive, negative or neutral sentiment. The algorithm detects the sentiment from words and expressions, such as 'reduce', 'remove', 'problem' would show a negative sentiment, whereas 'increase', 'resolve', and 'good' would show a positive sentiment. For details about the algorithm, see https://dev.havenondemand.com/apis/analyzesentiment#overview

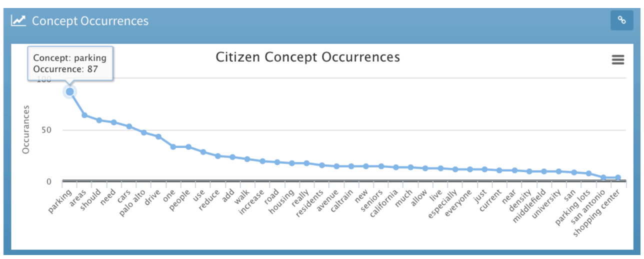

Figure 4. Entity occurrences in the data.

Entity occurrences. Expressions and words are extracted from the data, and shown in the order of the amount of occurrences, as Figure 4 shows. The concept extraction allows us to extract the key terms and their occurrences across the three datasets. For details about the algorithm, see https://dev.havenondemand.com/apis/extractentities#overview

Find similar entity. The analysis shows the association between the unit of analysis, which are the crowd's comments or ideas. For details about the algorithm, see https://dev.havenondemand.com/apis/findsimilar#overview

3.3 Data and methods

3.3.1 Data gathering methods

The analysis for this paper draws on data from the Transportation element of the Comprehensive Plan update. The crowdsourcing period for the Transportation element in the Comprehensive Plan finished in October 2015. Then the local government and the CAC members held monthly meetings to discuss the Plan and the community inputs. We chose the Transportation element as a starting point for the analysis because it is the first element in the policy update that has gone through a full cycle of feedback and revisions.

To gather crowdsourced data, we downloaded the online comments on the Digital Commenter on Transportation Comprehensive Plan. There were 160 comments from 132 users. Since each submission contained several ideas, we separated those into unique ideas that represented the units of analysis. As a result, 182 ideas were identified. Examples of the crowd ideas are "We have never actually instituted density minimums but really need to. We should no longer be building single homes or gigantic condos within half a mile of train station. We should make this actually happen." "We should work with companies like Lyft and Uber to create "on-demand" shuttles. I.e. cars or vans could pick up multiple passengers where and when they need it, instead of running a mindless shuttle that is inflexible and often not useful."

The CAC members offered their suggestions to the Transportation Comprehensive Plan during their monthly meetings, the latest one was 21 June 2016. We turned each speech act into an entry in an Excel spreadsheet, resulting to total of 132 speech acts. In June 2016, the city released their revised Transportation Comprehensive Plan announcing that the plan "incorporates community feedback received to date." In this draft plan, the government highlighted all the edited and the added policies and programs. We downloaded all these edited and added policies/programs and put them into a spreadsheet.

3.3.2 Analysis framework

To examine to what extent citizens voices matter in crowdsourcing policymaking, we compared crowdsourced input, CAC's input and the revised policies in the Transportation element. We focused on the impact of the civic input on the policy changes. CAC's input represents the first "filter" of the crowdsourced inputs since the CAC members raise their suggestions based on crowdsourced input. The revised policies are the second level "filter" of crowdsourced input since they incorporate the feedback from CAC, and secondarily, from the online crowd. Thus, we compare and contrast three datasets: crowdsourced input, CAC input and revised policies. All the datasets consist of unstructured data, and thus the first step is to turn unstructured data into structured data for analysis. We will introduce our human coding method and unsupervised machine learning method in the section below.

3.3.3 Data analysis

3.3.1.1 Quantifying Unstructured Data

To quantify unstructured data, we analyzed and categorized the crowdsourced input, the CAC input and the revised policies by using an open coding method (Strauss and Corbin, 1998). We analyzed all the ideas and developed the first round of categories. After the first round of coding, some categories were merged and some new categories emerged in the process of making the categories as exclusive and complete as possible.

Two researchers conducted the coding and another researcher supervised and evaluated the coding. Our unit of analysis is each crowdsourced idea. We classified each unique idea in the three data sets under a main category and a subcategory. Unclear ideas and category allocations were discussed and resolved collaboratively, and in most cases, clarity could be found. However, in some cases, some comments were off-topic, unclear, or did not contain any concrete ideas. We categorized those ideas in a category called 'Others'. We also compared our categories with the key issues that are discussed in the city's Transportation element in the Comprehensive Plan. As a result of the analysis, five main categories emerged: Big Picture infrastructure, Public transit, Private transit, Non-motor powered transit, and Special

Needs. Under each main category, we used the same open coding method to develop subcategories, which offer richer information about the details of each main category. We applied the same classification of main categories and subcategories to the crowdsourced input, CAC's inputs and the revised policies for consistency. Our coding method was presented in a CAC meeting, and it received positive feedback from the CAC members, the city staff and Palo Alto residents. After the coding stage, we counted the percentage of each main category in all three datasets. We will show the comparison in the Findings section in this paper.

3.3.1.2 Analyzing unstructured data with machine learning

In the second step of the analysis, we used Civic CrowdAnalytics tools to analyze the data. Two Natural Language Processing tools "Concept Extraction" and "Sentiment Analysis" were deployed to analyze the unstructured data. We first drew upon concept extraction algorithm*,* which allows extracting key terms and their occurrences across the three datasets. It enables comparing key terms across the datasets to see whether the revised policies diverge or converge with the crowdsourced input or CAC's suggestions. Compared to the manual coding that applies category names on each data unit, Civic CrowdAnalytics' concept extraction feature compares the units of analysis, that is words (i.e: the nouns and verbs in each suggestion) from the three datasets. We not only want to know *what* the crowd/CAC/city suggests for the policy, we are also interested to know *how* they talk about their suggestions. Therefore, we used sentiment analysis, which algorithm provides information about each input.

This tool identifies all the positive and negative sentiments in a suggestion, and the topics corresponding to the positive and negative sentiments in a suggestion. That is; what are the topics in suggestions that the citizens show negative or positive sentiments about. It also provides an aggregated sentiment score for the data unit. Applying this tool, we generated an aggregated sentiment score for each suggestion in the dataset of crowdsourced input. The scale of the sentiment is from -1 to 1. The negative score is a negative sentiment and the positive score means a positive sentiment. The closer the score is to -1 or 1, the more extreme the public sentiment is.

4. FINDINGS

4.1 The impact of the volume of the crowdsourced input

To examine whether the volume of crowdsourced input about certain topics influences the CAC input and the policy changes, we calculated the percentage of each main categories for the three datasets. In other words, we analyzed if, for instance, the fact that the crowd input contained a high number of demands for new measures to decrease rush hour traffic would translate into equal amount of new measures addressing the respective problem in the CAC and the revised policy. The findings are presented in the following, and summarized in Table 1a. Table 1a shows that comparing the order of proportion, shown in percentages, of the main categories in crowdsourced input, CAC's input and the revised policy, we find that the order of the percentage of main categories between the crowdsourced input and the revised policy is the same while the order of the percentage of main categories between the crowdsourced input and the CAC input is quite different. For instance, in the crowdsourced input, the Big Picture Infrastructure category accounts for most of the suggestions, followed by Public Transit, Private Transit, Non-Motor Powered

Table 1a. Comparing the main categories in percentages in crowdsourced input, CAC's transcript and the revised policy

	Crowdsourced Input	**CAC's Input**	**Revised Policy**
Big Picture Infrastructure	27.07%	24.24%	38.94%
Non-Motor Powered Transit	15.47%	9.85%	4.42%
Public Transit	25.97%	37.88%	27.43%
Private Transit	23.76%	28.03%	26.55%
Special Needs to Senior Citizens	7.73%	0.00%	2.65%
N (unique idea count)	184	132	113

Table 1b. Comparing the top main Subcategories under Big Picture Infrastructure

	Crowd Inputs	**CAC inputs**	**Revised Policy**
Top 1 subcategory	Traffic calming (road safety) (57.14%)	Road (56.25%)	Road safety (28.57%)
Top 2 subcategory	Road design (14.29%)	Transport Evaluation Models and Methods (40.63%)	Road design (14.29%)
Top 3 subcategory	Increasing housing density to reduce car trips (8.16%)		Transport Evaluation Models and Methods (14.29%)

Table 1c. Comparing the top main Subcategories under Non-motor powered transit

	Crowd Inputs	**CAC inputs**	**Revised Policy**
Top 1 subcategory	Infrastructure improvement (89.71%)	Infrastructure improvement (84.62%)	Infrastructure improvement (85.71%)
Top 2 subcategory	More bike supply options (10.29%)		Education program for safety biking (14.29%)

Transit, and finally, Special Needs for Senior Citizens. In the revised policy, the order of the magnitude is the same. This indicates that when the crowd raised more suggestions on certain issue, the government also revised more policies on the issue accordingly. However, CAC members' understanding of the importance of issues are quite different from the crowd. One possible reason is the background of CAC members: they are more elite and expert-oriented than ordinary residents. Nevertheless, as Table 1a shows, the government policy changes followed the pattern of the crowd.

Second, we discovered two interesting percentage differences when comparing our three datasets. The first difference lies in the Big Picture Infrastructure category. The crowdsourced input of this category accounts for 27% of the total crowdsourced input whereas in the policy changes, this category accounts for 38.94%, indicating that the city in fact paid greater attention to this category compared to others. Thus, there is a positive difference of 12%. Conversely, looking at the Non-Motor Powered Transit category and the Special Needs to Senior Citizens category, both the city government and CAC pay relatively little attention and the percentage does not match the percentage in the crowdsourced input. Thus, there is a negative gap.

Since there are large differences in the aforementioned two main categories, we further examined the differences between the crowdsourced suggestions, CAC suggestions and the policy changes in those main categories. Therefore, we analyzed the subcategories and identified the three top percentage subcategories in the three datasets, as shown in Table 1b and Table 1c. We found that in the Big Picture Infrastructure category, road design is among the common top percentage subcategories in the three datasets, indicating that the crowd, the CAC and the city government all pay great attention to this issue. The difference is that the revised policy pays more attention to the road safety issues (the top percentage issue) while the crowd cares more about the traffic calming issues to smooth the traffic (the top percentage issue).

In the Non-Motor Powered Transit category, comparing the three dataset, the similarity is that the crowd, the CAC and the city government all regard infrastructure building for bikes as the top subcategories. Yet a difference is that the city government also regarded education programs for safe biking important, which is not shown in the crowd and the CAC ideas. Therefore, a potential reason to explain the negative difference in Table 1a is compared to other main categories, the city government prioritized the issue that the crowd cared most about (Big Picture Infrastructure) and both the city government and the CAC paid least attention to the issue that the crowd cared less (non-motor powered transit).

4.2 The impact of the sentiments of the crowdsourced input

To examine the impact of the sentiment in the crowdsourced input on CAC and revised policy, we run the average sentiment score of the crowdsourced input for each main category. The sentiment represents the direction of the tone (positive/neutral/negative) as well as the magnitude of the tone (extreme/not extreme) in a particular expression about an issue. For instance, a crowdsourced idea stating "The amount of bike lanes are too inadequate and inconvenient" would show as negative sentiment in the analysis, whereas a statement: "The amount of bike lanes is adequate enough in the city and provides many convenience for the residents" would show as a positive sentiment.

Table 2. Crowd sentiment score and percentage of neutral suggestions.

	Average Sentiment Score	Percentage of neutral suggestions in each category
Big Picture Infrastructure	0.114	34.7%
Non-Motor Powered Transit	0.195	57.14%
Private Transit	0.287	44.68%
Public Transit	0.206	53.49%
Special Needs	0.093	42.86%

Table 3: The extreme (negative) sentiment topics in Big Picture Infrastructure

Example	Returned negative words and scores	Returned topic
I am concerned about two traffic trouble spots that I pass through on Alma St. in the downtown area: At Alma, near University Ave., pedestrians cross freely from both sets of stairs (north & south of the hump) at Cal-train Station. There is no cross-walk there. This needs some study and solutions.	Trouble (-0.53)	traffic
I do think that traffic calming measures in College Terrace have been effective. However, I am worried about the impact of future high-density office development without adequate parking included.	am worried about (-0.68) without adequate	the impact of future high-density office development; parking

Table 2 shows the average sentiment scores of the main categories in the crowdsourced input. Although overall we do not see a strong sentiment among citizens when raising suggestions, we do see that the sentiment score of the Big Picture Infrastructure category is relatively low compared to other main categories, indicating a relative negative sentiment. Table 2 (column 3) also offers us a clearer picture. We calculated the percentage of neutral-sentiment suggestions in the main categories in the crowdsourced input. The percentage of neutral sentiment in the Big Picture Infrastructure category is the lowest while the percentage of neutral sentiment in the Non-Motor Powered Transit is the highest. This provides a potential reason on why the city government pays extra attention to the big picture infrastructure and pays less attention to the non-motor powered transit: the government cares more about crowd opinions that show strong like/dislikes about the government policies, because it shows citizens' dissatisfaction.

The sentiment analysis in Civic CrowdAnalytics also provides details to investigate on the negative sentiment topics among each category. In the Big Picture Infrastructure category, we looked at the topics of the most extreme (negative) sentiments (score <-0.5). Examples in Table 3 show that the crowd is very sensitive about the road design and development issues in Palo Alto. However, the machine learning tool doesn't identify the topics with a 100% accuracy. The accuracy between machine analysis and human coding in main categories is 78.79%, in first level sub-category 58.59% and second level sub-category 51.52%, when comparing the manually-coded and machine-analyzed results.

4.3. Filtering or mirroring the public will?

We explored further whether government revised policies and CAC inputs diverge or converge with the crowd input. Through exploring the divergence and convergence, we are able to know what parts of the crowd suggestions are reflected in the CAC and revised policies. Instead of using the manual coding categories to explore this question, we utilized text analysis tools since it will explore the key terms in each dataset and return the occurrences of those key terms.

Table 4a. Overall high frequency key terms comparison.

	Top five key terms and occurrences
Crowdsourced Input	parking(87), cars(53), drive(43), people(33), walk(21), road(18)
CAC's Input	downtown (80), traffic congestion (55), sustainable transportation (49), paid parking (41), community (38)
Revised Policy	parking(37), improvement (28), development (27), service (20), downtown (12), vehicle/routes/community(12)

Table 4b: Big Picture Infrastructure: High frequency key terms comparison.

	Top key terms and occurrences
Crowdsourced Input	Cars (19), driving (16), road (12)
CAC's Input	Development (16), traffic congestion(15), traffic safety (10)
Revised Policy	traffic (22), improvement (14), safety (12)

Using the concept extraction feature, we identified the top five key terms and their occurrences in each of the three datasets. We found similarities and differences between the most prevalent key terms in each dataset. The similarity is, for instance, 'Parking' is a key word that has high occurrence in crowdsourced input, CAC's input and the revised policy. The difference between the crowd input and the CAC input is the crowd concern more about private transit issues (cars/drive/parking) while CAC concerns more about Big Picture Infrastructure issues (traffic congestion/sustainable traffic). This indicates that ordinary citizens raise relatively more specific suggestions that relate to their own life while CAC members (experts/elites) raise more abstract suggestions that relate to a broader community. The difference between the crowd input and the revised policy is that the city government (besides caring about the private transit issue such as parking) also cares about the general development issues such as service and community.

Comparing with the human coding result (Table 1a, 1b and 1c), Civic CrowdAnalytics provides us several extra findings about the difference between the crowd input, the CAC input and the revised policy. First, as mentioned above, both the CAC members and the government take the broader community development into consideration while the crowd raises more specific suggestions. Second, we see that in the revised policy, there are many occurrences about the word "improvement", indicating the city's determination and willingness to make changes. When analyzing the key terms in the Big Picture Infrastructure category across three datasets, we also found that the text analysis result echoes with our manual coding findings (Table 1b). In Table 1b, we see that the revised policy pays great attention to road safety, and Table 4b below shows that safety and traffic are among the words with high occurrence. Again, the analysis with text analysis tool provides some additional information. For instance, in the crowdsourced input in the Big Picture Infrastructure category, Table 1b (manual coding) just tells us that the crowd cares about traffic calming the most. The key terms in Table 4b below shows us that "car", "driving" and "roads" are the top high occurrence terms. In fact, the combination of car, driving and roads reflect the concerns of the traffic calming which calls for a smoother driving through using road signs and etc. Thus, the Natural Language Processing tool makes the manual human coding results more narrative.

5. DISCUSSION

To what extent citizens' voices matter in crowdsourced policymaking, our findings indicate several results. First, the results suggest that whether citizen voices are incorporated into the policy depend on the amount and the sentiment of their suggestions. When citizens have more demands with a stronger tone, the government pays more attention. Therefore, for the citizen suggestions to be transferred to the policy, both the amount of their input and the tone matter. Citizens' demands that are weak in sentiment or in quantity will be filtered when the policy is drafted by the city government. This finding reminds us about the power of collective action.

Second, considering the role of CAC, which is designed to mirror the public will, while at the same time utilize their expertise to filter the online crowd's will, our result indicates that in terms of representing the crowd's will, CAC's attention on issues are not influenced by the crowd demands in volume and the tone. This is shown in its order of the importance of main categories (issues) that are quite different from the online crowd. In terms of using their expertise to filter the crowd's will, we see that their suggestions are more abstract and community-oriented than the crowd's, indicating the possibility of them using expertise in judging, which issues should be incorporated into the final policy. However, we observed that the city government seems to follow more on the suggestions from the larger public (the online crowd) rather than from the filtered public (CAC).

Regardless, it remains unclear why certain suggestions from the crowd are adapted to the policy, whereas some or not. Furthermore, it remains unknown why the citizen representative body, CAC, doesn't reflect the crowd's input. If the citizens do not feel that their voices are heard and taken into account, it will hamper their motivation to participate. This doesn't mean that the crowd's ideas should be adapted to the policy as is, but the online participants should hear the reasons for the policy decisions.

The Natural Language Processing methods were useful in the analysis of the crowdsourced data, to a certain degree. The analysis tools can mine a large dataset fast, and after training the algorithm, it can categorize the data up to about 80% accuracy rate. That resolves, partially, the analysis and synthesis issue in crowdsourced policymaking. However, training the algorithm takes disproportionally long time compared to its benefit. This is partially due to the small size of the dataset: one needs to code at least half of the data when training the algorithm, and still the accuracy rate could be better. But the larger the dataset, the more meaningful it is to train the algorithm in the beginning. Furthermore, once the algorithm is trained, it can analyze several datasets about similar topics with improved performance, so the city can use it constantly in analyzing their civic datasets. The more the tool is used, the better the analysis results will be, because the algorithm gets trained and can capture as many features as in the training dataset.

6. CONCLUSION

To examine to what extent the citizens' voices matter in crowdsourcing policymaking, we used a new tool called Civic CrowdAnalytics, which uses Natural Language Processing methods to determine the crowd sentiment toward the main categories and to identify high frequency terms across the three datasets. We also used manual coding to compare between the proportions of each main category in the crowdsourced input, CAC's input, and the revised policy. The findings show that the crowd's input is reflected in the policy but, rather surprisingly, the filter consisting of other citizens, Citizens Advisory Committee's input in the policy reflected the crowd's opinion less than the actual policy changes did.

The use of NLP methods shows that they can be useful as knowledge discovery and data analysis tools in crowdsourced policymaking. They can help the city staff members to analyze and synthesize crowdsourced civic input more efficiently. However, the computational methods still have several issues that prevent them from creating a full benefit to the users. The accuracy rates of data categorization should be improved, particularly in the more granular subcategory level. Furthermore, because the training of the algorithm requires human involvement in the beginning of the data analysis, NLP methods should be considered as tools that need frequent use in order to become useful as civic technologies. To make the training process the most efficient, it would be useful for the cities to share their data and results online, so that other cities and actors could use the already trained algorithm for similar topics they are running the analysis on. For instance, cities crowdsourcing feedback for urban transportation plan could share their data online, and the ones who have already trained their algorithms, could pass the algorithms on to other cities, the work would be mutually beneficial. However, despite of the advances in automated analysis, it is clear that there will always be a need for manual, human-powered analysis methods. Every crowdsourced process is different, and the data varies: the topics that surface in a crowdsourced transportation policy process in Palo Alto most likely differ, to a certain degree, from topics that the crowd brings up somewhere else. This restricts the possibilities to have NLP algorithms, which would work globally in all datasets, even when the topics would be the same.

In our future research and design agenda, we will conduct a series of user-testing studies with Civic CrowdAnalytics to examine its fittingness in the end-users' workflow. We will iterate the applications' features and UI based on user-feedback. We will also work on improving the accuracy of the analytics and testing the reliability of automated analysis by comparing the results with manual coding. We will also explore the possibility to develop an openly accessible repository for algorithms trained for analyzing civic data.

7. REFERENCES

[1] Aitamurto, T. 2016. Collective Intelligence in Law Reforms: When the Logic of the Crowds and the Logic of Policymaking Collide. 49th Hawaii International Conference on System Sciences, HICSS. Kauai, January 5-8, 2016. IEEE Transactions, pp. 2780-2789.

[2] Aitamurto, T. and Landemore, H. 2016. Crowdsourced Deliberation: The Case of the Law on Off-Road Traffic in Finland. *Policy&Internet* 7 (2) 174–196.

[3] Aitamurto, T. and Landemore, H. 2015. Five Design Principles for Crowdsourced Policymaking: Assessing the Case of Crowdsourced Off-Road Traffic law in Finland. *Journal of Social Media for Organizations,* vol. 2, no. 1, pp. 1-19.

[4] Aitamurto, T., Landemore, H., Lee, D. & Goel, A. 2014. Crowdsourced Off-Road Traffic Law Experiment In Finland. Report about idea crowdsourcing and evaluation. Publications of the Committee for the Future, the Parliament of Finland. 1/2014. Helsinki, Finland.

[5] Brabham, D. 2013. Crowdsourcing. Cambridge: MIT Press.

[6] Convertino, G., Sándor, Á., & Baez, M. 2013. Idea Spotter and Comment Interpreter: Sensemaking tools for Idea Management Systems. In ACM Communities and Technologies Workshop: Large-Scale Idea Management and Deliberation Systems Workshop.

[7] Edwards, M. 2001. Social Policy, Public Policy — From Problem to Practice. Allen & Unwin, Sydney.

[8] Estlund, D. 2008. Introduction: Epistemic Approaches to Democracy. *Episteme,* 5, 1-4 doi:10.3366/ E1742360008000191

[9] Estellés-Arolas, E., & González-Ladrón-De-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189-200.

[10] Frey, B. S. (1994). Direct democracy: Politico-economic lessons from Swiss experience. *American Economics Review,* vol. 84, no. 2, pp. 338-342.

[11] Fung, A. (2015). Putting the public back into governance: the challenges of citizen participation and its future. *Public Administration Review*, *75*(4), 513-522.

[12] Harding, M., Knowles, B., Davies, N., & Rouncefield, M. 2015, April. HCI, Civic Engagement & Trust. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 2833-2842). ACM.

[13] Holzinger, A. (2013, September). Human-Computer Interaction and Knowledge Discovery (HCI-KDD): What Is

the Benefit of Bringing Those Two Fields to Work Together?. In *International Conference on Availability, Reliability, and Security* (pp. 319-328). Springer Berlin Heidelberg.

[14] Howe, J. 2008. Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business. New York, NY: Crown Business.

[15] Howlett, M., M. Ramesh, and A. Perl. 1995. Studying Public Policy: Policy Cycles and Policy Subsystems. Toronto: Oxford University Press.

[16] Kriplean, T., Morgan, J. T., Freelon, D., Borning, A., & Bennett, L. 2011, May. ConsiderIt: Improving structured public deliberation. In CHI'11 Extended Abstracts on Human Factors in Computing Systems (pp. 1831-1836). ACM.

[17] Landemore, H. 2013. Democratic Reason: Politics, Collective Intelligence, And The Rule of The Many. Princeton University Press.

[18] Landemore, H. 2014. Inclusive Constitution-Making: The Icelandic Experiment. Journal of Political Philosophy. 23 (2), 166–191.

[19] Lourenço, R. P. 2015. An Analysis of Open Government Portals: A Perspective of Transparency for Accountability. *Government Information Quarterly*, *32*(3), 323-332.

[20] Noveck, B. S. 2015. Smart Citizens, Smarter State: The Technologies of Expertise and the Future of Governing. Harvard University Press.

[21] Peters, G. 1999. American Public Policy: Promise and Performance. Chappaqua, NY. Chatham House/Seven Rivers.

[22] Prpić, J., Taeihagh, A. and Melton, J. 2015. The Fundamentals of Policy Crowdsourcing. *Policy & Internet* 7 (3) 340–361. doi: 10.1002/poi3.102

[23] Pateman, C. (1970). Participation and Democratic Theory. Cambridge University Press.

[24] Pateman, C. 2012. Participatory Democracy Revisited. Perspectives on Politics 10 (7), 1–19.

[25] Soares de Faria. F. C. 2013. The open parliament in the age of the Internet: Can the people now collaborate with legislatures in lawmaking? Câmara dos Deputados, Edições Câmara.

[26] Smith, G. 2009. Democratic Innovations: Designing Institutions for Citizen Participation. Cambridge, UK: University Press.

[27] Strauss, A., and Corbin, J. 1998. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage.

[28] Westerski, A., Iglesias, C. A., & Garcia, J. E. 2012, October. Idea relationship analysis in open innovation crowdsourcing systems. In Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2012 8th International Conference on (pp. 289-296). IEEE.