Smart Governance through Opinion Mining of Public Reactions on Ordinances

Manish Puri*, Aparna S. Varde* †, Xu Du†, Gerard de Melo‡
* Department of Computer Science, Montclair State University, Montclair, NJ, USA
† Environmental Management Ph.D. Program, Montclair State University, Montclair, NJ, USA
‡ Department of Computer Science, Rutgers University, New Brunswick, NJ, USA
{purim1, vardea, dux3}@montclair.edu, gdm@demelo.org

Abstract—**This work focuses on the area of Smart Governance in Smart Cities, which entails transparency in government through public involvement. Specifically, it describes our research on mining urban ordinances or local laws and the public reactions to them expressed on the social media site Twitter. We mine ordinances and tweets related to each other through their mutual connection with Smart City Characteristics (SCCs) and conduct sentiment analysis of relevant tweets for analyzing opinions of the public on local laws in the given urban region. This helps assess how well that region heads towards a Smart City based on (1) how closely ordinances map to the respective SCCs and (2) the extent of public satisfaction on ordinances related to those SCCs. The mining process relies on Commonsense Knowledge (CSK), i.e., knowledge that is obvious to humans but needs to be explicitly fed into machines for automation. CSK is useful in filtering during tweet selection, conducting SCC-based ordinance-tweet mapping and performing sentiment analysis of tweets. This paper presents our work in mapping ordinances to tweets through single or multiple SCCs and opinion mining of tweets along with an experimental evaluation and a discussion with useful recommendations.**

Keywords — Big Data; Classification; Commonsense Knowledge; Data Mining; Local Laws; Machine Learning; Sentiment Analysis; Smart Cities; Social Media; Urban Policy

I. INTRODUCTION

This paper centers on enhancing Smart Governance, which falls under the umbrella of Smart Cities. Specifically, we seek to analyze tweets about ordinances or local laws in a given urban region, which represent opinions of people on related topics. This aids in understanding people's reactions to the respective urban policies addressed in these ordinances. An important goal of this work is to assess how well the concerned urban area is progressing towards being a Smart City based on the ordinances passed and the public reactions to them. Fig. 1 shows different Smart City Characteristics, as widely accepted in the literature [1, 2].

Twitter is one of the biggest sources for data mining, with about 350 million users and over 500 million tweets sent per day on a variety of topics. Hence, Twitter is a valuable source of data on public reactions to ordinances. Its micro-blogging nature is useful with respect to the brevity of the information to be analyzed, since each tweet is limited to 280 characters.

We map tweets to corresponding ordinances through their mutual connection with respective Smart City Characteristics based on a measure of semantic relatedness. A trivial keyword matching approach of trying to connect ordinances directly to tweets does not suffice as they both contain intricate and heterogeneous natural language. Moreover, ordinances and tweets both constitute big data, as there are thousands of ordinances and millions of tweets. Hence, obtaining a direct mapping is challenging. Existing techniques from the field of machine learning [4] are not useful to learn these mappings, as they need significant volumes of data to train the models. Ours is pioneering work in the area of ordinance mining and hence we do not possess such large volumes of training data for our mapping task.

Fig. 1. Smart City Characteristics [3]

Instead, our proposed mapping technique takes into account generic connections between ordinances and tweets, through SCCs. We rely on a transitive property: "If ordinances map to one or more SCCs and if tweets map to the same SCCs, the ordinances are likely to be broadly related to the respective tweets". This is due to the finite nature of classical sources of SCC data [1, 2] which possess a limited set of identifying features that can be used for mapping. Hence, this transitive mapping approach seems more feasible, since by using this, we can bypass having to map millions of tweets to thousands of ordinances directly.

Since a single ordinance or tweet can map to one or more SCCs, we develop an algorithm for SCC mapping accordingly. In the process of ordinance–tweet mapping, we make use

2375-0197/18/$31.00 ©2018 IEEE
DOI 10.1109/ICTAI.2018.00131

IEEE computer society

of Commonsense Knowledge (CSK) from sources such as WebChild [5] and WordNet [6]. The use of CSK is vital to measure semantic relatedness in a more informed way. The terms encountered in classical SCC sources may not directly appear in relevant ordinances and tweets. For example, an ordinance or tweet may contain the term "Pre-Kindergarten for all", which pertains to the characteristic of Smart People (since one feature of this SCC is "21st century education"). Humans can intuitively make the connection upon reading the content of the ordinances or tweets and the SCC features. However, to automate the mapping, this knowledge needs to be explicitly fed to the algorithm, which is done by deploying concepts, properties and relationships in the CSK source WebChild [5].

Likewise, CSK also helps in filtering out unwanted tweets from among the millions of tweets initially obtained through the use of terms in SCC Domain KBs (derived using CSK and SCC sources). Accordingly, after finding relevant tweets and mapping them to their respective ordinances, our next step is sentiment analysis. Here we conduct a sentiment polarity-based classification of tweets, using SentiWordNet [7] derived from the CSK source WordNet [6], in order to gauge public opinion. CSK plays a role here by connecting relevant terms to appropriate sentiment words, thereby capturing subtle human judgment. The outcome of the sentiment analysis can be used to provide feedback to urban agencies on their policies.

Our research is thus potentially useful to urban agencies for assessment in relevance to Smart Cities. By the identification of SCCs addressed in local laws, information can be provided on how well urban policies are aiming towards Smart City development. Further, the knowledge discovered by mining data on Twitter and mapping ordinances to tweets is helpful to urban agencies to evaluate public contentment. This would help them determine their Smart City public appeal and accordingly enforce appropriate legislation to enhance urban management as needed. This work thus falls under the characteristic of Smart Governance (or Smart Government), which embodies the involvement of the public in decision-making and transparency of the whole governing process.

The rest of this paper is organized as follows. Section II overviews related work in the area. Section III describes our approach on mapping ordinances and tweets through SCCs. Section IV focuses on sentiment analysis of tweets. Section V summarizes our experimental evaluation, while Section VI presents discussion and challenges. Section VII states the conclusions and ongoing work.

II. RELATED WORK

Recently, there has been significant interest in Smart Cities, and a number of developments have occurred in this field. For example, in Barcelona, buses are now configured to run on optimal routes for better power consumption [2]. In Amsterdam, there are canal lights that adjust their brightness automatically depending on how often they are used by pedestrians [2]. These initiatives fall under the category of Smart Mobility, whereas other works in this area consider decisions for autonomous vehicles [8] or saving trips in delivery and pickup [9]. Copenhagen (ranked number one in the 2017 Smart City Index) has buildings with sensors for air quality and climate control and smart meters for intelligent control of energy consumption [10]. Health and safety issues, considering AQI (Air Quality Index) standards for human health, are addressed in [11]. Such works bridge the areas of Smart Living and Smart Environment. Further research affects the Smart Environment domain by relying on cloud computing solutions rather than on-site storage solutions for mid-sized data centers [12]. In Smart Economy, cost savings and profit distribution has been studied [9], as well as important issues for business and household taxation [13] and data center cost savings [14]. Another study [15] considers urban policy data, modeling it using data warehousing and XML databases, and conducts preliminary data mining using classical techniques such as association rules and decision tree classifiers [4]. This heads towards Smart Governance.

Various studies have been conducted on data mining from social media. An important piece of work in this field focuses on unsupervised modeling and trend detection in social media [16]. In an experiment in 2015, a fuzzy-based method was used to pre-process and analyze Twitter hashtags so as to study trends in hashtag popularity [17]. However, such approaches are not feasible in our current project, which involves mapping tweets to a pre-existing set of ordinances without existing training data. Another example of research in this field focuses on supervised classification. While standard methodologies exist [4], these require massive training sets to account for the short length and variability of tweets. Another approach is in the TweetSift system [18], which classifies tweets by topic and uses external entity knowledge and word embeddings that are topic-enhanced. These lead to topic-specific word embeddings such that different senses of equivocal words obtain different representations. This process also comes with the assumption that the knowledge bases provide very relevant signals regarding different entities, such as particular uses on Twitter. In contrast to this, our approach uses generic Commonsense Knowledge and does not need labeled training data. In addition, works by other researchers have not considered the setting of ordinances and Smart City Characteristics. Hence, our work is among the first to explore this area.

The use of Commonsense Knowledge in the domain of Smart Cities is in its early stages as well, with important opportunities to enhance a number of research areas. Works in the field [19, 20] describe the use of CSK for various tasks such as: enhancement of autonomous vehicles that can contribute to Smart Mobility; and machine translation in writing aids that have the potential to impact the Smart People characteristic. The use of CSK would enhance the reasoning capabilities of machines, which would enable them to solve several challenging problems. The use of such CSK-enabled machines in various aspects of automation in Smart Cities would make them even smarter. Our work in this paper addresses the theme of Commonsense Knowledge in Smart Cities. We make use of CSK in various parts of the ordinance and tweet mining process, including mapping as well as sentiment analysis.

III. APPROACH FOR MAPPING

A. SCC Based Mapping Process

Our proposed approach for mapping tweets and ordinances to each other through SCCs is illustrated in Fig. 2. The main source for the SCCs in our approach is a technical report [2] published by the Vienna University of Technology. This report describes Smart Cities as having six different Smart City Characteristics: Smart Mobility, Governance, People, Living, Economy and Environment.

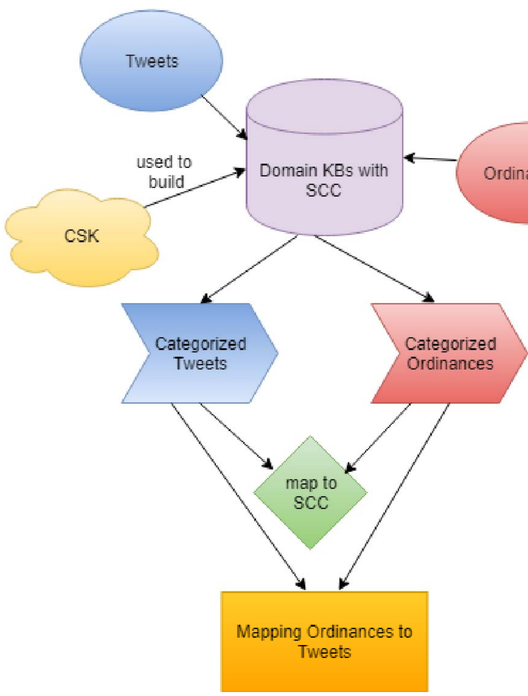

Fig. 2. Illustration of the Mapping Process

Our own work in this research project is on transparency in government through public involvement, which falls under the realm of *Smart Governance*. Hence, we enumerate its specifics, just to give an example of SCC features.

Smart Governance

- Transparency in government
- Optimization of public service and administration
- Direct involvement in public policies
- Citizen participation
- Positive and open communication channel with citizens
- More informed decisions by feedback and engagement

B. Role of Commonsense Knowledge

Considering the SCC Smart Governance elaborated herewith, if ordinances or tweets relate to any of the specifics described above, they are most likely related to Smart Governance. However, most tweets and ordinances may not directly contain these specific features such as *citizen participation* in their descriptions and hence machines may not be able to recognize them. Humans, through common sense, will be able to spot such features and thus infer that an ordinance or tweet maps to one or more SCCs. Hence, to automate the mapping process, we make use of a commonsense knowledge (CSK) repository called WebChild [5]. This consists of various commonsense concepts derived from vast quantities of data available on the Web, as well as their properties and relationships. A snapshot of the WebChild browser is given in Fig. 3, which depicts results related to the concept *nation*.

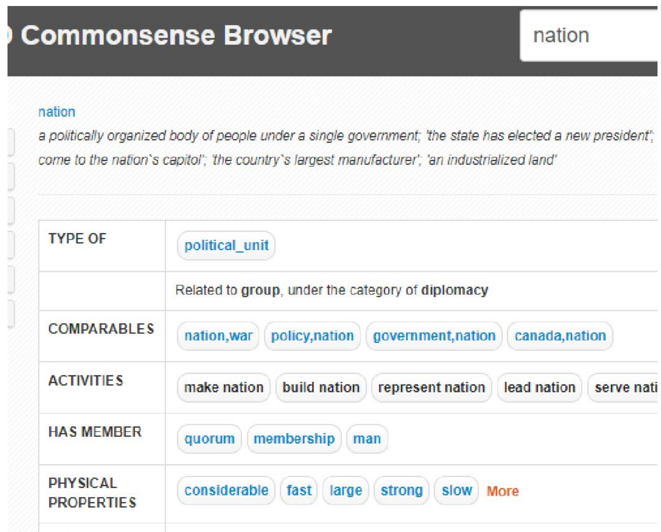

Fig. 3. Partial Screenshot of WebChild

In order to induce a mapping of ordinances and tweets using Smart City Characteristics, we construct domain specific knowledge bases (Domain KBs) using various SCC sources guided by CSK. Sample Domain KBs for three SCCs are shown in Fig. 4. WebChild is the primary source for CSK in our research. We also use the lexical database WordNet [6] while building the KBs. By using these SCC Domain KBs, CSK principles are used to find connections between terms x in every ordinance text O and SCCs. This is denoted by $C(O, x)$. For example, if the ordinance text includes the term *unemployment*, the CSK concepts are useful to find its semantic relatedness with the SCC Smart Economy. This is because the terms in the corresponding SCC Domain KB have been derived from classical SCC sources as well as the CSK source WebChild to help make the connection. These Domain KBs guide the mapping of ordinances and tweets through SCCs in our algorithm, as outlined next.

C. Mapping with Single or Multiple SCCs

An ordinance or a tweet can relate to one or more SCCs. Accordingly, each ordinance / tweet is mapped to SCCs with the most relevant number of features. Mapping to different

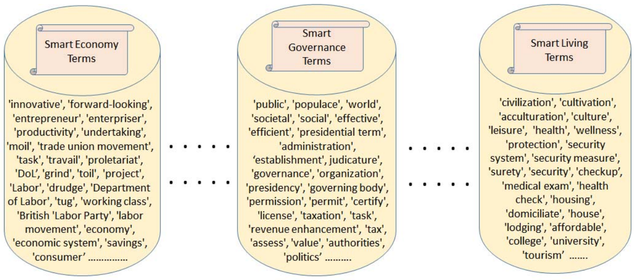

Fig. 4. Sample SCC Domains

SCCs is based on the weights of relevant terms assigned to them. For instance, if an ordinance/tweet has two terms related to Smart Economy and one related to Smart Governance, the ratio of mapping Smart Economy:Smart Governance is 2:1. If an ordinance or tweet consists of a significant number of terms related to a single SCC, the other terms may be ignored as they may not be relevant to the intention of the tweet. This is determined by a threshold, which can be adjusted as needed.

Based on this discussion, our proposed algorithm for mapping ordinances tweets through SCCs is as follows.

ALGORITHM 1: ORDINANCE-TWEET–SCC MAPPING

for each SCC S_i **do**:
    build domain KB K_i
$A \leftarrow \emptyset$
for each ordinance O_i **do**:
    for each SCC S_j **do**:
        $L_{i,j} \leftarrow \sum_{x \in K_j} C(O_i, x)$
    $A \leftarrow A \cup \{(O_i, S_j) \mid j = \text{argmax}_j L_{i,j}\}$
for each tweet T_i **do**:
    for each SCC S_j do:
        $M_{i,j} \leftarrow \sum_{x \in K_j} C(T_i, x)$
    $A \leftarrow A \cup \{(T_i, S_j) \mid j = \text{argmax}_j L_{i,j}\}$
$\theta \leftarrow \{(O_i, T_k) \mid \exists S_j : (O_i, S_j) \in A \wedge (T_k, S_j) \in A\}$
return θ

Hence, this algorithm incorporates the transitive property such that if ordinances map to one or more SCC(s) and tweets map to the same SCC(s), then the ordinances broadly map to the tweets. This can thus be used to determine the actual ordinance to tweet mapping on a broad scale, which is the final output of our mapping process. Note that we do not deal with the finest levels of granularity in this ordinance-tweet mapping as of now. We maintain a more generic connection at the level of relevance to SCCs, since an important aspect of this work entails heading towards Smart Cities.

IV. SENTIMENT ANALYSIS OF TWEETS

A. *Process of Sentiment Analysis*

Sentiment analysis of the text involves determining the emotion expressed in a particular piece of writing. The specific task of sentiment polarity classification focuses on assessing whether the sentiment is "positive", "negative" or "neutral", and sometimes the extent to which it heads in that respective direction, i.e., "strongly positive" etc. This often serves the purpose of opinion mining, i.e., discovering knowledge from people's opinions or reactions. Sentiment analysis has a number of applications in different areas as follows.

- **Business:** Used by companies to obtain product, service and brand satisfaction from customers
- **Politics:** To gauge the population's interest in various political events
- **Social events:** To understand people's reactions to economic and social events in general

B. *Opinion Mining using CSK*

In our research, we conduct sentiment analysis to discover knowledge specifically with respect to opinion mining of tweets on ordinances. This is conducted after the mapping of ordinances to tweets (as explained in the previous subsection). The primary database used for Sentiment Analysis in this work is SentiWordNet [7]. The SentiWordNet source has been built for guiding sentiment classification and opinion mining. This is an enhanced version of the CSK source WordNet [6]. It groups words into synonym sets (synsets) annotated by how positive the terms are. Accordingly, words are classified as positive, negative or neutral based on polarity of terms.

In SentiWordNet, different meanings exhibited by the same word can have different sentiment scores. For example, the word *estimable* when relating to computation has a neutral score of 0.0, while the same word in the sense of *deserving respect* is assigned a positive score of 0.75. The process we deploy for sentiment analysis of tweets constitutes a semi-supervised learning method using SentiWordNet. Through this, subtle human judgment through commonsense in understanding emotions is embodied in the mining processes with specific reference to context.

C. *Algorithm for Polarity Classification*

Based on the given discussion, our proposed algorithm for sentiment analysis of tweets through polarity classification is as follows.

ALGORITHM 2: TWEET POLARITY CLASSIFICATION

for each tweet t_i **do**:
    if not (t_1 relevant according to SCC KB):
        continue (with next tweet)
    map t_i to ordinances using Algorithm 1
    $W_i \leftarrow$ set of words in t_i
    for each $w \in W_i$ **do**:
        $s_w \leftarrow$ polarity score of w in SentiWordNet
    $s_i \leftarrow \sum_{w \in W_i} s_w$
return final polarity scores s_i for relevant t_i

Based on this algorithm, we classify thousands of tweets that we obtain from Twitter. Note that the selection of relevant tweets and also the mapping of tweets to their respective ordinances is guided by CSK. The SCC Domain KBs derived from WebChild and WordNet serve to filter out unwanted tweets as a first step, followed by the mapping of tweets to relevant ordinances using SCCs as a next step. The results of our polarity classification using this approach are presented in the experimental evaluation section, after the results of ordinance and tweet mapping through SCCs.

V. EXPERIMENTAL EVALUATION

A. Ordinance to SCC Mapping

The source of the ordinances in our experiments herewith is the NYC metropolitan legislative council website [21]. A partial snapshot of this is seen in Fig. 5. Consider the following sample ordinance obtained from this NYC source.

Fig. 5. NYC Ordinance Website

Sample Ordinance: *A Local Law to amend the administrative code of the city of New York, in relation to recycling outreach, education and enforcement; and to repeal subdivisions d and e of section 16-305 and section 16-311 of the administrative code of the city of New York, relating to source separation of recyclable materials and recycling centers.*

The mapping of this sample ordinance to its relevant SCC(s) is shown in Table I.

TABLE I
MAPPING OF A SAMPLE ORDINANCE TO SCC(S)

Smart City Characteristics	Occurrences
Economy	0
Environment	4
Governance	0
People	0
Mobility	0
Living	1

Thus, it can be seen that we get the following mapping results for this ordinance: Smart Economy = 0%, Smart Environment = 80%, Smart Governance = 0%, Smart People = 0%, Smart Mobility = 0%, Smart Living = 20%. Hence, we observe that the given ordinance maps most closely with Smart Environment, but also to some extent with Smart Living. Based on the threshold used in this execution, it is mapped to Smart Environment:Smart Living with a ratio of 80:20, stating the outputs as percentages.

B. Tweet to SCC Mapping

We extract location-specific geo-tagged tweets posted by the general public from NYC and map these to SCCs using the process described in the previous section. We mine approximately 5,000 tweets in the experiments shown here. An small random subset of tweets as mined before mapping (regardless of the their relevance to ordinances) is given below.

Tweet T1: *"RT @OccAware: @RachelNotley I am contacting you to tell your government does not have my support to put BC land, water and our tourist economy at risk while trashing indigenous rights"*

Tweet T2: *"RT @verge: Elon Musk made history launching a car into space. Did he make art too?"*

Tweet T3: *"Concerns about HB1319 the predatory lending bill. High interest rates, short-term loans for poor. Veterans groups social service gps have spoken against."*

Tweet T4: *"Adding plants to your home or office has so many benefits such as reducing stress and increasing productivity! Not to mention they add a personal touch to your surroundings. "*

Tweet T5: *"RT @CoachKCullen: A1 a strong non-digital learning environment is needed before you introduce digital learning environment. Bad practices can just multiply the more technology is introduced satchat"*

These tweets are subjected to mapping using Algorithm 1. Each tweet can have terms related to one or more SCC(s). Different SCC terms from the input tweets are considered and accordingly weights are assigned to each SCC. These determine the relevance of the tweets to the SCC(s) based on the threshold levels used in the experiments.

We have developed a GUI that accepts a tweet as the input and determines the closest matching SCC(s) as the output, or returns *"No matches"* if absolutely no SCC gets matched during the mapping process. Fig. 6 shows an example of tweet to SCC mapping using this GUI.

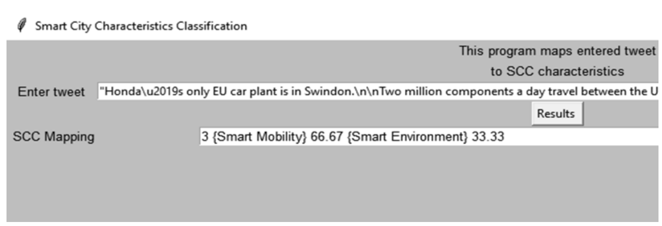

Fig. 6. Screenshot of GUI for Tweet to SCC Mapping

C. SCC Mapping Assessment

After the mapping is conducted, its outcomes are assessed by domain experts from the department of Earth and Environmental Studies. Ground truth is defined by the experts such that a mapping is identified as correct if is agrees with the judgment of the expert and incorrect otherwise. For example, if the system identifies the mapping as Smart Economy:Smart Governance with a ratio of 60:40 and the expert agrees, this is considered to be a correct mapping. On the other hand, if the expert disagrees with the single or multiple SCC(s) identified in the mapping, or considers their ratios to be highly inappropriate, this is treated as an incorrect mapping. Assessment is then conducted using the standard *Precision* metric [4] as the ratio of correct mappings to all mappings. Thus, *Precision = Correct Mappings / (Correct Mappings + Incorrect Mappings)*. Based on this evaluation, Fig. 7 shows a chart summarizing the domain expert assessment of tweet and ordinance mappings with respect to the SCCs.

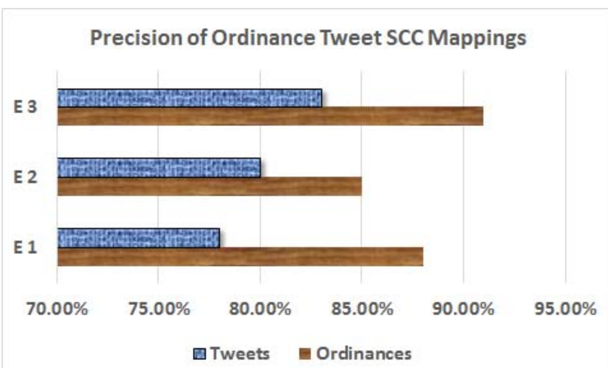

Fig. 7. Summary of Mapping Assessment

The evaluation of the mapping results obtained can be interpreted as follows. With respect to Expert 1, around 88% of the ordinances are correctly mapped to their respective multiple SCCs, while approximately 78% of the tweets are appropriately mapped. Hence, the ordinance to tweet mapping precision on a broad range would be at best 78%, considering their mutual connection with multiple SCCs, as per the ground truth defined by this expert. Likewise, on the whole, the ordinance to multiple SCC mapping precision is observed to be in the range of the higher 80s, while that of the tweets to multiple SCC is around the lower 80s. Thus, the ordinance to tweet mapping precision through mutually relevant SCCs would be at best in the lower 80s.

Note that this is an enhancement over our early work [22], where we considered mapping of ordinances and tweets to single SCCs only, based on the closest match. It was concluded therein [22] that the mapping precision needs to be higher to conduct opinion mining of tweets with respect to relevant ordinances and SCCs. Thus, we proposed enhanced algorithms in this paper and also refined domain KBs on SCCs with more intricate CSK concepts.

D. Ordinance to Tweet Mapping Output

The precision ranges obtained in the ordinance to SCC and tweet to SCC mappings are considered acceptable by our domain experts to proceed with further work. Hence, it is feasible to use these mappings in order to output the ordinance–tweet mappings. We present examples herewith of broadly related correct ordinance to tweet mappings as determined through their mutual SCC connections.

Mapping Example 1: In this excerpt, both the ordinance and tweet map to Smart Economy.

Ordinance: *A Local Law to amend the administrative code of the city of New York, in relation to authorizing an increase in the amount to be expended annually in seven business improvement districts and two special assessment districts.*

Tweet: *@DowntownNYC is one of NYCs largest Business Improvement Districts, which works to enhance the quality of life in #LowerManhattan. Attend our #YLG #SecretsofSuccess on 5/11 ft. @JessLappin for an exclusive talk about what it's like to lead a #BID.*

Mapping Example 2: Here, the ordinance and tweet both map to Smart Mobility as well as Smart Living.

Ordinance: *A Local Law to amend the administrative code of the city of New York, in relation to parking violations issued for the failure to display a muni-meter receipt.*

Tweet: *The new NYC Parking Ticket: Pay or Dispute app makes the process of paying or disputing a violation easier and faster. #nycpayordispute*

Likewise, various such ordinance-tweet mappings through their semantic relatedness with the SCC(s) set the stage for sentiment analysis of tweets to analyze public opinion.

E. Results of Sentiment Analysis on Tweets

In the experiments shown here, we use tweets from the NYC region after filtering out unwanted ones guided by SCC domain KBs derived through CSK. We map these to ordinances through our proposed mapping approach in Algorithm 1. Sentiment analysis of these tweets is then conducted using Algorithm 2. Examples of these are summarized next.

NYC Tweet Example 1:

"#FairFares is just common sense. What it will do is level the playing field so every #NYC resident can access the @MTA. Pretty simple idea. Not only is it the right thing to do but it will also help to grow our economy. #nowisthetime"

This tweet maps to Smart Mobility and Smart Economy. For this tweet, the net score is 0.43. It obtains a positive score of 0.72 and a negative score of -0.29.

NYC Tweet Example 2:

"#NewYork's statue of Liberty was RED before pollution turned it green"

This tweet maps to Smart Environment. For this tweet, the net score is -0.21. Its positive score is 0.15, while its negative score is -0.36.

Likewise, based on several examples, the combined results of sentiment analysis on tweets are obtained. These are illustrated in Fig. 8. The pie chart in this figure summarizes the overall public reactions as determined from tweets across all

SCCs. Since the percentage of positive tweets is the greatest among these, we can conclude that the people of NYC seem to approve of the concerned policies (to a greater extent than complaining about them or being neutral). However, positive sentiments are expressed by less than half the public, which means that there is potential for improvement with respect to heading towards a Smart City.

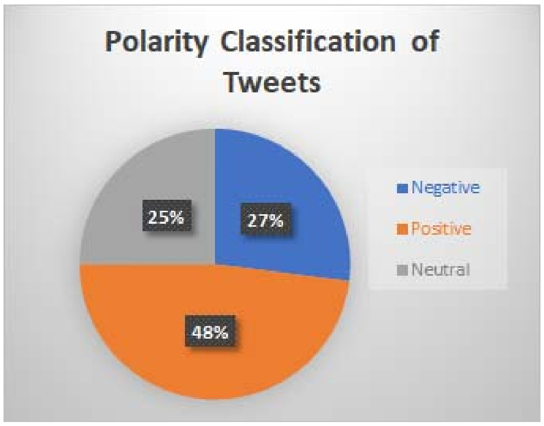

Fig. 8. Polarity Classification of Tweets on all SCCs

Additionally, we analyze specific public contentment for each SCC and obtain the results as shown in Table II. These numbers help urban agencies assess how satisfied people are on matters pertaining to each individual SCC. For example, here the public seems least satisfied with issues related to Smart Environment. Thus, feedback can be provided to urban agencies that they need to address policies to make the environment smarter, e.g., develop more energy efficient systems.

TABLE II
PUBLIC CONTENTMENT FOR EACH SCC

Aspect	Positive Polarity
Smart Economy	47%
Smart Mobility	48%
Smart Environment	33%
Smart Governance	45%
Smart People	52%
Smart Living	56%

On the whole, public contentment seems fine considering the various SCCs in our sample. Hence, from this analysis, we can infer that based on our data used and the accuracy of our results that the metropolitan region of NYC seems to tend fairly well towards being a Smart City, with scope for further enhancement. A summary of these results can be provided as feedback to urban agencies to help them make decisions in outlining further legislative policies.

VI. DISCUSSION AND CHALLENGES

The overall evaluation in this paper reveals that many ordinances and tweets get correctly mapped to each other on a broad level of semantic relatedness. Domain experts confirm that this is acceptable for current use, although it could get better in the future. Accordingly, upon closer observation we find that a few mappings of ordinances and tweets are imprecise with respect to their SCC identification. Tweet mapping particularly needs further refinement. Moreover, some ordinance–tweet mapping outputs are found to be imprecise, in spite of their correct mutual SCC connection being identified in the ordinance–SCC and tweet–SCC mappings.

In order to address these issues, we need further research with respect to the refinement of the content in the SCC Domain KBs and also the levels of granularity in ordinance to tweet mappings. SCCs can still be used as a mutual connection; however, this could be done with more specific features that refer to their actual details. Direct mapping of ordinances to tweets can be a next step addressing intricate natural language and other aspects. Note that our proposed SCC-based mapping approach substantially narrows down the sample space for potential direct ordinance–tweet mappings. In the absence of this approach, the large quadratic number of pairs is very challenging, given ordinances and tweets in the order of thousands and millions, respectively.

Sentiment analysis of tweets has also been assessed by our domain experts through a process similar to that of mappings (details not shown herewith). The polarity classification is found to be accurate for the most part, i.e., if the overall classification of a tweet is positive as identified by our approach, it indeed expresses a positive sentiment with respect to ground truth defined by experts etc. Due to this, we have considered it feasible to use the results of sentiment analysis for summarizing the public reactions on ordinances as shown in Figs. 10 and 11. We can therefore use these outcomes to provide recommendations to urban agencies. However, it is to be noted that we have found a few incorrect polarity classifications. This can be attributed to the fact that tweets in general do not follow a systematic grammar structure, making it difficult to derive semantic patterns in some cases.

Based on our work, we outline the following challenges in dealing with tweets for mapping and polarity classification.

- The usage of informal language in tweets makes it difficult for pre-processing using basic Natural Language Processing (NLP) techniques.
- There is rampant usage of acronyms in Twitter, which do not relate to standard vocabulary sources.
- Tweets show ambiguous characteristics with respect to NEE (Named Entity Extraction) as well as NED (Named Entity Disambiguation).

Addressing these challenges and the other ongoing tasks is nontrivial and constitutes further work. Ongoing research involves addressing these issues for enhanced knowledge discovery from ordinance and tweets in the future.

VII. CONCLUSIONS

In this paper, we mine ordinances and their public reactions to gauge how well a given urban region tends towards a

Smart City. The novelty of this work includes: (1) being among the first to address ordinance mining, especially for Smart Cities; (2) implementing single and multiple SCC-based ordinance to tweet mapping; and (3) deploying commonsense knowledge in mining (for tweet selection, mapping processes and polarity classification). The overall challenges include: (1) dealing with intricate natural language in ordinances and tweets; (2) handling big data of the order of thousands and millions respectively; and (3) considering further issues in tweets, e.g., acronyms and ambiguity.

We evaluate our work with real data from NYC sources. The ordinance to SCC mapping precision is found to be in the higher 80% range while tweet to SCC mapping precision is in the lower 80s on an average. The polarity classification of tweets suggests that the majority of the public is satisfied with the topics that the ordinances cover. Yet, positive sentiment amount to lower than 50%, implying scope for improvement. Through this analysis, feedback can be provided to urban agencies for policy decisions. This work contributes to Smart Governance, by public involvement entailing transparent decision-making.

Our ongoing work seeks to enhance the mapping precision and to address finer levels of granularity in the ordinance–tweet mapping. We aim to improve tweet sentiment analysis, incorporating advanced concepts in NLP and Machine Learning. The long-term goal of our research is to aid urban regions in enhancing legislation related to Smart Cities. This constitutes multidisciplinary work in Artificial Intelligence and Environmental Management.

ACKNOWLEDGMENTS

We thank Robert Taylor, Clement Alo and their PhD students from the Department of Earth and Environmental Studies at Montclair State University (NJ) for useful inputs from a domain-specific angle. Early work in this area started when Aparna Varde was a visiting researcher at Max-Planck-Institut für Informatik (Saarbrücken, Germany) in Gerhard Weikum's Databases and Information Systems group. Gerard de Melo's research at Rutgers University (NJ) is funded in part by ARO grant no. W911NF-17-C-0098 (DARPA SocialSim program).

REFERENCES

[1] The IEEE Smart Cities Technical Community. https://smartcities.ieee.org/, 2018.

[2] TU Wien (Vienna University of Technology). European smart cities. Technical report, Vienna, Austria, 2015.

[3] Smart City Consortium. Background. https://smartcity.org.hk/index.php/aboutus/background.

[4] Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco, CA, 2011.

[5] Niket Tandon, Gerard de Melo, Fabian Suchanek, and Gerhard Weikum. Webchild: Harvesting and organizing commonsense knowledge from the web. In *ACM International Conference on Web Search and Data Mining*, WSDM, pages 523–532, 2014.

[6] George Miller and Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, MA, 1998.

[7] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *International Conference on Language Resources and Evaluation, LREC*, May 2010.

[8] Katherine Rose Driggs-Campbell, Victor Shia, and Ruzena Bajcsy. Decisions for autonomous vehicles: Integrating sensors, communication, and control. In *International Conference on High Confidence Networked Systems*, HiCoNS '14, pages 59–60. ACM, 2014.

[9] Yang Yu, Qi Lou, Jiafu Tang, Junwei Wang, and XiaoHang Yue. An exact decomposition method to save trips in cooperative pickup and delivery based on scheduled trips and profit distribution. *Computers and Operations Research Journal*, 87:245 – 257, 2017.

[10] EasyPark Group. The 2017 smart cities index. https://easyparkgroup.com/smart-cities-index, 2017.

[11] Xu Du, Onyeka Emebo, Aparna Varde, Niket Tandon, Sreyasi Nag Chowdhury, and Gerard Weikum. Air quality assessment from social media and structured data: Pollutants and health impacts in urban planning. In *IEEE International Conference on Data Engineering (ICDE) - Workshops*, pages 54–59, 2016.

[12] Michael J. Pawlish, Aparna S. Varde, and Stefan A. Robila. The greening of data centers with cloud technology. *International Journal of Cloud Applications and Computing*, 5(4):1–23, October 2015.

[13] G. Leef. *Smart Economics: Commonsense Answers to 50 Questions about Government, Taxes, Business and Households*. Foundation for Economic Education, 2007.

[14] Michael Pawlish and Aparna Varde. Free cooling: A paradigm shift in data centers. In *IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 347–352, 2010.

[15] Xu Du, Diane Liporace, and Aparna Varde. Urban legislation assessment by data analytics with smart city characteristics. In *IEEE Ubiquitous Computing, Electronics and Mobile Communications Conference (UEMCON)*, pages 20–25, 2017.

[16] Liangjie Hong and Brian D. Davison. Empirical study of topic modeling in twitter. In *Workshop on Social Media Analytics*, SOMA '10, pages 80–88. ACM, 2010.

[17] L. A. Zadeh, A. M. Abbasov, and S. N. Shahbazova. Analysis of twitter hashtags: Fuzzy clustering approach. In *2015 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC)*, pages 1–6, 2015.

[18] Quanzhi Li, Sameena Shah, Xiaomo Liu, Armineh Nourbakhsh, and Rui Fang. Tweetsift: Tweet topic classification based on entity knowledge base and topic enhanced word embedding. In *ACM International on Conference on Information and Knowledge Management*, CIKM, 2016.

[19] Priya Persaud, Aparna Varde, and Stefan Robila. Enhancing autonomous vehicles with commonsense: Smart mobility in smart cities. In *IEEE International Conference on Tools with Artificial Intelligence (ICTAI) - Smart Cities Workshop*, pages 1008–1012, 2017.

[20] Niket Tandon, Aparna S. Varde, and Gerard de Melo. Commonsense knowledge in machine intelligence. *ACM SIGMOD Record*, 46(4):49–52, 2017.

[21] The New York City Council. Legislative research center web page. http://legistar.council.nyc.gov/, 2018.

[22] Manish Puri, Xu Du, Aparna Varde, and Gerard de Melo. Mapping ordinances and tweets using smart city characteristics to aid opinion mining. In *The Web Conference, WWW Companion Volume, Satellite Track on Social Sensing and Enterprise Intelligence SSEI*, pages 1721–1728. W3C, 2018.