W
Western Illinois University

Problem Identification in Community Economic Development via Natural Language Processing

Adee Athiyaman[1]

Introduction

During June 2021, a news reporter contacted the author and asked for explanations about a county that has experienced increases in sales-tax revenues during the Covid-19 shutdown; a deviation from the norm for the region[2]. The reporter believed that revenue increases are from legalized marijuana sales in Illinois, but was aware that other counties in the region did not experience increases in tax revenues. In decision making literature, the inference of the reporter is termed 'bottom-up decision making'; in other words, the reporter relied on an objective attribute (for example, that the state has legalized recreational marijuana sales) to frame his inference about changes to county tax revenues.

An economist with considerable experience in economic theory may use the 'shift in demand curve' concept[3] to argue that factors such as mobility of county residents (intra-county travel) could influence increases in retail sales, and thus change a county's tax revenues[4]. This is the 'top-down' approach to decision making; the economist knows, based on his knowledge of the subject and experience in observing events like the above, what are the probable causes for increases in a region's tax revenues.

The case illustrates that someone has to notice a problem or opportunity before solving it. How does one identify problems to study (research) in community economic development (CED)? Is it possible to use artificial intelligence (AI) technologies such as

---

[1] Professor, Illinois Institute for Rural Affairs, Western Illinois University

[2] This deviation from expectation or gap is often called an opportunity or benefit; in general, a positive or desired gap from expectation is called an opportunity and a negative gap a problem.

[3] Demand curve for a good or service would shift when the ceteris paribus assumption relating price and quantity variables is relaxed.

[4] See, https://www.galesburg.com/story/business/economy/2021/06/21/galesburg-showed-retail-gains-2020-despite-covid-while-other-communities-saw-sharp-declines/7067233002/

natural language processing (NLP) to gain insights into CED questions that require answers? This paper addresses these and other related questions.

Conceptual Model

The lexical definition of CED is: a community – a group of people in a geographical area with similar values or attitude towards life, and economic development – the modifier, 'economic', alludes to the type of development that is the focus of the community. The principles of categorization suggest that as a concept, CED is at the 'basic' or the meso level of categorization; a subordinate concept of CED would be, 'new firm starts' in the community; 'national economic development' would be the superordinate concept. Since CED processes involve both organism (people) and organizations (CED agencies)[5], CED is a living system[6].

A living system exhibits the following characteristics: open – has permeable boundaries or geographical borders, takes in inputs (for example, imports of goods and services from outside its boundary), and has critical subsystems that are integrated and act as a whole with a purpose[7] (see Table 1).

Salient Subsystems for CED

One of the major functions of CED is information provision to stakeholders, for example, informing potential entrepreneurs about human capital in the community[8]. The "input transducer" subsystem monitors the external environment for threats and opportunities and transmits the scanning data to the "decoder" subsystem. The structure of the decoder subsystem could be a centralized committee of lawmakers or decentralized project teams working on information pertaining to different capital stocks such as human, cultural, social, etc. The output of the decoder subsystem, status report on capital stock in the community, is transmitted to the "decider" subsystem.

The decider subsystem then develops plans and makes decisions that guide the CED efforts. Two subsystems provide information from the CED system back to the environment: the "encoder" subsystem converts CED plans into operational tactics, such as creating more recreational parks in the community, and the "output transducer" subsystem communicates these tactics to relevant audiences, for example, potential residents are informed about recreational amenities in the community and businesses are made aware of the skills and attributes of the human capital in the region.

---

[5] See, for example, the organization responsible for economic development in McDonough County, IL; https://www.maedco.org/

[6] Miller, J. G. (1978). Living Systems. New York: Mc-Graw Hill.

[7] Athiyaman, A. (2011). Biofuels: What is on Practitioners' Minds? Academy of Marketing Studies Journal, 15(1), 1-21.

[8] Athiyaman, A. & Runge, A. (2020). Implementing CED: Normative Prescriptions with Applications to Rural Illinois. Research Brief, 2(1). Available online: http://www.iira.org/wp-content/uploads/2020/03/EDImplementation_Jan2020-1.pdf

Table 1: Salient Subsystems of CED

Subsystem	Function
Input transducer	It senses environmental signals and brings information into the CED system.
Decoder	The environmental information is decoded into a language for internal use.
Decider	It is the executive; it plans, directs, and controls the activities of the entire system.
Encoder	Converts CED plans into operational tactics.
Output transducer	Implements operational plans; for example, public information campaigns are implemented using both "paid" and "owned" media.

In summary, the subsystems listed in Table 1 assist in the survival of the CED system; the decider receives information inputs from the input transducer and the decoder subsystems. These inputs inform the construction of strategic and operational plans, which get implemented by the encoder and the output transducer subsystems.

The implication for conceptualizing CED as a living system is that we have a framework to categorize research problems or opportunities in the study area. For example, the generic need of CED practitioners to understand the impact of changes in public policy on a community can be categorized as a "decoder" problem; a solution would be to analyze relevant data and gain insights on policy impacts.

Table 2 suggests that problems are not perceived directly, only their symptoms are perceived through sensors. A problem can be a change from a routine situation to a deteriorating one; for instance, the Covid-19 pandemic has changed the way people live and work[9]; the problem could also be a shift from a recovering situation to a deteriorating one; for instance, Covid-19 numbers have started to climb again in the nation, for the fourth time since March 2020[10]. A situation can have a single problem or it can have multiple issues; for instance, a community's image, defined as residents' beliefs about the community, could be negative because of a single problem such as "poor infrastructure" or the image could be poor because of multiple

---

[9] Athiyaman, A. (2020). The economic impact of Covid-19 on Illinois Counties, Research Brief, 2(9), May 4, 1-18.
[10] https://www.aarp.org/health/conditions-treatments/info-2021/covid-4th-wave.html.

issues in the community (high crime rate, lack of policing, etc.). Finally, a situation could signal potential problem; an example would be low childhood vaccination rates in the community could signal potential problems such as measles outbreaks.

Table 2: Problems, Symptoms, and Sensors

Problem ⟶	Symptoms ⟶	Sensors
Routine ⟶ Deteriorating	Speed of change	
Recovering ⟶ Deteriorating	Number and variety	Practitioners and Applied Researchers
Single Issue	Trajectory	
Multiple Issues	Bifurcation points	
Potential Issue(s)	Absence of an event	
Weakness in a plan		
Weakness in action		

Table 2 shows that problems are experienced as symptoms and the latter are sensed or perceived by us, practitioners or applied researchers. Speed of change and number and variety of changes to a situation are easily recognized; for instance, the spread of Covid-19 in communities is easily recognized by looking at the Covid-19 tracking data provided by the John Hopkins University[11]. Similarly, the trajectory of the Covid-19 spread could be gauged by being sensitive to changes in the infection rates. A bifurcation point represents a temporary state that can evolve into one of several stable states. Again, using the Covid-19 pandemic as an example, the pandemic could affect travel and tourism: people may delay nonessential travel, or may avoid travel completely and rely on virtual tours such as the Vatican's tour of the Sistine Chapel[12] for sensual gratification (a salient drive for travel). The update rate or changes in tourist numbers have to be monitored to detect (predict) shifts in the 'nature state'. Finally, sometimes, an important piece of

_____

[11] https://coronavirus.jhu.edu/us-map

_____

[12]
https://m.museivaticani.va/content/museivaticani-mobile/en/collezioni/musei/cappella-sistina/tour-virtuale.html

information is the absence of an event; the negative event can often be sensed as a disconfirmed expectation.

Methodology

The focus was on statements[13] about CED problems that have been reported in the popular press. The NewsBank[14] database was used to identify publications that contained the terms "community development" in the headline and the terms "problem or opportunity" in the lead paragraph; search dates were restricted to the period 2000 – 2021. In all, 261 publications were identified (Figure 1).

Figure 1: Data: Source, Search Filters, and Results
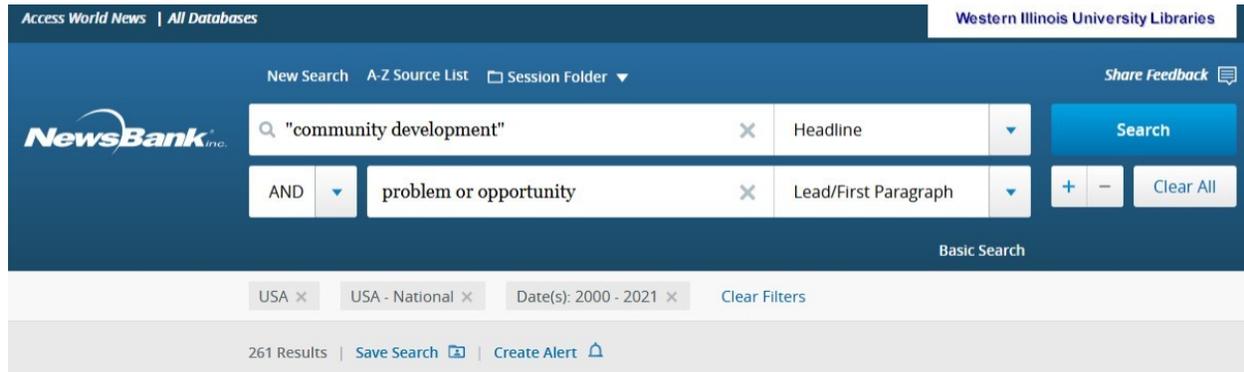
These 261 publications were combined into a corpus and was analyzed as follows: first, the corpus was tokenized (broken down by terms), and tagged according to their lexical category; the lexical categories were formed into chunks of noun phrases; next, conceptual entities were identified, noun phrases; finally, relations among the concepts (named entities) were explored using a simple predicate and the resulting sentences were presented as areas for research and inquiry. Table 3 is a summary of methods used for information extraction and data analysis.

---

[13] Statements are assertions of propositions – concepts in relations. Cognitive science suggests that a state of affair is expressed in statements (sentences) that are codified as propositions in one's thought. Schematically, State of Affair ☽ Sentences ☽ Propositions (conceptual relations in mind).

[14] A newspaper database; see https://www.newsbank.com/about-newsbank.

Table 3: Information Extraction: Tasks, Methods, and Tools

| Task	Method	Tools
Compiling a corpus of 261 popular press articles, all in pdf format (see Figure 1).	Merge 261 pdf files into a single document.	Third-party python library, pyPdf.
Split the corpus into terms and tag their part-of-speech.	Tokenize sentences in the corpus and identify their lexical categories.	The NLTK algorithm; the python implementation of NLTK was used.
Tagging of tokens.	Segment at the chunk level. For example, the multi-token, "the homeless population", can be chunked as a noun phrase consisting of a determiner, an adjective, and a noun.	RegexpParser, a Python algorithm[15].
Relation extraction.	Triples of the form X, α, Y, where X and Y are concepts (for example, housing and income) and α is the string of words that intervenes between X and Y. Based on Table 2, the verbs worse, decline, and fall were used as links.	The NLTK algorithm.
Syntactic "Trees" to represent conceptual relations.	Relations extracted are portrayed as trees with S as the parent and noun phrase, etc. as children.	NLTK draw algorithms.

---

[15] He, K., Hong, N., Lapalme-Remis, S., Lan, Y., Huang, M., Li, C., & Yao, L. (2019). Understanding the patient perspective of epilepsy treatment through text mining of online patient support groups. Epilepsy & Behavior, 94, 65-71.

Results

The top 50 concepts mined from the publications include 'housing' – the most mentioned concept, and 'money' – the least mentioned noun phrase (Figure 2); 'transportation' and 'transit' were mentioned more often than 'business' and 'industry'.

Figure 2: Salient Concepts (Noun Phrases)
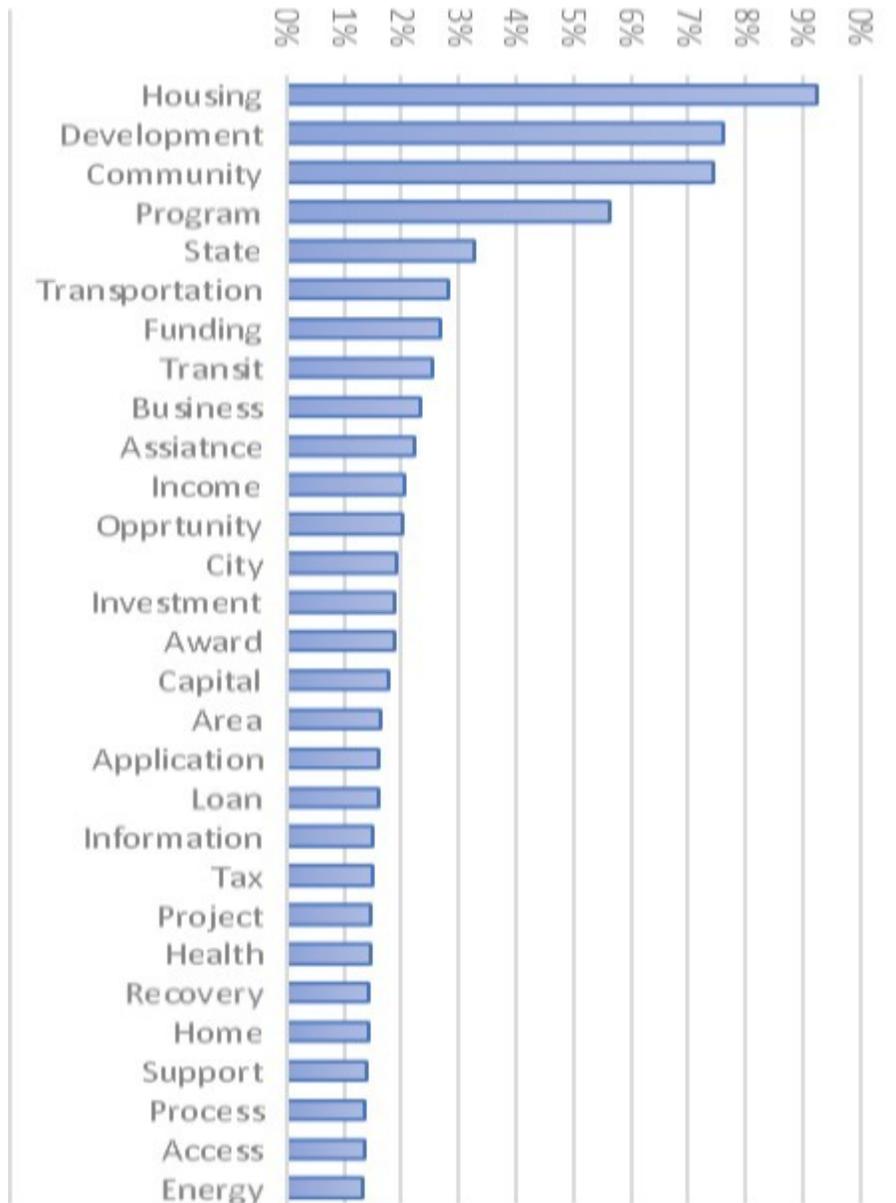
The conceptual relations highlighted "deteriorating" problems:

> Despite recent declines in home prices, the nation faces a huge shortfall of decent, affordable housing.

Single issues:

> In fact, in many communities, some of the other subsidized developments are in as bad a shape, or worse, than much of the public housing stock.

Multiple and Potential Issues:

> Specific housing hazards include exposure to allergens that may cause or worsen asthma, lead-based paint hazards, mold, and excess moisture and indoor air quality.

Figure 3 is a schematic representation of the "deteriorating" problem statement given above.

Figure 3: Tree Representation of Problem
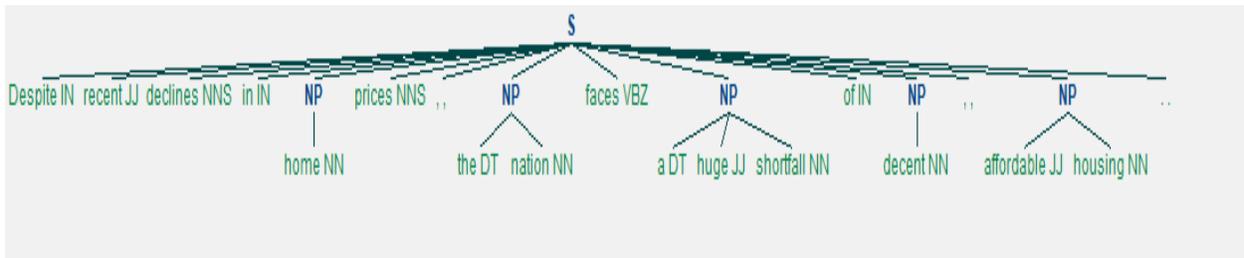
To facilitate the use of this natural-language-processing approach to CED problem detection, an interactive computer application is offered. Labeled "CEDProb.EXE", the application can be executed in Windows OS. Interested readers can download the associated zipped file, which contains the application and the data file, unzip and run the software. The outputs include a list of 50 salient noun phrases (concepts) and a printout of contextual relations (problem statements).

Finally, Table 4 shows the author's classification of a few statements into the five CED functions. Note that the functions are assumed to be oblique or correlated; put simply, a problem could be clustered into more than one CED function.

Table 4: (Problem) Statements Clustered into CED Functions

Input Transducer
Despite recent declines in home prices, the nation faces a huge shortfall of decent, affordable housing.

Output Transducer
Specific housing hazards include exposure to allergens that may cause or worsen asthma, lead-based paint hazards, mold, and excess moisture and indoor air quality.

Input Transducer; Decider
In fact, in many communities, some of the other subsidized developments are in as bad a shape, or worse, than much of the public housing stock.

Mississippi Center for Justice urges this subcommittee and HUD to: * all possible means to ensure that Mississippi cures the deficiencies and shortfalls in its housing programs outlined in this testimony; * reverse the diversion of housing funds to finance the expansion of the State Port at Gulfport and require Mississippi to finish housing first.

These problems fall into the following categories: Inexperience and inefficiencies of government entities administering affordable housing programs; additional requirements imposed by the CDBG statute upon government administered housing programs; and difficulty resolving issues associated with very low-income homeowners such as duplication of benefits, gap funding requirements, and heirship or other lack of clear title.

Decider; Encoder; Output Transducer
Another 1,147 units are financed with CDBG and LIHTC, which reduces Mississippi's projection of 5,823 units to 3,182, well below what is required to make up for shortfalls in small rental housing.

Decider
And consumers cut spending across the board, as they tried to rebuild savings after the shocks to their wage income and the declines in the stock market and housing values.

This Subcommittee may wish to see included performance targets which speak to the critical need to better coordinate transportation with housing, land and social equity goals - all objectives which fall within your subcommittee's jurisdiction.

Encoder
These plans fall into three categories: 1) the affordable restructuring of the mortgages to preserve homeownership; 2) rent-to purchase options through a temporary third-party owner; or 3) transitional support to new housing for owners who cannot sustain their tenure in the foreseeable future.

Output Transducer
Through the Recovery Act and FY 2009 Appropriations Acts, you were able to end the program funding shortfall that had forced HUD to sign partial-year, rather than full-year, contracts with the owners of this housing.

Note: * denotes statements that may contain additional tokens that are irrelevant for the discussions.

Summary and Conclusion

This paper highlights the use of natural language processing algorithms to mine popular press articles about CED; specifically, it highlights a methodology to extract CED problems, topics for research. A total of 261 articles were searched and 50 salient concepts were extracted; it is interesting to note that concepts such as housing and transportation appear more frequently than concepts such as industry and business.

A relational analysis of concepts using simple predicates worse, decline, and fall as conceptual links suggests that community housing is a top-of-mind issue for CED practitioners. Although exploratory in nature, the AI approach of this paper offers a promising approach to understanding the research needs of CED practitioners.