

Quantifying Luhmann

A Semi-Supervised Approach to Automatic Detection of Social Systems



Lukas
Gienapp

Leipzig University



Arno Simons

DZHW



Benno Stein

Bauhaus-Universität Weimar



Martin
Potthast

Introduction

Overview

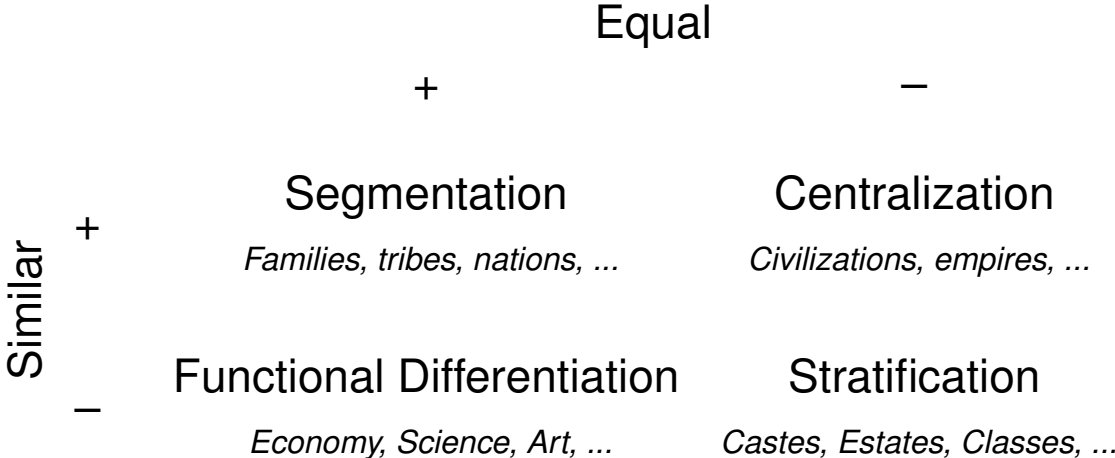
- Observations
 - Various “Frame-like” concepts in the social sciences
 - Each such concept has been claimed to capture a list of empirical instances
- Questions
 1. Can we detect the empirical instances of such concepts computationally?
 2. How much of the underlying theory of the concept do we actually need for detection?

Authors	Concept	Instances
Luhmann	Social Systems/Codes	Politics, Economy, Science, Art, Law, ...
Dryzek	Environmental Discourses	Problem Solving, Sustainability, ...
Boltanski & Thevenot	Economies of Worth	Inspired, Domestic, Fame, ...
Bourdieu	Habitus	Academic, Economic, Artistic, ...
...

Social Systems Theory

Luhmann's Functional Differentiation of Society [Roth, 2015]

Transition from a primarily “stratified” to a “functionally differentiated” society:



Social Systems Theory

Luhmann's Functional Differentiation of Society [Roth and Schütz, 2015]

System	Code	Medium	Program	Function
Science	Un-/True	Truth	Theory	Verification
Legal	Un-/Lawful	Norm	Law	Standardization
Economy	Non-/Payment	Money	Price	Distribution
Art	Innovative/Imitative	Style	Fashion	Creation
Politics	Government/Opposition	Power	Ideology	Limitation
Religion	Immanent/Transcendent	Faith	Confession	Revelation
Education	Un-/Placeable	Vita	Curriculum	Formation
Sport	Success/Failure	Achievement	Goal	Mobilization
Health	Ill/Healthy	Illness	Diagnosis	Restoration
Mass Media	Non-/Informative	Medium	Topic	Multiplication

Social Systems Theory

Luhmann's Functional Differentiation of Society

System	Book	Year	Size [†]
Science	Die Wissenschaft der Gesellschaft	1990	244
Legal	Das Recht der Gesellschaft	1993	252
Economy	Die Wirtschaft der Gesellschaft	1994	236
Art	Die Kunst der Gesellschaft	1997	449
Politics	Die Politik der Gesellschaft	2000	278
Religion	Die Religion der Gesellschaft	2000	19
Education	Das Erziehungssystem der Gesellschaft	2002	169
Sport			
Health	<i>New systems identified in the literature</i>		
Mass Media			
Moral	Die Moral der Gesellschaft	2008	731

[†] in 90 word passages

Social Systems Theory

“The biggest empirical challenge is Luhmann’s concept of meaning, which ... appears as a selection process in which a system has to choose particular options from a limited arsenal of possibilities ... Mathematical quantification finds no starting point here anymore ... Social systems theory becomes empirically operationalizable by putting the treatment of contingency—the figure of ‘determining the undetermined’—center stage.”

[Vogd, 2007:8]

→ Challenge accepted!

Classifying Social Systems

Operationalization

- Multi-class fuzzy label classification
 - categorical labels (soc. systems = classes)
 - mixed membership possible
 - fuzzy membership possible
- Levels of granularity
 - classifying documents into categories
 - classifying terms into categories
- Data
 - limited ground truth (Luhmann's books)
 - more labeled data is expensive to create
 - domain adaption required



Classifying Social Systems

Operationalization: Seed-Guided Text Classification

- **Idea:** integrate auxiliary knowledge into a topic modeling process
 - semi-supervised approach
 - each class is characterized by a set of **seed words**
 - seeds = prior information to influence convergence

- **Requirements:**
 1. a seed-guided topic modeling approach
 2. a representative seed word set

- **Application where supervised training data is unavailable:**
 - complex operationalization
 - fuzzy labeling is hard to annotate directly
 - required amount of data is high

Classifying Social Systems

Operationalization: Seed Guided Multi-Label Topic Model (SMTM) [Zha & Li. 2019]

- ❑ Semi-supervised generative classification model
- ❑ Two kinds of topics: a **general** and multiple **category** topics
 - general topic captures general semantic information of the corpus
 - category topics correspond to desired classification
 - documents are modeled as mixture of general and category topics
- ❑ Input variables: document-word matrix, seed term set for each category
- ❑ Latent variables:
 - document-category matrix, each cell is probability a topic occurring in a document
 - category-word-matrix, each cell is probability of a word being generated by a topic
- **Benefit**: model convergence can be influenced; classification possible
- **Problem**: how can we obtain a representative seed word set?

Classifying Social Systems

Selecting Seed Words

- Source: Experts → **Luhmann's books**
Pointwise mutual information ranks highly descriptive words per social system
- Amount: Few seed words with high document coverage suffice [Liu et al., 2016]

System	Seed words
Science	bewusstsein, kommunikation, beobachtungen, wissen, erkenntnis, moment, operation, ...
Legal	recht, rechtstheorie, geltung, normen, regeln, schrift, unrecht, entwicklung, juristen, ...
Economy	geld, wirtschaft, preise, knappheit, arbeit, zahlung, steuerung, unternehmen, markt, ...
Art	kunst, kunstwerk, wahrnehmung, welt, kunstwerke, ordnung, zweiter, form, formen, ...
Politics	macht, politik, ursachen, wirkungen, gewalt, sanktionen, entscheidung, stimmen, ...
Religion	religion, gott, transzendenz, immanenz, religiÃ¶se, berÃ¼cksichtigung, geist, ...
Education	erziehung, lebenslauf, unterricht, schÃ¼ler, bildung, schulen, ungewissheit, kinder, lehrer, ...

Classifying Social Systems

Selecting Seed Words

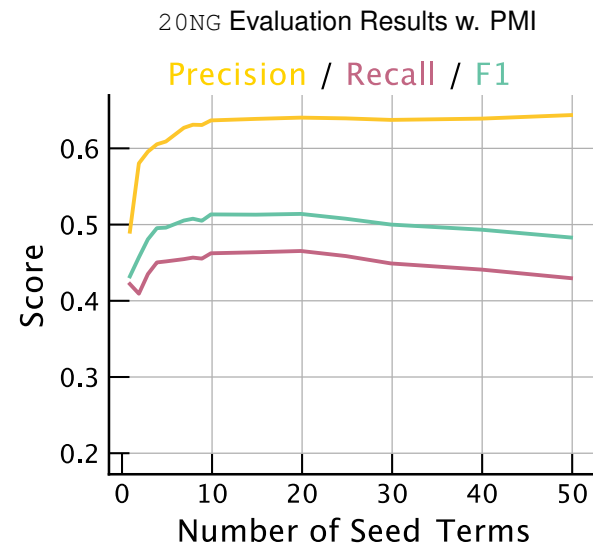
- Source: Experts → **Luhmann's books**
Pointwise mutual information ranks highly descriptive words per social system
- Amount: Few seed words with high document coverage suffice [Liu et al., 2016]

System	Seed words (translated)
Science	consciousness, communication, observations, knowledge, insight, moment, operation, ...
Legal	law, legal theory, authority, norms, rules, writing, injustice, development, lawyers ...
Economy	money, economy, prices, scarcity, work, payment, taxation, companies, market, ...
Art	art, artwork, perception, world, order, second, shape, forms, ...
Politics	power, politics, causes, effects, violence, sanctions, decision, votes, ...
Religion	religion, god, transcendence, immanence, religious, consideration, spirit, ...
Education	upbringing, resume, teaching, pupil, education, schools, uncertainty, children, teacher, ...

Classifying Social Systems

Experiment 1: Seed Term Selection Heuristic and Amount

- Model evaluation on existing datasets for parameter tuning and comparison against SOTA
- Experimental Setup:
 - Datasets: 20NG (news text), Ohsumed (med. abstracts)
 - Seed Methods: PMI, Saliency, TFIDF
 - Human Seed Baselines: [X. Li et al. 2018], [Zha & C. Li 2019]
- PMI is best out of multiple computational methods
- PMI on par with human selection
- Few seed terms suffice ($10 < n < 20$)



Classifying Social Systems

Experiment 2: Validation and Comparison to Supervised Baselines

20NG Dataset

	Name	Representation	Acc.	Prec.	Rec.	F ₁
Seed-G.	SMTM	Bag-of-Words	0.498	0.672	0.494	0.545
	Seed Term Overlap	Bag-of-Words	0.370	0.533	0.365	0.387
	Random Assignment	Bag-of-Words	0.051	0.050	0.050	0.050
Supervised	Multinomial Bayes	Bag-of-Words	0.687	0.683	0.674	0.671
	SVC	TFIDF	0.663	0.672	0.652	0.655
	Decision Tree	TFIDF	0.402	0.400	0.393	0.394
	Guidotti et al. 2021	Sparse Tensors	0.864	0.863	0.856	0.856
	Gupta et al. 2020	Multi-Sense Emb.	0.862	0.862	0.862	0.862
	Yamasada et al. 2019	Bag-of-Entities	—	—	—	0.862

→ Competitive with **basic** supervised baselines; needs no training data

Classifying Social Systems

Experiment 3: Classifying Passages from Luhmann's Books

- Setup
 - Luhmann's books split into 2653 passages, 90 terms each
 - 7 social systems, 15 seed terms each
 - SMTM model to classify passages to social systems
- Results
 - Good scores across all systems
 - High precision at acceptable recall

System	F₁
Science	0.565
Legal	0.648
Economy	0.655
Art	0.691
Politics	0.554
Religion	0.354
Education	0.656
Average	0.589

→ **Conclusion:** Social system classification with SMTM is feasible

→ **Open:** How about real-world text data?

- Approach: Ongoing transfer experiments to Wikipedia, DMOZ, scientific articles, ...