

2022 OECD Hackathon

To what extent are countries' green transition goals, as set out in their strategies, reflected in their STI policies?

Vincent Gu, Yifan Liu, Daniel J. Marchetto, Sergio Pelaez,
Matteo Zullo

Georgia Institute of Technology

June 7, 2022

Outline

- 1 Intro
- 2 Neural Network Classification
- 3 Topic Model
- 4 Clustering
- 5 Cross Dataset Mapping
- 6 Conclusion

Intro

IPCC Sixth Assessment Report on Climate Change

- Devastating consequences if 1.5°C warming by 2040

Paris Climate Goals: countries pledge ambitious climate action

- How are countries translating ambitions into policies?



Neural Network Classifier

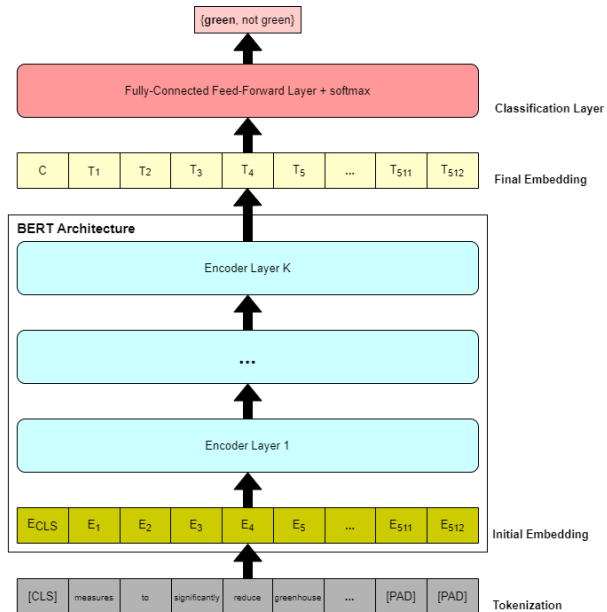
- Classification of strategy documents
 - ① Created ground truth dataset
 - ② Split 313 documents into 380,276 sentences
 - ③ Classified sentences using neural network
- Ground truth dataset
 - Manually coded 800 sample sentences: 1 if environment related, 0 if not related

id	text	label
1	Within the framework, the design of a Climate Change Observatory is being promoted.	1
2	Progress is being made towards an open access approach to information and data.	0
3	MinCienCia assumes leadership in promoting scientific and technological-based innovation.	0
4	One of the most profound global issues is the sustainability of human life on the planet.	1

Neural Network Classifier

Bidirectional Encoder Representations Transformer (BERT)

- 1 Handle non-linear classifications
 - E.g., “transition”, “sustainable”
- 2 Pre-trained bidirectionally
 - Better understanding of natural language
- 3 Outperform non-deep learning ML classifiers
 - Marchetto et al. (2021), Zullo et al. (2022)



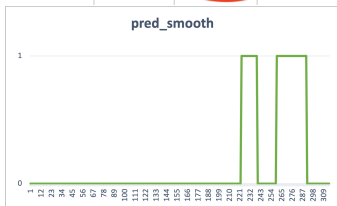
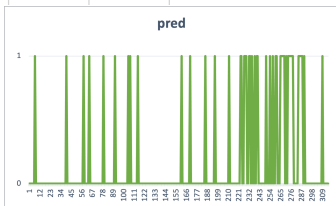
Measures to significantly reduce greenhouse gas emissions with regard to Austria's transition...

Figure: BERT neural network architecture

Neural Network Classifier

- Classification output
 - Train model (0.47 ss), predict 380,276 texts (21 mm)
 - Output predictions, smooth them using moving average

doc_id	country	text	pred	probability	pred_smooth
ES_B2P2	Spain	Social networks are information channels th	1	0.523	1
ES_B2P2	Spain	Embedded : Sensors that allow to optimise	0	0.594	1
ES_B2P2	Spain	Process improvement using embedded sens	1	0.610	1
ES_B2P2	Spain	In addition , it provides reliability and precis	0	0.641	1
ES_B2P2	Spain	The incorporation of these sensors also give	1	0.549	1



Topic Model

- Given a corpus, what latent space can be constructed to explain generative process of text?
- Classical techniques treat text as generated by random Dirichlet processes approximating latent word-topic spaces
- What is a latent topic? Latent semantic spaces?

Topic Model

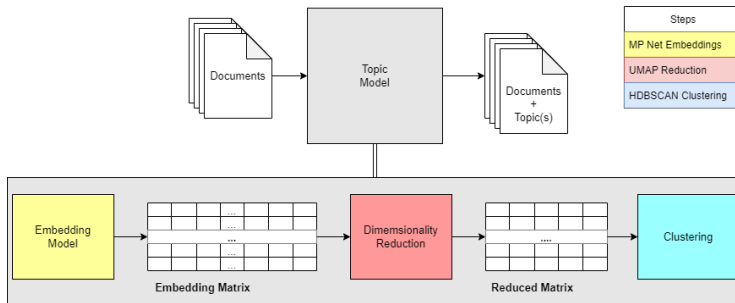


Figure: Model procedure based on BERTopic

Topic Model

- Data Preparation
 - Removed punctuation & words with < 2 characters
 - Increased classification threshold to 0.55
 - Excluded sentences with < 5 words (final $n = 13,199$)
- Algorithms' Parameters
 - UMAP - 10 neighbors & components
 - HDBSCAN - 20 minimum cluster size; ℓ_1 distance
 - TF-IDF - unigrams & bigrams
- Results
 - 70 topic clusters, 46 clearly relating to green transition

Topic Model

Topic	Count	Name	Green
-1	6195	-1_the_and_of_energy	0
0	706	0_hydrogen_fuel_of hydrogen_for	1
1	622	1_vehicles_transport_electric_road	1
2	457	2_name name_name_emergency_credit	0
3	266	3_climate_climate change_change_extreme	1
4	249	4_waste_recycling_materials_recycled	1
5	247	5_we_there_you_this	0
6	221	6_water_groundwater_irrigation_of water	1
7	208	7_batteries_lithium_battery_storage	1
8	190	8_heat_heating_pumps_heat pumps	1
9	171	9_emissions_greenhouse gas_greenhouse_gas emissions	1
10	168	10_materials_nickel_graphite_minerals	0
11	158	11_quantum_silicon_semiconductor_wafers	0
12	133	12_efficiency_energy efficiency_energy_consumption	1
13	122	13_gas_natural gas_natural_calorific	1
14	120	14_innovation_science_technology_science and	0
15	117	15_power_grid_the_transmission	1
16	115	16_forests_forest_carbon_trees	1
17	115	17_oil_prices_crude oil_crude	1
18	114	18_soil_erosion_the soil_soils	1
19	111	19_nuclear_nuclear power_fuel_spent	1
20	110	20_renewable_renewable energy_of renewable_energy	1
21	101	21_biomass_wood_biomass is_of biomass	1
22	89	22_buildings_energy_building_energy efficiency	1
23	88	23_ecosystems protection_and restoration_and ecosystems_of biodiversity	1
24	87	24_biofuels_biogas_waste_from	1

Figure: Sample Topic Model output

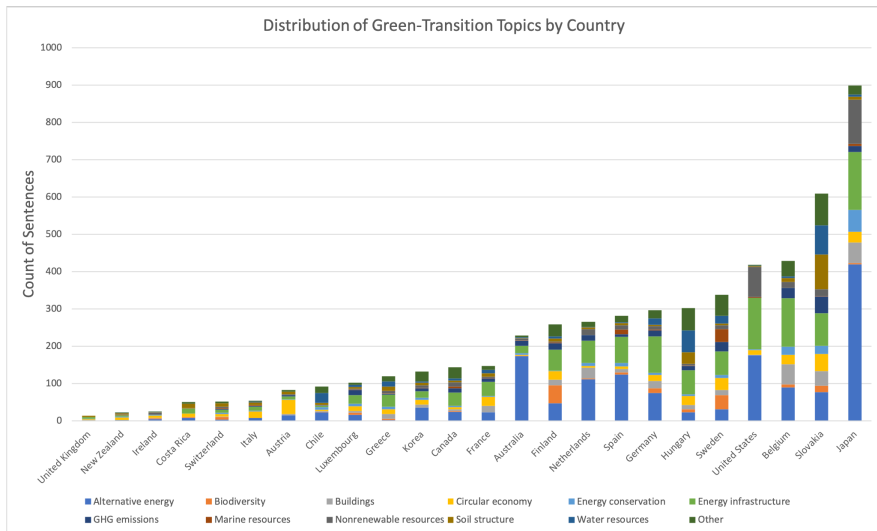


Figure: Counts of sentences by label from topic model

Clustering

Green Intensity Ratio (GIR)

$$GIR = \frac{\text{No. of "green" sentences}}{\text{No. of all sentences}}$$

- How *aggressively* countries describe transition goals

Green Broadness Ratio (GBR)

$$GBR = \frac{\text{No. of "green" topics mentioned}}{\text{No. of all possible green topics}}$$

- How *broad* transition goals are represented in the text

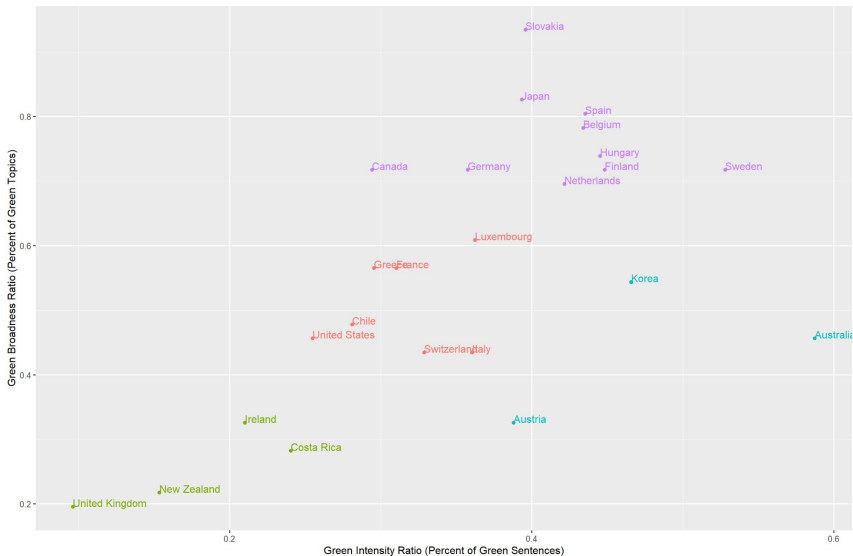


Figure: Green Intensity Ratio (GIR) vs Green Broadness Ratio (GBR): Clusters

Cross Dataset Mapping



Figure: Extent of Green Intensity

Cross Dataset Mapping

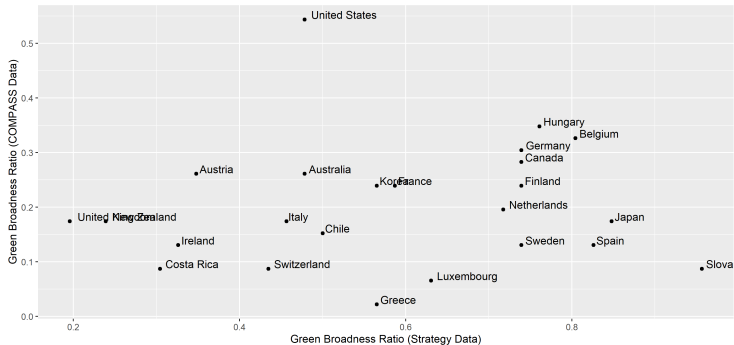


Figure: Extent of Green Broadness

Cross Dataset Mapping

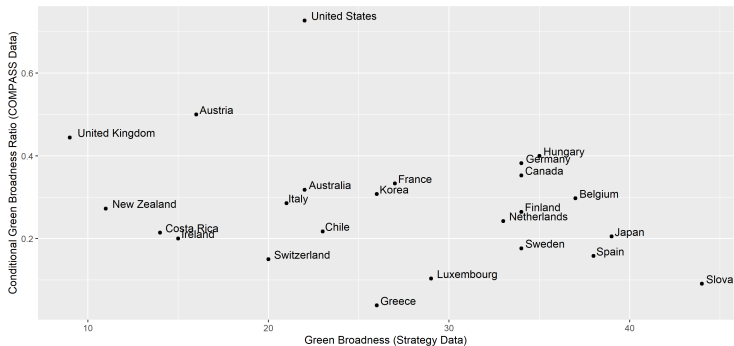


Figure: Conditional Extent of Green Broadness

Conclusions

Green Strategy

- Cross-country analysis
 - Depth and breadth of green transition positively related
 - The UK, New Zealand, and Ireland are least committed
 - Slovakia, Japan, and Spain are most committed
- Green transition macrotopics
 - Most prominent green transition topics are **alternative energy, energy infrastructure, and buildings**, with huge variation across countries

Conclusions

Green Initiative: Cross Dataset Mapping

- *Little evidence* that intensity of green transition strategies is reflected into intensity of green transition initiatives
- *We do not find* that broadness of green transition strategies is reflected into broadness of green transition initiatives after conditioning on those having in strategy

End - Thank you for your time!

Any questions or comments?

Appendix

Topic	Count	Name	Green	Topic	Count	Name	Green
-1	6195	-1_the_and_of_energy	0	35	58_35_housing_rental_homes_of_rental	0	
0	706	0_hydrogen_fuel_of_hydrogen_for	1	36	55_36_solar_solar_cells_solar_energy_panels	1	
1	622	1_vehicles_transport_electric_road	1	37	55_37_diseases_virus_infectious_the_virus	0	
2	457	2_name_name_name_emergency_credit	0	38	52_38_circular_circular_economy_economy_the_circular	0	
3	266	3_climate_climate_change_change_extreme	1	39	52_39_digital_computing_computing_power_consumption	0	
4	249	4_waste_recycling_materials_recycled	1	40	50_40_tax_energy_tax_tax_on_fuels	1	
5	247	5_we_there_you_this	0	41	47_41_agriculture_livestock_agriculture_livestock_farms	1	
6	221	6_water_groundwater_irrigation_of_water	1	42	46_42_marine_resources_marine_and_marine_and_protection	1	
7	208	7_batteries_lithium_battery_storage	1	43	44_43_landscape_protected_protected_areas_of	1	
8	190	8_heat_heating_pumps_heat_pumps	1	44	44_44_data_the_data_information_of_data	0	
9	171	9_emissions_greenhouse_gas_greenhouse_gas_emissions	1	45	43_45_jobs_skilled_workers_to_attract	0	
10	168	10_materials_nickel_graphite_minerals	0	46	42_46_manufacturing_industries_industry_services	0	
11	158	11_quantum_silicon_semiconductor_wafers	0	47	41_47_water_quality_been_identified_quality_and_identified	1	
12	133	12_efficiency_energy_efficiency_energy_consumption	1	48	41_48_carbon_carbon_dioxide_dioxide_co2	1	
13	122	13_gas_natural_gas_natural_calorific	1	49	35_49_digital_of_digital_digitization_the_digital	0	
14	120	14_innovation_science_technology_science_and	0	50	34_50_coastal_preservation_preservation_of_water_resources	1	
15	117	15_power_grid_the_transmission	1	51	33_51_smart_grids_smart_grids_storage	1	
16	115	16_forests_forest_carbon_trees	1	52	31_52_effects_damage_effects_are_their_effect	0	
17	115	17_oil_prices_crude_oil_crude	1	53	31_53_protein_proteins_products_milk	0	
18	114	18_soil_erosion_the_soil_soils	1	54	30_54_air_dust_air_quality_pollution	1	
19	111	19_nuclear_nuclear_power_fuel_spent	1	55	29_55_woman_man_man_in_standing	0	
20	110	20_renewable_renewable_energy_of_renewable_energy	1	56	29_56_tourism_of_tourism_the_tourism_climate_change	1	
21	101	21_biomass_wood_biomass_is_of_biomass	1	57	28_57_research_needs_the_research_needs_what_what_are	0	
22	89	22_buildings_energy_building_energy_efficiency	1	58	27_58_demand_energy_demand_growth_grow	1	
23	88	23_ecosystems_protection_and_restoration_and_ecosystems_of_biodiversity	1	59	27_59_fossil_fossil_fuels_fuels_fossil_energy	1	
24	87	24_biofuels_biogas_waste_from	1	60	26_60_structures_cladding_thermal_cladding_the	0	
25	81	25_species_biodiversity_climate_change_change	1	61	26_61_energy_consumption_consumption_consumption_has_decreased	1	
26	78	26_sustainable_environmental_environment_sustainable_consumption	1	62	25_62_heating_emissions_reduction_emissions_of	1	
27	71	27_climate_energy_energy_transition_transition	1	63	24_63_the_measure_significant_greenhouse_to_lead_measure	1	
28	70	28_regions_region_areas_country	0	64	24_64_productivity_labour_employment_growth	0	
29	66	29_prices_growth_inflation_slowed	0	65	24_65_degradation_have_no_risks_environmental_degradation_risks_of	1	
30	65	30_offshore_offshore_wind_wind_wind_power	1	66	23_66_and_limitation_limitation_limitation_of_pollution_to	1	
31	63	31_final_final_energy_energy_consumption_consumption	1	67	23_67_health_public_health_population_public	0	
32	62	32_broadband_spectrum_technologies_radio	0	68	23_68_coal_coal_consumption_coal_fired_power_plants	1	
33	62	33_geothermal_geothermal_energy_geothermal_power_deep	1	69	21_69_hydropower_pumped_hydroelectric_hydraulic	1	
34	58	34_digital_transitions_transitions_contribution_transitions_contribution_to	0				