

# Classification of sentences as *about-climate* and *not-about-climate* according to distance measures

[https://maelfabien.github.io/machinelearning/NLP\\_5/#implementation](https://maelfabien.github.io/machinelearning/NLP_5/#implementation)

The article above explains the method employed for classification. The difference is that instead of word vectors, we employ BERT sentence embeddings.

The embedding that represents *about-climate* class was made by averaging the embeddings of paragraphs of the texts annotated as “major”; the embedding for *not-about-climate* class was made by averaging “peripheral” texts.

The examples below display the results from splitting different articles into paragraphs, encoding these paragraph into embeddings, and classifying the paragraph as *about-climate* (green) or *not-about-climate* (red) by calculating the Euclidean distance between the paragraph’s embedding and the two embeddings that represent each class, mentioned above. The paragraph is classified according to which class embedding it’s closer to.

If the embeddings are the same, Euclidean distance is 1; the more dissimilar they are, the closer it is to 0. So in the example below, “Darwin and culture” was assigned as *not-about-climate*, since its embeddings are closer to it (0.3) than to *about-climate* (0.15).

Darwin and culture

CC:0.15428501

NOT CC:0.3056256

Here’s the code:

<https://github.com/lukaborec/Sentiment-and-argument-mining/blob/main/Sentence-embeddings.ipynb>

## Problems with sentence-wise distance calculation:

### Unclear results

- Such approaches will be crucial in facing current global challenges, from recessions to pandemics and climate change.

CC:0.7804254

NOT CC:0.6549018

What are the “such approaches”?

- Unfortunately, this is an uphill battle.

CC:0.5257464

NOT CC:0.45393652

What is an uphill battle?

## Obvious misclassifications

- There are compelling reasons to protect rainforests from deforestation and further human encroachment, something that developing and developed countries alike are increasingly realizing.

CC:0.6673375

NOT CC:0.69131446

## Problems with paragraph-wise distance calculation:

### Low level of granularity

- A fitting illustration is the management of fisheries and marine ecosystems. Recent history is replete with examples of how overfishing has led to a collapse of important fisheries. These problems have resulted from a combination of inadequate scientific information on stock abundance and recommended catches, insufficient government regulation of the conduct of fisheries and of allowable quotas, and international competition for access to resources, often acted out in open defiance of international law. The global long-term sustainable yield of commercial fishing is estimated to be some 100 million tons. But there are indications that it might be higher if overfishing is curbed to allow stocks to recover, if renegades are curtailed, and if scientifically sound practices of policing and inspection are supported. As the knowledge base widens and common interests become more clearly perceived, there is really no reason why we should not be able to optimize the long-term yield through sensible management systems.

CC:0.7925511

NOT CC:0.65176725

While this paragraph is clearly about climate change – overfishing and the devastation of fish habitats – but we need to find a way to deduce which sentences of the paragraph are relevant. One solution is to calculate word embeddings.

## Misclassifications

- I recently came across an article written by a Norwegian scientist during the 1970s, when I was Norway's Minister of the Environment. In the article he argued that there was no such problem as acid rain and that “facts” and “science” did not belong in the arena of politics and

policy. This assertion was counter to my own beliefs and made me react strongly. Politics that disregard science and knowledge will not stand the test of time. Indeed, there is no other basis for sound political decisions than the best available scientific evidence. This is especially true in the fields of resource management and environmental protection.

CC:0.72114587

NOT CC:0.7555592

This paragraph is classified as negative although it indirectly expressed the author's stance on environmental protection. However, paragraph-wise classification seems to be making far fewer mistakes than sentence-based classification.

## Low difference between the two distances

- In ocean management, as in most other areas of human endeavor, close cooperation between scientists and politicians is the only way to move forward. Science must underpin our policies. If we compromise on scientific facts and evidence, repairing nature will be enormously costly—if possible at all.

CC:0.78352535

NOT CC:0.7789918

- As caretakers of our common future, we have the responsibility to seek scientifically sound policies, nationally as well as internationally. If the long-term viability of humanity is to be ensured, we have no other choice.

CC:0.72157943

NOT CC:0.7165551

The difference between CC and NOT CC scores tends to be low, here it's only .1. Solution: more data?

## Classification of a text classified as “peripheral”

4611173b-peripheral.txt

Darwin and culture

CC:0.15428501

NOT CC:0.3056256

A new series of essays traces the astounding variety of reactions to the theory of evolution.

CC:0.33830497

NOT CC:0.5159374

The public reception of scientific ideas depends largely on two factors: people's ability to grasp factual information and the cultural lens through which that information is filtered. The former is what scientists tend to focus on when they give popular accounts of issues such as climate change. The assumption is that if they explain things very, very clearly, everyone will understand. Unfortunately, this is an uphill battle. The general public's average capacity to weigh facts and numbers is notoriously poor – although there is encouraging evidence that probabilistic reasoning can be improved by targeted education early in life (see page 1189).

CC:0.7764772

NOT CC:0.72297084

disagree

Even more crucial, however, are the effects of the cultural lens. Over the coming month, Natures Opinion pages will explore particularly vivid examples of these effects in the world's widely divergent reactions to Charles Darwin's ideas about evolution in the late nineteenth and early twentieth centuries (see page 1200).

CC:0.6058615

NOT CC:0.6377132

In England, for example, the Church reacted badly to Darwin's theory, going so far as to say that to believe it was to imperil your soul. But the notion that Darwin's ideas 'killed' God and were a threat to religion was by no means the universal response in the nineteenth century.

CC:0.59054923

NOT CC:0.5830164

Darwin's theory reached the world at a time when many people were looking for explanations for social, political and racial inequalities, and in many parts of the world were wondering how to improve their lot in the face of Europe's global imperialism. So from Egypt to India, China and Japan, many religious scholars embraced Darwin's ideas, often showing how their own schools of thought had anticipated the notion of evolution. Against the threat of Western imperialism and Western charges of 'backwardness', it was to their advantage to highlight the rationality of their creed.

CC:0.55030996

NOT CC:0.73679364

In China, Darwin's ideas were seen as supporting Confucians' belief in the perfectibility of the cosmic order. Evolutionary theory also became fodder for political movements of revolution and reform, and eventually laid the groundwork for communism. Latin American politicians initially reacted to Darwin's ideas by attempting to entice white Europeans to emigrate and intermarry with local populations, believing that this would 'improve the stock'. But after two world wars had made European culture look less impressive, Latin America began to see its racial diversity as an advantage, and moved towards a social view that favoured a homogeneous blend of cultures.

CC:0.56121254  
NOT CC:0.671714

In nineteenth-century Russia, meanwhile, a tendency to distrust rabid, dog-eat-dog capitalism helped incline **naturalists** away from a view of evolution that emphasized competition between species. Instead they embraced a 'theory of mutual aid, an account that focused on the role of cooperation in ensuring survival in a harsh environment.

CC:0.5845236  
NOT CC:0.55780554

The lesson for today's scientists and policy-makers is simple: they cannot assume that a public presented with 'the facts' will come to the same conclusion as themselves. They must take value systems, cultural backdrops and local knowledge gaps into account and frame their arguments accordingly. Such approaches will be crucial in facing current global challenges, from recessions to pandemics and climate change. These issues will be perceived and dealt with differently by different nations — not because they misunderstand, but because their understanding is in part locally dependent.

CC:0.7170604  
NOT CC:0.658403

Darwin once said: "But then with me the horrid doubt always arises whether the convictions of man's mind, which has been developed from the mind of the lower animals, are of any value or at all trustworthy." Researchers and policy-makers would do well to mimic his humility when presenting science, and remember how people's minds truly work.

CC:0.5944788  
NOT CC:0.62920904

## Classification of an unannotated text

NatureOCR/279001a0.ocr.txt

Costs and benefits of carbon dioxide  
CC:0.3406849  
NOT CC:0.42073855

The release of carbon dioxide to the atmosphere by the burning of fossil fuels is, conceivably, the most important environmental issue in the world today. Whatever direction global energy policies take in the future, it is indisputable that carbon dioxide concentrations in the atmosphere will continue to rise. There is still uncertainty about the ultimate destination of carbon dioxide. It seems that roughly half the fossil fuel output has remained in the atmosphere, and early workers supposed that the remainder was consumed by the oceans and the biosphere. But the role of the biosphere is

now a matter of hot debate. Some research has suggested that far from being a sink for carbon dioxide, the biosphere (through deforestation and changing land use) could actually be a source. Other work suggests the contrary, or that the role of the biosphere has actually changed with time. But it is inescapable that atmospheric concentrations have already climbed by 15% as a result of man's activities during this century and there seems little doubt that concentrations would be double present values around the middle of the next century if current growth rates for the use of fossil fuels (over 4% per annum) were to persist. This is unlikely, of course, given the depletion of energy resources, but at least the figure gives some sort of guide for realistic modelling.

CC:0.71440387

NOT CC:0.6532393

Whatever the uncertainties about future emissions and the biosphere, there is no disagreement amongst researchers on the qualitative impact that an increase in carbon dioxide will have on climate: mean annual surface temperature will rise, and the rises will be greater at high latitudes. There is also consensus that the hydrological cycle would become more active—with precipitation and evaporation levels both rising. Beyond this there is still scope for quantitative disagreement, but a commonly quoted figure is that a doubling of atmospheric carbon dioxide would result in a world global annual mean surface temperature rise of 2 to 3 °C, with marked latitudinal asymmetry. As yet, however, no model adequately accounts for changes in the ice-covered regions of the world or in the hydrosphere (particularly ocean currents), and there is considerable room for disagreement regarding the importance of feedback effects arising from changes in cloud cover.

CC:0.7022232

NOT CC:0.54372305

With so much uncertainty around, is it irresponsible and premature to widen the debate at this stage from meteorologists and climatologists to those with interests in the consequences of climatic change—agriculturalists, glaciologists, oceanographers, economists, sociologists, political scientists and so on? Surely not, provided that sensible perspectives are maintained. A recent workshop jointly sponsored by the American Association for the Advancement of Science and the US Department of Energy has been attempting to lay transdisciplinary foundations for a federally supported research programme on the impact of increasing atmospheric carbon dioxide content and it is not too early for other nations (or more reasonably groups of nations, such as the European Economic Community) to take similar initiatives. Even if large amounts of money were not immediately forthcoming, there are still some links across the specialist boundaries which ought to be made now.

CC:0.74489176

NOT CC:0.75211614

mistake?

In the long run the United Nations presumably has to get in on the act, and the United Nations Environmental Programme will shortly be setting up a carbon dioxide committee. At first sight scientists might despair at the thought of yet another area in which there will be politicised conflict between industrialised nations, large-scale releasers of carbon dioxide, and the developing world, involuntary recipients of the consequences. But careful reading of what climatologists and

meteor-logists have to say by way of prediction makes it clear that there could be as many benefits as losses as a result of temperature and rainfall changes—and that some parts of the world may even become cooler.

CC:0.7480972

NOT CC:0.7604522

mistake

There is no clear indication that the animal and plant kingdoms will as a whole prosper more or less in a changed climate. And there may be direct carbon dioxide effects, such as changes in the rates of photosynthesis and respiration, increases in the efficiency of plant water use and changes in nitrogen fixation rates. To be sure, the most widely publicised effect of a substantial global warming is the danger of the West Antarctic Ice Sheet breaking loose and melting, with highly predictable effects on sea level. But for the rest the picture is complex and by no means universally gloomy. The sooner some of the complexities are unravelled, the sooner the carbon dioxide problem can be intelligently injected into discussions of world energy strategies.

CC:0.8233782

NOT CC:0.62991905