Manuscripts and newspapers: Using machine learning to map the sciences and identify political bias

Alan Medlar and Dorota Glowacka
SCIENTIFIC LITERATURE
Motivation

- We perform exploratory search user studies on scientific literature
- Full-text retrieval has higher recall (and lower precision) over searching bibliographic records (title, abstract, etc.)
- Could using abstracts for retrieval impact experimental results?
Abstracts vs. full-text

- **Bioinformatics**: article sections (e.g., results) besides abstract provides better representations of certain biological concepts

- **Medicine**: clinical decisions based solely on abstracts results in worse patient outcomes

- **General**: well-known differences between established/emerging fields, theoretical/ applied fields, individual fields tend to have own style/expectations
Research questions

• **RQ1:** How well do abstracts represent the full-text of a paper in different CS subfields?

• **RQ2:** If there are differences between subfields, could this impact (perceived) retrieval performance?
Data preprocessing

- **35,137 CS papers** from arXiv (2007-mid 2015)

- 23% papers associated with > 1 category (40 possible author-assigned categories)

- Extracted 6.7 sections per article (SD = 2.7)

- Classify sections as abstract, introduction, background, related work, methods, results, discussion, conclusions and back matter

- Classified **53%** of sections based on headings **that occurred at least twice** (35% of headings were unique)

- Classified 3.3 sections per article (SD = 1.2)
Representation

- **Full-text** and **sections** represented using probabilistic topic models

- Topic model inferred from full-text + used to predict individual sections (100 topics)

- Multi-sections merged using element-wise summarisation and normalised

- Representativeness metric

- **KL divergence**: two discrete probability distributions, $P$ and $Q$, "how much information is lost when $Q$ is used to approximate $P"$

\[ D_{KL} (P||Q) = \sum_i P(i) \log_2 \left( \frac{P(i)}{Q(i)} \right) \]
Abstract representativeness is subfield-specific

- The degree to which abstracts represent the full-text is subfield-specific

- Higher mean KL divergence (less representative abstract) appear to be more theoretical...

- Lower mean KL divergence (more representative abstract) appear to be more applied...

- Theoretical abstracts tend to be shorter, but KL divergence is not correlated with abstract length ($R^2 = 0.003, p < 2.2 \times 10^{-16}$)
Abstract representativeness is subfield-specific

- The degree to which abstracts represent the full-text is **subfield-specific**

- **Higher mean KL divergence** (less representative abstract) *appear* to be more theoretical...

- Lower mean KL divergence (more representative abstract) *appear* to be more applied...

- Theoretical abstracts tend to be shorter, but KL divergence is not correlated with abstract length ($R^2 = 0.003, p < 2.2 \times 10^{-16}$)
Abstract representativeness is correlated with retrieval performance

• What is the impact on retrieval?
  • Generate queries that **disproportionately favour specific subfields** (most informative features from multi-class SVM, removed duplicates, manually removed junk)
  • 1,257 queries (33.1 per category, SD = 14.5)
  • Retrieve top-100 results using full-text and abstracts - calculate precision@100
  • Precision@100 negatively correlated with KL divergence ($R^2 = 0.38$, $p = 3.65 \times 10^{-5}$)
Section-wise representativeness captures domain structure

• Are these trends random, or related to some underlying structure?

• Calculate KL divergence between all 8 sections and full-text

• Do hierarchical clustering (complete-linkage clustering with Euclidean distance)

• **Theoretical** and **applied** subtrees, deeper subtrees make sense

• Obvious errors explained by high variance (e.g. Operating Systems) or corpus bias (e.g. Networking and Internet, 23% associated with Information theory as well)
Section-wise representativeness captures domain structure

- Are these trends random, or related to some underlying structure?

- Calculate KL divergence between all 8 sections and full-text

- Do hierarchical clustering (complete-linkage clustering with Euclidean distance)

- **Theoretical** and **applied** subtrees, deeper subtrees make sense

- Obvious errors explained by high variance (e.g. Operating Systems) or corpus bias (e.g. Networking and Internet, 23% associated with Information theory as well)
Section-wise representativeness captures domain structure

• Are these trends random, or related to some underlying structure?

• Calculate KL divergence between all 8 sections and full-text

• Do hierarchical clustering (complete-linkage clustering with Euclidean distance)

• **Theoretical** and **applied** subtrees, deeper subtrees make sense

• Obvious errors explained by high variance (e.g. Operating Systems) or corpus bias (e.g. Networking and Internet, 23% associated with Information theory as well)
Section-wise representativeness captures domain structure

- Are these trends random, or related to some underlying structure?

- Calculate KL divergence between all 8 sections and full-text

- Do hierarchical clustering (complete-linkage clustering with Euclidean distance)

- **Theoretical** and **applied** subtrees, deeper subtrees make sense

- Obvious errors explained by high variance (e.g. Operating Systems) or corpus bias (e.g. Networking and Internet, 23% associated with Information theory as well)
Section-wise representativeness captures domain structure

- Are these trends random, or related to some underlying structure?

- Calculate KL divergence between all 8 sections and full-text

- Do hierarchical clustering (complete-linkage clustering with Euclidean distance)

- **Theoretical** and **applied** subtrees, deeper subtrees make sense

- Obvious errors explained by high variance (e.g. Operating Systems) or corpus bias (e.g. Networking and Internet, 23% associated with Information theory as well)
Current work

• Unexpectedly interesting results...

• In progress:
  • **Downloaded and preprocessed all of arXiv (~1.5M → ~1M)**
  • **CS, Physics, CS + Physics (monophyletic?)**
  • Better use of topic models
  • Improved tree building (phylogenetics)
POLITICAL BIAS
Motivation

• Facebook et al. recommend news articles, but **obscure** the publisher...

• Clicking on articles is telling Facebook "*more news about Brexit please!*"

• Can we automatically identify **political bias** in news articles?
Left & right wing framing of Article 50 high court verdict
James Slack CBE (currently Prime Minister’s official spokesperson)

Left & right wing framing of Article 50 high court verdict

The Guardian

Turmoil for May as judges rule that parliament must decide on Brexit

Daily Mail

Fury over 'out of touch' judges who defied 17.4m Brexit voters and could trigger constitutional crisis
Theresa May is heading for a rebellion over her Brexit strategy, after the high court ruled that Britain could not leave the European Union without the permission of parliament.

Three senior judges declared yesterday that the government could not press ahead with triggering article 50 of the Lisbon treaty, the formal process for beginning Brexit, without first consulting MPs and peers in the Commons and Lords.

The decision, made after a legal challenge brought in the wake of the EU referendum in June, is a dramatic setback for the prime minister, who had argued she had the personal authority to begin the process without a parliamentary vote.

Downing Street has said it will challenge
Left & right wing framing of Article 50 high court verdict

Theresa May is heading for a rebellion over her Brexit strategy, after the high court ruled that Britain could not leave the European Union without the permission of parliament.

Three senior judges declared yesterday that the government could not press ahead with triggering article 50 of the Lisbon treaty, the formal process for beginning Brexit, without first consulting MPs and peers in the Commons and Lords.

The decision, made after a legal challenge brought in the wake of the EU referendum in June, is a dramatic setback for the prime minister, who had argued she had the personal authority to begin the process without a parliamentary vote.

Downing Street has said it will challenge the court’s ruling a chance for MPs to put the national interest first and halt Brexit before it wreaks any more havoc, says Polly Toynbee

A momentous constitutional decision was taken by the high court of England and Wales yesterday. A prime minister’s absolute power to do what they like, when they like, regardless of laws and treaties, was struck down. Theresa May cannot tear up our right to be EU citizens without the authority of parliament.

Judges, wisely, do not generally want to usurp the power of elected governments to govern. Laws made by judges are a poor substitute for those made by the people.
Left & right wing framing of Article 50 high court verdict
Left & right wing framing of Article 50 high court verdict

Theresa May is heading for a rebellion over her Brexit strategy, after the high court ruled that Britain could not leave the European Union without the permission of parliament.

Three senior judges declared yesterday that the government could not press ahead with triggering article 50 of the Lisbon treaty, the formal process for beginning Brexit, without first consulting MPs and peers in the Commons and Lords.

The decision, made after a legal challenge brought in the wake of the EU referendum in June, is a dramatic setback for the prime minister, who had argued she had the personal authority to begin the process without a parliamentary vote.

Downing Street has said it will challenge the ruling, and Brussels legal experts have said the government could go ahead with triggering article 50.

The court’s ruling is a chance for MPs to put the national interest first and halt Brexit before it wreaks any more havoc, says Polly Toynbee.

A momentous constitutional decision was taken by the high court of England and Wales yesterday. A prime minister’s absolute power to do what they like, when they like, regardless of laws and treaties, was struck down. Theresa May cannot tear up our right to be EU citizens without the authority of parliament.

Judges, wisely, do not generally want to usurp the power of elected governments to govern. Laws made by judges are a poor substitute for those made by the representatives of the people.

MPS last night tore into an unelected panel of ‘out of touch’ judges for ruling that embittered Remain supporters in Parliament should be allowed to frustrate the overwhelming verdict of the British public.

The Lord Chief Justice and two senior colleagues were accused of putting Britain on course for a full-blown ‘constitutional crisis’ by saying Brexit could not be triggered without a Westminster vote.

The judgment by Lord Thomas – a founding member of the European Law Institute, a club of lawyers and academics aiming to ‘improve’ EU law – throws into chaos Mrs May’s timetable for invoking article 50 in March next year.

Senior MPs – led by an ex-justice minister – said it was an outrage that an ‘unholy alliance’ of judges and embittered Remain backers could thwart the wishes of 17.4 million Leave voters.

They warned that Mrs May could be forced to hold an election early next year if the courts did not back down. Leave
Left & right wing framing of Article 50 high court verdict

Theresa May is heading for a rebellion over her Brexit strategy, after the high court ruled that Britain could not leave the European Union without the permission of parliament.

Three senior judges declared yesterday that the government could not press ahead with triggering article 50 of the Lisbon treaty, the formal process for beginning Brexit, without first consulting MPs and peers in the Commons and Lords.

The decision, made after a legal challenge brought in the wake of the EU referendum in June, is a dramatic setback for the prime minister, who had argued she had the personal authority to begin the process without a parliamentary vote.

Downing Street has said it will challenge.

The court’s ruling is a chance for MPs to put the national interest first and halt Brexit before it wreaks any more havoc, says Polly Toynbee.

A momentous constitutional decision was taken by the high court of England and Wales yesterday. A prime minister’s absolute power to do what they like, when they like, regardless of laws and treaties, was struck down. Theresa May cannot tear up our right to be EU citizens without the authority of parliament.

Judges, wisely, do not generally want to usurp the power of elected governments to govern. Laws made by judges are a poor substitute for those made by Parliament.

MPS last night tore into an unelected panel of ‘out of touch’ judges for ruling that embittered Remain supporters in Parliament should be allowed to frustrate the overwhelming verdict of the British public.

The Lord Chief Justice and two senior colleagues were accused of putting Britain on course for a full-blown ‘constitutional crisis’ by saying Brexit could not be triggered without a Westminster vote.

The judgment by Lord Thomas – a founding member of the European Law Institute, a club of lawyers and academics aiming to ‘improve’ EU law – throws into chaos Mrs May’s timetable for invoking article 50 in March next year.

Senior MPS – led by an ex-justice minister – said it was an outrage that an ‘unholy alliance’ of judges and embittered Remain backers could thwart the wishes of 17.4 million Leave voters. They warned that Mrs May could be forced to hold an election early next year if the courts did not back down.

Turn to Page 2
Left & right wing framing of Article 50 high court verdict

Theresa May is heading for a rebellion over her Brexit strategy, after the high court ruled that Britain could not leave the European Union without the permission of parliament.

Three senior judges declared yesterday that the government could not press ahead with triggering article 50 of the Lisbon treaty, the formal process for beginning Brexit, without first consulting MPs and peers in the Commons and Lords.

The decision, made after a legal challenge brought in the wake of the EU referendum in June, is a dramatic setback for the prime minister, who had argued she had the personal authority to begin the process without a parliamentary vote.

Downing Street has said it will challenge the court's ruling - which is a chance for MPs to put the national interest first and halt Brexit before it wreaks any more havoc, says Polly Toynbee.

A momentous constitutional decision was taken by the high court of England and Wales yesterday. A prime minister's absolute power to do what they like, when they like, regardless of laws and treaties, was struck down. Theresa May cannot tear up our right to be EU citizens without the authority of parliament.

Judges, wisely, do not generally want to usurp the power of elected governments to govern. Laws made by judges are a poor substitute for those made by elected representatives.

The Lord Chief Justice and two senior colleagues were accused of putting Britain on course for a full-blown 'constitutional crisis' by saying Brexit could not be triggered without a Westminster vote.

The judgment by Lord Thomas – a founding member of the European Law Institute, a club of lawyers and academics aiming to 'improve' EU law – throws into chaos Mrs May's timetable for invoking article 50 in March next year.

Senior MPs – led by an ex-justice minister – said it was an outrage that an 'architect of judges' and embittered Remain backers could thwart the wishes of 17.4 million Leave voters. They warned that Mrs May could be forced to hold an election early next year if the courts did not back down. Leave voters could then be asked to choose between their prime minister and the judges.
Theresa May is heading for a rebellion over her Brexit strategy, after the high court ruled that Britain could not leave the European Union without the permission of parliament.

Three senior judges declared yesterday that the government could not press ahead with triggering article 50 of the Lisbon treaty, the formal process for beginning Brexit, without first consulting MPs and peers in the Commons and Lords.

The decision, made after a legal challenge brought in the wake of the EU referendum in June, is a dramatic setback for the prime minister, who had argued she had the personal authority to begin the process without a parliamentary vote.

Downing Street has said it will challenge

The court's ruling is a chance for MPs to put the national interest first and halt Brexit before it wreaks any more havoc, says Polly Toynbee

A momentous constitutional decision was taken by the high court of England and Wales yesterday. A prime minister's absolute power to do what they like, when they like, regardless of laws and treaties, was struck down. Theresa May cannot tear up our right to be EU citizens without the authority of parliament.

Judges, wisely, do not generally want to usurp the power of elected governments to govern. Laws made by judges are a poor substitute for those made by elected representatives.

MPS last night tore into an unelected panel of 'out of touch' judges for ruling that embittered Remain supporters in Parliament should be allowed to frustrate the overwhelming verdict of the British public.

The Lord Chief Justice and two senior colleagues were accused of putting Britain on course for a full-blown 'constitutional crisis' by saying Brexit could not be triggered without a Westminster vote.

The judgment by Lord Thomas - a founding member of the European Law Institute, a club of lawyers and academics aiming to 'improve' EU law - throws into chaos Mrs May's timetable for invoking article 50 in March next year.

Senior MPs – led by an ex-justice minister – said it was an outrage that an ' unholy alliance' of judges and embittered Remain backers could thwart the wishes of 17.4 million Leave voters. They warned that Mrs May could be forced to hold an election early next year if the courts did not back down. Leave
Left & right wing framing of Article 50 high court verdict

Theresa May is heading for a rebellion over her Brexit strategy, after the high court ruled that Britain could not leave the European Union without the permission of parliament.

Three senior judges declared yesterday that the government could not press ahead with triggering article 50 of the Lisbon treaty, the formal process for beginning Brexit, without first consulting MPs and peers in the Commons and Lords.

The decision, made after a legal challenge brought in the wake of the EU referendum in June, is a dramatic setback for the prime minister, who had argued she had the personal authority to begin the process without a parliamentary vote.

Downing Street has said it will challenge the court’s ruling.

The court’s ruling is a chance for MPs to put the national interest first and haul Brexit before it wreaks any more havoc, says Polly Toynbee.

A momentous constitutional decision was taken by the high court of England and Wales yesterday. A prime minister’s absolute power to do what they like, when they like, regardless of laws and treaties, was struck down. Theresa May cannot tear up our right to be EU citizens without the authority of parliament.

Judges, wisely, do not generally want to usurp the power of elected governments to govern. Laws made by judges are a poor substitute for those made by those elected.

MPS last night tore into an unelected panel of ‘out of touch’ judges for ruling that embittered Remain supporters in Parliament should be allowed to frustrate the overwhelming verdict of the British public.

The Lord Chief Justice and two senior colleagues were accused of putting Britain on course for a full-blown ‘constitutional crisis’ by saying Brexit could not be triggered without a Westminster vote.

The judgment by Lord Thomas – a founding member of the European Law Institute, a club of lawyers and academics aiming to ‘improve’ EU law – throws into chaos Mrs May’s timetable for invoking article 50 in March next year.

Senior MPs – led by an ex-justice minister – said it was an outrage that an unholy alliance of judges and embittered Remain backers could thwart the wishes of 17 million Leave voters. They warned that Mrs May could be forced to hold an election early next year if the courts did not back down. Leave
Left & right wing framing of Article 50 high court verdict

The court's ruling is a chance for MPs to put the national interest first and halt Brexit before it wreaks any more havoc, says Polly Toynbee.

A momentous constitutional decision was taken by the high court of England and Wales yesterday. A prime minister's absolute power to do what they like, when they like, regardless of laws and treaties, was struck down. Theresa May cannot tear up our right to be EU citizens without the authority of parliament. Judges, wisely, do not generally want to usurp the power of elected governments to govern. Laws made by judges are a poor substitute for those made by elected representatives, as Mr Justice Grieve pointed out in his dissenting judgment. The court's decision is a clear statement that the national interest cannot be surrendered to the desire of one government.

MPS last night tore into an unelected panel of 'out of touch' judges for ruling that embittered Remain supporters in Parliament should be allowed to frustrate the overwhelming verdict of the British public. The Lord Chief Justice and two senior colleagues were accused of putting Britain on course for a full-blown 'constitutional crisis' by saying Brexit could not be triggered without a Westminster vote. The judgment by Lord Thomas - a founding member of the European Law Institute, a club of lawyers and academics aiming to 'improve' EU law - throws into chaos Mrs May's timetable for invoking article 50 in March next year. Senior MPs, led by an ex-justice minister - said it was an outrage that an 'unholy alliance' of judges and embittered Remain backers could thwart the wishes of 17.4 million Leave voters. They warned that Mrs May could be forced to hold an election early next year if the courts did not back down. Leave
Approach

• Data sets:
  • Left-wing British newspaper articles (Guardian)
  • Right-wing British newspaper articles (Daily Mail)

• Build two language models

• Calculate the likelihood ratio on a per word basis:
  • $1.0 = \text{neutral}$
  • $< 1.0 = \text{left-wing}$
  • $> 1.0 = \text{right-wing}$
Preliminary results (with broken data)

- **Left-wing** articles about UK politics
- **Right-wing** articles about UK or politics
- Model issues (see house of commons)

What is the original 'backstop' in the Withdrawal Agreement? Variously described as an insurance policy or safety net, the backstop is a device in the Withdrawal Agreement intended to ensure that there will not be a hard border between Northern Ireland and the Republic of Ireland, even if no formal deal can be reached on trade and security arrangements. It would mean that if there were no workable agreement on such matters, Northern Ireland would stay in the customs union and much of the single market, guaranteeing a friction-free border with the Republic. This would keep the Good Friday agreement intact. Both the UK and EU signed up to the basic idea in December 2017 as part of the initial Brexit deal, but there have been disagreements since on how Brussels and allies across Europe. Johnson suggested the EU's position was influenced by the manoeuvres of Conservative MPs who have been examining legislative methods to stop no deal in the [House of Commons](https://www.parliament.uk), including former cabinet ministers like Philip Hammond. Downing
Acknowledgements

Chien-yu Huang
Arlene Casey

Yu An
Yang Liu