Review of “Automatically Classifying the Role of Citations in Biomedical Articles”

Journal Link:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041379/

Citation Schema Modification

A.
- A1. Relation between the citing and the cited articles
- A2. Relation between two cited articles
- A3 Relation between two propositions in one cited article

B.
- B1. Related to knowledge
- B2. Related to experimental protocol

C.
- C1 No author's comments
- C2 Author's comments:
  - C2a reported speech and discovery
  - C2b epistemic modality
  - C2c evaluation
    - C2c1 positive evaluation
    - C2c2 quantitative evaluation
    - C2c3 first to discover
  - C2d adversative
    - C2d1 contrast
    - C2d2 conflict
    - C2d3 specify
    - C2d4 obstacle
  - C2e consistency
    - C2e1 similar features
    - C2e2 similar opinions, results and methods
  - C2f cause

1. **Background/Perfunctory**
2. **Contemporary**
3. **Contrast/Conflict**
4. **Evaluation**
5. **Explanation of results**
6. **Material/Method**
7. **Modality**
8. **Similarity/Consistency**

Modify because:
Confusion about the hierarchy scheme
Poor inter-annotator agreement

Citation Scheme:
Hierarchical Schema
Minimized overlapping features

Two annotators label citations based on cue words

Table 1:

Inter-annotator agreement over the 10 articles that were annotated by the two annotators

Category	Kappa Value	F1-Score (%)
Background/Perfunctory	0.49	73.20
Contemporary	0.87	86.96
Contrast/Conflict	0.67	72.22
Evaluation	0.53	59.13
Explanation	0.49	53.93
Method	0.89	90.65
Modality	0.60	65.30
Similarity/Consistency	0.48	52.38
Average	0.63	69.22

Open in

Kappa Value

Cohen's Kappa	Interpretation
0	No agreement
0.10 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 0.99	Near perfect agreement
1	Perfect agreement

Accuracy from Model

Table 2:

Performance of the SVM and MNB algorithms for each category. The models were evaluated using a 10-fold cross validation paradigm. All values in %. SVM: Support Vector Machine, MNB: Multinomial Naïve Bayes

Category	**Number Feature**	**Best model**	**Recall**	**Precision**	**F1-Score**	**Accuracy**
Background/ Perfunctory	150	SVM	66.4±5.4	80.5±7.1	72.7±5.7	78.8±4.4
Contemporary	25	SVM	86.4±16.1	95.9±6.6	89.7±8.3	99.0±0.9
Contrast/Conflict	50	SVM	79.1±8.5	88.5±6.6	83.4±7.1	92.9±3.1
Evaluation	200	SVM	71.3±9.8	46.7±8.1	55.8±6.6	89.5±1.5
Explanation	100	SVM	82.6±11.9	58.6±11.8	68.0±9.9	95.3±1.4
Method	250	MNB	89.5±6.4	73.3±10.4	80.0±5.7	94.0±1.4
Modality	100	SVM	86.0±8.1	78.7±5.1	81.9±4.8	93.8±1.8
Similarity/Consistency	50	SVM	79.2±9.9	61.8±6.7	68.8±4.6	94.6±1.0
Average	-	-	80.1	73.0	75.0	92.2

Open in a separate window

Confusion

- We used unigrams (individual words) and bigrams (two consecutive words) as features to train the algorithms.
- We trained and tested the algorithms with the top 25, 50, 100, 150, 200, 250, and 500 features. Features were ranked using mutual information.

Role of linguistic features for classification

Researcher	Features used	Model Used	Accuracy
Agarwal et al.	• Uni-grams • Bi-grams	SVM and MNB	92.2%
Sugiyama et al.	• Proper nouns • Previous and next sentence • Uni-gram • Bi-gram	SVM and MaxEnt	88.2%
Wang et al.	• 48 groups of Cue phrase	• High number of cue phrase identifies better than low number of cue phrase	
Small	• Hedging words (“using”)	• Word “using” has higher predictability for method and non-method section	

- SVM model outperform other classifier in determining citations accurately

Identifying citing sentences in research papers using supervised learning

Sugiyama et al.

ME: Maximum Entropy

Table 1. Accuracy obtained byME and SVM.

Feature	Accuracy (ME)	Accura cy (SVM)
(1) Unigram	0.876	0.879
(2) Bigram	0.827	0.851
(3) Proper Noun	**0.882**	**0.882**
(4) Previous and Next Sentence	**0.882**	**0.882**
(5) Position	0.875	0.877
(6) Orthographic	0.878	0.880
(7) All [(1) - (6)]	0.876	0.878

https://www.researchgate.net/publication/224138996_Identifying_citing_sentences_in_research_papers_using_supervised_learning

Knowledge Flows, Patent Citations and the Impact of Science on Technology

- 29 pages
- Cited by 90

ARTICLE

Knowledge Flows, Patent Citations and the Impact of Science on Technology

Nomaler, önder ; Verspagen, Bart

Economic systems research, 2008, Vol.20 (4), p.339-366

PEER REVIEWED OPEN ACCESS

Online Access Available ›

Abstract

Technological innovation depends on knowledge developed by scientific research. The number of citations made in patents to the scientific literature has been suggested as an indicator of this process of transfer of knowledge from science to technology. We provide an intersectoral insight into this indicator, by breaking down patent citations into a sector-to-sector matrix of knowledge flows. We then propose a method to analyze this matrix and construct various indicators of science intensity of sectors, and the pervasiveness of knowledge flows. Our results indicate that the traditional measure of the number of citations to science literature per patent captures important aspects of intersectoral knowledge flows, but that other aspects are not captured. In particular, we show that high science intensity implies that sectors are net suppliers of knowledge in the economic sector, but that science intensity does not say much about pervasiveness of either knowledge use or knowledge supply by sectors. We argue that these results are related to the specific and specialized nature of knowledge.

Overview

Defined ***matrix*** and ***indicators*** to represent and quantify the knowledge flow(from science to technology)

Defined the knowledge flow table/matrix to measure the flow of science(non-patent paper) to industry(patents)

Defined indicators to score the sectors in terms of their role in knowledge transfer system

Proposed a methodology, related to economic input-output analysis, that can be used to analyze the intersectoral knowledge flow.

Knowledge flow representation-matrix

Old way	New way
• Number of direct citations from science paper and patent	• Knowledge flow table • Defined the matrix using v'(percentage of citations from academic paper) and A(percentage of citations from other patents) • Summation of direct and indirect citations from past

Knowledge flow representation-indicators

Indicator	Explanation
Backward multipliers	Measures the amount of patent-to-patent citations that are necessary to pull one additional unit of science knowledge from the past
Forward multipliers	Measures the amount of patent-to-patent citations that are necessary to carry forward one additional unit of science knowledge into future
Net science multipliers	Relative strength of the patents in terms of their double role in performing as sources and sinks at the same time
Self-use	It is the share of sector j-specific knowledge in the composite knowledge mix of the sector j patents
Self-supply	how much a sector generates internal knowledge versus knowledge that is used by other sectors
Use pervasiveness	Indicates the variety in inter-industrial backward citation linkage between industry j and all other industry
Supply pervasiveness	Indicates the variety in inter-industrial forward citation linkage between industry j and all other industries, and is a measure of how pervasively industry I influence the other industries in the in system

Three Main Findings

high science intensity

also implies net supply of knowledge to other sectors

scientific knowledge is highly specialized and specific

the number of science citations per patent is not a good indicator

for the knowledge pervasiveness of sectors

Appendix A

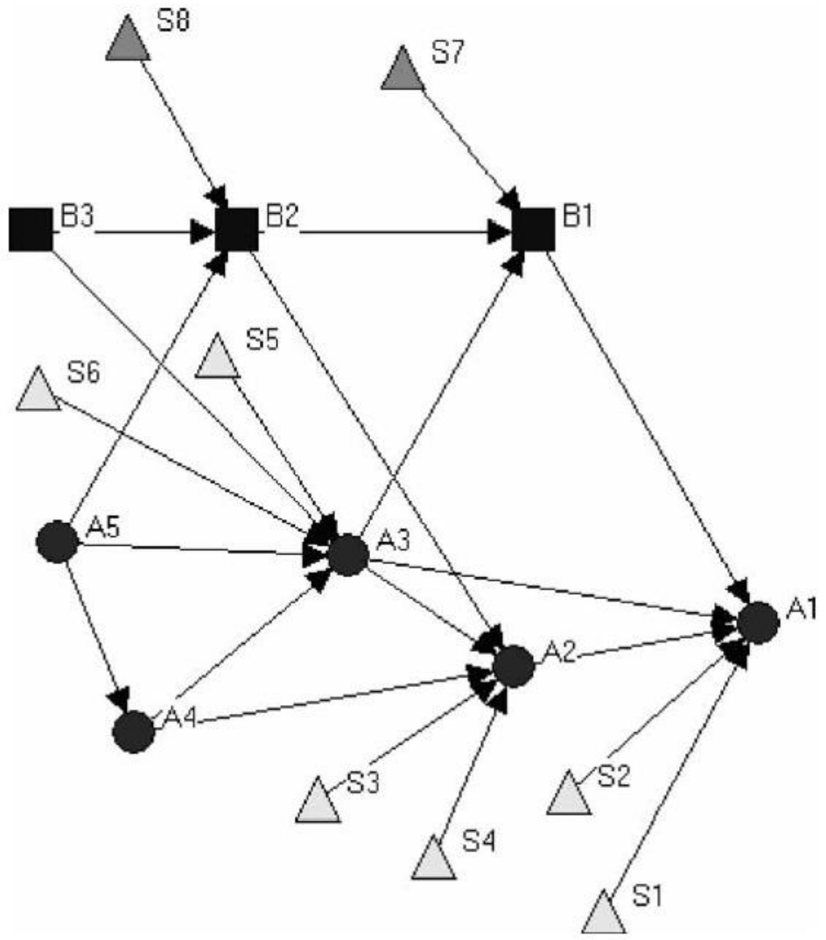

Figure 1. Stylized patent-citation network graph

Patent-citation network graph

- Nodes represent ideas
- Type A and B are technology/industry/ patent
- Type S is an academic paper

348 O. Nomaler & B. Verspagen

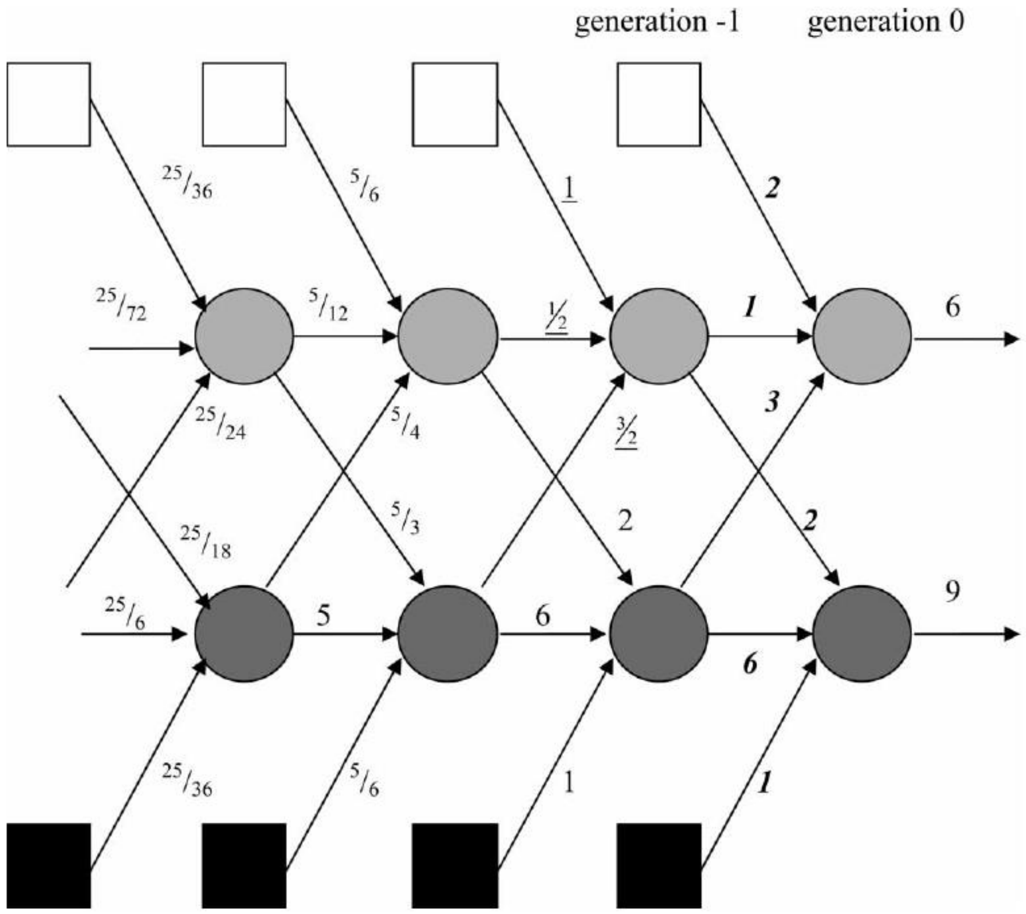

Figure 2. Stylized citation network graph used to illustrate formal approach

Appendix B

- Used this graph to explain how the idea of input-output analysis is used to define the knowledge flow table

Appendix C

- Construction of citation flow table

Table 2. The construction of the citations-flow table

INTERMEDIATE CITATION FLOWS IN PERIOD $-\infty$ TO t: $(n \times n)$ The matrix **F**, with elements $a_{ij} \cdot x_j$	**TOTAL SCIENCE KNOWLEDGE PROVIDEDTO POST-t PATENTS** $(n \times 1)$ $\mathbf{g} = \mathbf{c}$ **(by assumption): TOTAL CITATIONS MADE IN YEAR t**	**CITATIONS RECEIVED IN PERIOD $-\infty$ TO t:** $(n \times 1)$ $\mathbf{y} = \mathbf{Fi} + \mathbf{g} = (\mathbf{I}-\mathbf{A})^{-1}\mathbf{g} = \mathbf{x}$
TOTAL SCIENCE KNOWLEDGE TAKEN IN PERIOD $-\infty$ TO t: $(1 \times n)$ $\mathbf{s}' = (\mathbf{Dg})' = [\hat{\mathbf{v}}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{c}]'$		
CITATIONS MADE IN PERIOD $-\infty$ TO t: $(1 \times n)$ $\mathbf{x}' = \mathbf{s}'\hat{\mathbf{v}}^{-1} = [(\mathbf{I}-\mathbf{A})^{-1}\mathbf{c}]'$		

Appendix D

Table 3. Indicator scores of the sectors (definitions are given in the text, all data refer to calculations made with 1992 data)

Sector	Science citations per patent	Backward multiplier	Forward multiplier	Net science multiplier	Self-use	Self-supply	Use pervasiveness	Supply pervasiveness
Electrical machinery	0.48	4.91	5.05	0.97	0.55	0.56	6.64	8.30
Electronics	1.05	3.39	3.90	1.14	0.78	0.69	5.11	6.65
Chemistry	0.62	4.67	5.17	1.03	0.52	0.50	3.60	8.10
Pharmaceuticals	1.91	2.84	3.98	1.34	0.81	0.61	3.89	4.92
Oil refining	0.18	6.76	6.01	0.43	0.18	0.43	4.62	6.17
Motor vehicles	0.13	9.11	7.38	0.54	0.29	0.53	10.04	7.44
Other transport	0.19	7.71	6.60	0.66	0.29	0.44	10.07	6.47
Ferrous basic metals	0.56	5.00	5.21	1.08	0.47	0.44	8.64	9.24
Non-ferrous basic metals	0.83	4.41	5.11	1.31	0.55	0.42	8.70	9.71
Metal products	0.21	6.76	6.08	0.72	0.29	0.40	10.39	8.32
Instruments	0.61	4.74	4.87	1.00	0.63	0.63	8.05	9.55
Computers and office equipment	0.98	3.61	4.00	1.10	0.74	0.68	4.26	5.71
Other machinery	0.18	6.95	6.14	0.59	0.28	0.47	8.71	10.26
Food products	1.43	3.25	4.34	1.33	0.59	0.44	3.07	6.10
Textiles	0.22	6.75	6.09	0.68	0.21	0.30	7.94	6.57
Rubber and plastic products	0.08	7.80	6.39	0.31	0.09	0.29	9.04	7.65
Stone, clay and glass products	0.53	5.42	5.65	1.16	0.40	0.35	10.03	9.21
Paper and printing	0.36	5.79	5.51	0.82	0.30	0.37	9.18	9.59
Other manufacturing	0.24	6.30	5.62	0.60	0.21	0.35	9.15	9.33

Left most column: number of non-patent citations per patent

Appendix F

- Correlations between indicators

Table 4. Correlation coefficients between the indicators

	Science citations per pat.	Backward multiplier	Forward multiplier	Net science multiplier	Self-use	Self-supply	Use pervasiveness	Supply pervasiveness
Science citations per pat.	1.00							
Backward Multiplier	−0.88	1.00						
Forward Multiplier	−0.85	0.98	1.00					
Net science multiplier	0.85	−0.89	−0.80	1.00				
Self-use	0.86	−0.88	−0.88	0.87	1.00			
Self-supply	0.54	−0.53	−0.60	0.46	0.82	1.00		
Use pervasiveness	−0.69	0.70	0.71	−0.46	−0.61	−0.52	1.00	
Supply pervasiveness	−0.46	0.22	0.27	−0.10	−0.29	−0.32	0.61	1.00