Quality of Bibliometric Databases: Accuracy in Classification of Document Types

Universitat de Girona

Quality of Bibliometric Databases: Accuracy in Classification of Document Types

D.A. Maisano, L. Mastrogiacomo, L. Ferrara and F. Franceschini

Politecnico di Torino, Dept. of Management and Production Engineering

STRUCTURED ABSTRACT

Purpose: Scholarly publications are usually classified into *document types* (DTs), which are predefined categories outlining their nature (e.g., *research articles*, *conference proceedings*, *reviews*, *short notes*, *letters*, *book chapters*, etc.). This research presents a new semi-automated methodology to assess the accuracy of DT classification in bibliometric databases, such as Scopus and Web of Science (WoS). The methodology can handle a relatively large amount of documents (on the order of tens/hundreds of thousands) and is adaptable to the different classes of DTs covered by the databases in use, without requiring an *a priori* definition of a correspondence between their DTs.

Methodological approach: The first phase of the proposed methodology is automated and exploits discrepancies in DT classifications by two competing databases (e.g., Scopus and WoS), in order to identify a subset of potentially misclassified documents, i.e., with possible DT-classification errors. The second phase involves the manual analysis of this subset of documents, resulting in the identification and attribution of DT-classification errors. The novel methodology is illustrated through a realistic application example.

Findings: The methodology is shown to be effective in identifying DT-classification errors, suggesting a path to improve the quality and reliability of bibliometric databases. With reference to the application example provided, Scopus and WoS have overall error rates around 1.7% and 1.2%, respectively. A similar analysis based on a larger sample of documents is still in progress.

Practical/social implications: By improving database accuracy, the academic community can benefit from more reliable bibliometric indicators, which can affect (at least to some extent) research funding, decision making and academic reputation.

Keywords: Bibliometric databases, Document type, Semi-automated analysis, Database accuracy.

Paper type: Research paper

INTRODUTION

Scientific publications are typically categorized into *document types* (DTs), i.e., predefined categories that outline their nature and primary characteristics (Harzing, 2013; Yeung, 2021). Common DTs include *research articles*, *conference proceedings*, *surveys*, *letters*, *book chapters*, etc. The classification of publications into significant DTs is a task performed by the publishers and/or bibliometric databases indexing them, so as to serve multiple purposes (Donner, 2017). Firstly, it contributes to the organisation of knowledge, aiding researchers in efficiently selecting and retrieving relevant information. Additionally, it plays a crucial role in research evaluation, differentiating scientific contributions for the construction of bibliometric indicators and assessments, e.g., DTs that are usually subject to less scrutiny during acceptance, such as *conference proceedings*, *short notes* and *letters*, may be excluded from some evaluations. Furthermore, most bibliometric indicators relating to scientific journals, like the well-known *impact factor*, are constructed by considering the DT of indexed documents (García-Pérez, 2010).

Unfortunately, DT classification is not always accurate, leading to instances where scientific contributions are mislabelled. The literature documents common mislabelling scenarios, such as misclassifying *research articles* as *reviews*, *letters* and *conference proceedings* as *research articles*, which indicates that some types of errors tend to be more frequent than others (Sigogneau, 2000; Yeung, 2019; Mokhnacheva, 2023). The misclassification of DTs can be somewhat fed by the absence of standardized rules, which results in subjectivity. The distinction between a *research article*, a *review*, or a *short note*, for example, can be quite subtle. Moreover, the DT-classification rules used by both bibliometric databases and publishers are not completely transparent, nor the metadata they exchange. Furthermore, there are quite often discrepancies in the nomenclature and definitions of DTs between different publishers and databases. These discrepancies become evident when comparing the official DT lists from major generalist databases, like Scopus and Web of Science (WoS) (Clarivate, 2024; Elsevier, 2024).

Relatively few studies on DT-classification errors are documented in the literature, forming a part of the broader category of research on errors in bibliometric databases (Donner, 2023; Franceschini et al., 2013; 2015; 2016; García-Pérez, 2010; Moed, 2005; Olensky et al., 2016; Valderrama-Zurián et al., 2015). These studies indicate that DT-classification errors are significant for both Scopus and WoS, amounting to a few percentage points (Yeung, 2021). It was also noted that specialized databases – like PubMed in the medical field – tend to have fewer errors compared to generalist databases such as Scopus and WoS, possibly due to a smaller and more homogenous set of indexed

documents (Yeung, 2019). The few existing studies on DT-classification errors unfortunately have (at least) a couple of limitations. First, they generally limit the investigation to relatively small samples of a few hundred or thousand documents. In addition, they only involve manual analysis of the documents of interest.

This research introduces a novel semi-automated methodology to investigate DT-classification errors on large quantities of documents. The methodology includes an initial phase to determine a relatively large *corpus* of publications, simultaneously indexed by the (competing) generalist Scopus and WoS databases. Having determined the DTs assigned by the aforementioned databases to each publication, a subset of publications with potentially inconsistent/discordant DTs is determined. Manual analysis of this subset is then performed, identifying any DT-classification errors and attributing them to the responsible database. The new methodology, which can be classified as "adaptive" since it requires no *a priori* correspondence between the DTs covered by one database and those covered by the other, is described through an application example.

NEW METODOLOGY

The following pseudo-code summarizes the basic steps of the new methodology for identifying DT-classification errors, which is described in detail in the rest of the section.

Start

1. **Identify a *corpus* of scientific publications.** Use specific criteria to select a relatively broad group of scientific publications (e.g., scientific output from research groups, institutions, scientific journals, etc., within a certain time window).
2. **Query Scopus and WoS.** Conduct separate queries in the Scopus and WoS databases using criteria that define the *corpus* of publications at step (1).
3. **Determine the "intersection" of publications indexed by both databases.** Using suitable unique identifiers (e.g., DOI code), identify the subset of publications resulting from the queries in step (2), indexed in both databases.
4. **Determine the DTs for each publication at step (3), according to both Scopus and WoS.** For each publication in the intersection subset, identify the DTs classified by Scopus and WoS respectively.
5. **Construct the concordance matrix.** Create a so-called *concordance matrix* (described later) to compare the DTs assigned by the two databases of interest. Arrange the relevant DTs (in columns for Scopus and in rows for WoS), in order to maximize the amount of (concordant) publications in the diagonal.

6. **Identify publications with discordant DTs.** Focus on the scientific publications outside the diagonal of the concordance matrix, as their DT-classifications are potentially discordant.
   6.1 *Manual analysis.* Manually analyse each discordant publication and identify the most plausible DT for it (i.e., the "true" DT).
   6.2 *Determine a DT-classification error for Scopus or WoS.* Determine which of the two databases is responsible for the (presumed) DT-classification error and record it.
   6.3 *Return to step (6.1).* Repeat the manual analysis until all publications with discordant DTs have been analysed.
7. **Determine error statistics.** Process the results of the analysis and construct database-error statistics.
   7.1 *Consider each specific database separately and report the relevant DT-classification errors in a so-called error table (described later).*
   7.2 *Determine different kinds of error statistics*, both from the overall perspective of a database of interest (ε) and from the perspective of a specific DT (contemplated by the database itself), i.e., calculate the rate of *false exclusions* (α) and the rate of *false classifications* (β) related to that DT.
   7.3 *Return to step (7.1).* Repeat the calculation of error statistics for the other database.

End

First, it is necessary to identify a relatively large sample of scholarly publications (*step 1*). By way of example, let us assume that the governing bodies of the two major universities in Turin (Italy) – Politecnico di Torino (a technical university of engineering and architecture) and Università di Torino (a generalist university) – have decided to undertake a campaign to verify the accuracy of the data reported in the Scopus and WoS databases, in anticipation of a forthcoming national evaluation exercise of the scientific output produced in 2019-2023. To this end, the publications produced in the five-year period of interest by the more than two thousand researchers affiliated with the two universities are considered. Then, both Scopus and WoS databases are queried (*step 2*) to extract data on the publications of interest (e.g., using co-author affiliation and issue years). This yields two sets of publications: one returned by Scopus and one returned by WoS. About 34 thousand articles were obtained from Scopus and 38 thousand from WoS. Let us note that the two sets are not of identical size because the two databases differ in terms of source coverage (e.g., one database may index some journals or conference proceedings that are not necessarily indexed by the other database and *vice versa*).

The "intersection" (*step 3*) between the above two sets, i.e., the subset of common documents, is then determined. The identification of these common documents can be accomplished using the relevant *digital object identifier* (or DOI code); therefore, documents without a DOI code are necessarily

excluded from the analysis. In the application example, we restricted the analysis to the DTs of *research article* (or more simply *article*), *conference proceeding*, *letter*, and *review*, neglecting for simplicity the other DTs (e.g., *book chapters*, *monographs*, etc.). Returning to the application example, cross-referencing the two sets of documents returned by Scopus and WoS yields an intersection subset, consisting of 26,405 total documents, classified by each database into corresponding DTs, as shown in Table 1.

Table 1 – Summary of DTs classified by Scopus and WoS with reference to the "intersection" subset, in the application example.

Scopus DTs		WoS DTs	
Article	21,793	*Article*	21,800
Review	2,445	*Review*	2,472
Conference paper	1,810	*Proceedings paper*	1,756
Letter	357	*Letter*	377
Total	**26,405**	**Total**	**26,405**

In this case the DT labels of the two databases are coincident, except for *conference paper*, as contemplated by Scopus, and *proceedings paper*, as contemplated by WoS[1]. The amounts of articles classified in the DTs of the two databases are very close, although the differences indicate possible discrepancies in the classification of the DTs, as better evidenced in the so-called concordance matrix (step 5). This matrix (exemplified in Table 2) is nothing more than a contingency table showing in column the DTs assigned by Scopus and in row the DTs assigned by WoS for the publications of interest. The rows and columns are sorted in such a way that the row totals and column totals are both sorted in descending order. In this way the matrix is *diagonal*, that is, with most of the papers placed on the main diagonal. At the same time, this sort of diagonalization establishes an empirical correspondence between the DTs covered by one and the other database. This correspondence can be called "adaptive" in that it does not require any *a priori* link between Scopus and WoS DTs (Owen, 2001). In other words, among the possible permutations between Scopus and WoS DTs, the proposed diagonalization ensures the highest degree of concordance. Off-diagonal elements indicate discordant DT classifications, denoting potential errors by (at least) one database. For this reason, these elements can be classified as "discordant" (and by extension the documents involved in them).

[1] In fact, expanding the study to other more specialized DTs, the differences between the databases would be more pronounced. For example, DTs covered by one database but not the other, such as *biographical-item*, *book review*, *expression of concern*, *fiction*, *meeting abstract*, *retraction*, *data paper*, and many others, can be observed (Clarivate, 2024; Elsevier, 2024).

Table 2 – Example of *concordance matrix* related to data collected for the application example. While the elements in the diagonal (in square brackets) denote DT classifications that are concordant between competing databases, those off-diagonal denote possible DT-classification errors.

DT classifications →		by Scopus				
↓		*Article*	*Review*	*Confer. paper*	*Letter*	**Row total**
by WoS	*Article*	[21,418]	297	71	14	**21,800**
	Review	323	[2,146]	2	1	**2,472**
	Proceed. paper	19		[1,737]	-	**1,756**
	Letter	33	2	-	[342]	**377**
	Column total	**21,793**	**2,445**	**1,810**	**357**	**26,405**

Manual analysis of each of the discordant publications (sub-step 6.1) is performed using all available information, such as abstract, data on the publisher's site, "fulltext", etc. This analysis is aimed at identifying the "true" DT and, consequently, detect the possible DT-classification error of the databases. Figure 1 exemplifies the heading of a paper classified by Scopus as a *conference paper* and by WoS as an *article*. Manual analysis revealed that the "true" DT is *article*, since the paper is a (complete) research contribution unrelated to any conference. The DT classification error is therefore attributed to Scopus in this case.

European Journal of Operational Research 275 (2019) 1020–1036

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

On the usefulness of soft OR models in decision making: A comparison of Problem Structuring Methods supported and self-organized workshops

Isabella M. Lami[a,*], Elena Tavella[b]

[a] Interuniversity Department of Regional and Urban Studies and Planning (DIST), Politecnico di Torino, Viale Mattioli 39, 10125 Turin, Italy
[b] Department of Food and Resource Economics, University of Copenhagen, Rolighedsvej 25, 1958 Frb. C, Denmark

ARTICLE INFO

Article history:
Received 6 October 2017
Accepted 12 December 2018
Available online 23 December 2018

Keywords:
Problem structuring
Model usefulness
Questionnaires
Reflective discussions

ABSTRACT

Applications of Problem Structuring Methods (PSMs) are extensively reported in the literature; however, their evaluation remains challenging. Moreover, scholars' disagreement on how to evaluate and compare PSMs constrains research that seeks to show whether certain PSMs are more useful than others within particular contexts and 'better than doing nothing'. Drawing on scholars' suggestion to evaluate PSMs through a link between action, outcomes and context, we address this gap by adopting an exploratory, quasi-experimental research design with MSc students to evaluate and compare three workshops supported by the use of (i) Strategic Choice Approach, (ii) Soft Systems Methodology and (iii) a non-PSM supported approach, and report on the findings of a quantitative and qualitative analysis. We explore the *usefulness* of PSMs for achieving outcomes that can support decision making. Our findings show variation in the usefulness of the three approaches and the different workshop outcomes.

Figure 1 – Example of document misclassified by Scopus as *conference paper* instead of *article* (https://doi.org/10.1016/j.ejor.2018.12.016).

Entomologia Experimentalis et Applicata

FOUNDED 1958

DOI: 10.1111/eea.12967

SPECIAL ISSUE: 6TH INTERNATIONAL ENTOMOPHAGOUS INSECTS CONFERENCE

Biological control of invasive stink bugs: review of global state and future prospects

Eric Conti[1]*, Gonzalo Avila[2,3], Barbara Barratt[3,4], Fernanda Cingolani[5], Stefano Colazza[6], Salvatore Guarino[7], Kim Hoelmer[8], Raul Alberto Laumann[9], Lara Maistrello[10], Guillaume Martel[11,12], Ezio Peri[6], Cesar Rodriguez-Saona[13], Gabriele Rondoni[1], Michael Rostás[14], Pio Federico Roversi[15], René F.H. Sforza[11], Luciana Tavella[16] & Eric Wajnberg[17,18]

[1]*Department of Agricultural, Food and Environmental Sciences, University of Perugia, Borgo XX Giugno, Perugia 06121, Italy,* [2]*The New Zealand Institute for Plant & Food Research Limited, Private Bag 92169, Auckland 1142, New Zealand,* [3]*Better Border Biosecurity, Christchurch New Zealand,* [4]*AgResearch, Invermay Research Centre, Private Bag 50034, Mosgiel 9053, New Zealand,* [5]*CEPAVE (CONICET-UNLP), Boulevard 120 e/ 60 y 64 s/n, La Plata 1900, Argentina,* [6]*Dipartimento di Scienze Agrarie, Alimentari e Forestali, Università degli Studi di Palermo, Viale delle Scienze, edificio 5, Palermo 90128, Italy,* [7]*Institute of Biosciences and Bioresources (IBBR), National Research Council of Italy (CNR), Corso Calatafimi 414, Palermo 90129, Italy,* [8]*USDA–ARS, Beneficial Insects Introduction Research Unit, 501 S. Chapel St, Newark DE, USA,* [9]*Embrapa Recursos Genéticos e Biotecnologia. PqEB Avda W5 Norte (Final), CEP 70770-917, Brasília DF, Brazil,* [10]*Dipartimento di Scienze della Vita, Centro BIOGEST-SITEIA, Università di Modena e Reggio Emilia, Via G. Amendola 2, Reggio-Emilia 42122, Italy,* [11]*USDA-ARS-European Biological Control Laboratory, Campus International de Baillarguet, 810, avenue du Campus Agropolis, Montferrier-sur-Lez 34980, France,* [12]*Montpellier SupAgro, Place Pierre Viala, 2, Montpellier 34000, France,* [13]*Department of Entomology, Rutgers University, New Brunswick NJ, 08901, USA,* [14]*Agricultural Entomology, Department of Crop Sciences, University of Göttingen, Grisebachstr. 6, Göttingen 37077, Germany,* [15]*CREA Consiglio per la Ricerca in Agricoltura e l'analisi dell'economia agraria, Via di Lanciola, 12/a, Cascine del Riccio, Firenze 50125, Italy,* [16]*Dipartimento di Scienze Agrarie, Forestali e Alimentari (DISAFA), Università degli Studi di Torino, Largo Paolo Braccini 2, Grugliasco TO, 10095, Italy,* [17]*INRA, 400 Route des Chappes, BP 167, Sophia Antipolis Cedex 06903, France, and* [18]*INRIA, Sophia Antipolis, Project Hephaistos,2004 Route des Lucioles, BP 93, Sophia AntipolisCedex 06902, France*

Accepted: 11 June 2020

Key words: biocontrol, chemical ecology, Hemiptera, invasive species, landscape management, pre-emptive classical biological control, semiochemicals, parasitoid, Pentatomidae, risk assessment, Tachinidae, Scelionidae

Abstract

Invasive stink bugs (Hemiptera: Pentatomidae) are responsible for high economic losses to agriculture on a global scale. The most important species, dating from recent to old invasions, include *Bagrada hilaris* (Burmeister), *Halyomorpha halys* (Stål), *Piezodorus guildinii* (Westwood), *Nezara viridula* (L.), and *Murgantia histrionica* (Hahn). *Bagrada hilaris*, *H. halys*, and *N. viridula* are now almost globally distributed. Biological control of these pests faces a complex set of challenges that must be addressed to maintain pest populations below the economic injury level. Several case studies of classical and conservation biological control of invasive stink bugs are reported here. The most common parasitoids in their geographical area of origin are egg parasitoids (Hymenoptera: Scelionidae, Encyrtidae, and Eupelmidae). Additionally, native parasitoids of adult stink bugs (Diptera: Tachinidae) have in some cases adapted to the novel hosts in the invaded area and native predators are known to prey on the various instars. Improving the efficacy of biocontrol agents is possible through conservation biological control techniques and exploitation of their chemical ecology. Moreover, integration of biological control with other techniques, such as behavioural manipulation of adult stink bugs and plant resistance, may be a sustainable pest control method within organic farming and integrated pest management programs. However, additional field studies are needed to verify the efficacy of these novel methods and transfer them from research to application.

Figure 2 – Example of a paper incorrectly classified by Scopus as an *article* and by WoS as a *review*, instead of *conference/proceedings paper* (https://doi.org/10.1111/eea.12967).

Very rarely, simultaneous errors of both databases can be observed. As an example, let us consider the document in Figure 2, classified by Scopus as an *article* and by WoS as a *review*. Manual analysis

found that this document is in fact a *conference paper* on a journal special issue, resulting in an error for both databases. Probably the Scopus classification error (i.e., *article* instead of *conference paper*) stems from the fact that this *conference paper* is published in a journal special issue. On the other hand, the WoS error (i.e., *review* instead of *proceedings paper*) stems from the fact that the word "review" is used both in the title and within the conclusions to say that the paper actually reviews more than a century of research on the biological control of invasive stink bugs. Additionally, an extensive reference list is presented (more than 200 references).

For the diagonal elements (in square brackets), the DT classifications in both databases are automatically assumed to be correct, given the general agreement between the counterpart databases. To test the plausibility of this assumption, a sampling of 0.5 percent of these concordant documents (i.e., $\lceil 0.5\% \cdot 25{,}643 \rceil = 129$) was conducted, without detecting any DT-classification error. This confirms the plausibility of the hypothesis of validity of DT assignments for these publications, in line with the concepts of *convergent validity*, and *wisdom of crowds* (Franceschini et al., 2022).

Having completed the manual analysis of the discordant documents (cf. step 6 and relevant sub-steps), the DT-classification errors found for each database are summarized in an *error table*, i.e., a contingency table illustrating in the columns the DT assignments made by the database of interest and in the rows the "true" (or correct) DTs resulting from the manual analysis. In practice, the main diagonal of the error table contains the correct DT classifications while the off-diagonal elements correspond to the incorrect DT classifications. Tables 3 and 4 show the error tables related to Scopus and WoS, with reference to the application example. Let us note that for a small portion of the DT-classification errors, the "true" DT is *other DTs*, which is different from the four DTs considered in this simplified study. In fact, this DT class includes all other DTs not covered in this study (e.g., *book review*, *note*, *short survey*, etc.). For example, Figure 3 exemplifies a publication in the field of forensic psychiatry, containing a brief report of a clinical case with some references to existing literature on the subject. The contribution lacks the consistency and originality to be considered a (research) *article*, nor does it provide a sufficiently thorough review of the state of the art to be considered a *review*. Unsurprisingly, it is classified by the journal as a *case report*. This publication is misclassified by Scopus as an *article* and by WoS as a *review*, whereas our manual analysis showed that the most plausible Scopus and WoS DTs would respectively be *note* and *other*, which would converge in the class *other DTs* in the present simplified analysis (Clarivate, 2024; Elsevier, 2024). Furthermore, it was noted that Scopus tends to classify *case-report* publications as *articles*, whereas WoS tends to classify them as *reviews*, resulting in a systematic classification error.

Table 3 – Example of *error table* for Scopus, with reference to the application example. Error statistics (see Eqs. 1, 2, and 3) are bolded.

		DT classification by Scopus					
		Article	*Review*	*Confer. paper*	*Letter*	Row total	$\boldsymbol{\alpha_{Scopus,\,DT}}$
"True" DT classification	*Article*	[21,503]	102	53		21,658	**0.7%**
	Review	233	[2,331]			2,564	**9.1%**
	Confer. paper	20	5	[1,757]		1,782	**1.4%**
	Letter	30	2		[357]	389	**8.2%**
	Other DT(s)	7	5			12	-
	Column total	21,793	2,445	1,810	357	26,405	
	$\boldsymbol{\beta_{Scopus,\,DT}}$	**1.3%**	**4.7%**	**2.9%**	**0.0%**		$\boldsymbol{\varepsilon_{Scopus} \cong 1.7\%}$

Table 4 – Example of *error table* for WoS, with reference to the application example. Error statistics (see Eqs. 1, 2, and 3) are bolded.

		DT classification by WoS					
		Article	*Review*	*Confer. paper*	*Letter*	Row total	$\boldsymbol{\alpha_{WoS,\,DT}}$
"True" DT classification	*Article*	[21,584]	75		3	21,662	**0.4%**
	Review	176	[2,386]			2,562	**6.9%**
	Proc. paper	23	3	[1,756]		1,782	**1.5%**
	Letter	16	1		[374]	391	**4.3%**
	Other DT(s)	5	7			12	-
	Column total	21,804	2,472	1,756	377	26,405	
	$\boldsymbol{\beta_{WoS,\,DT}}$	**1.0%**	**3.5%**	**0.0%**	**0.8%**		$\boldsymbol{\varepsilon_{WoS} \cong 1.2\%}$

Next, some indicators depicting database errors are constructed. An overall indicator of the DT-classification errors by a certain database (ε_{db}) is expressed by the ratio of the total number of off-diagonal elements (i.e., depicting erroneous DT classifications resulting from manual analysis) to all elements in the *error table* (i.e., the total number of DT classifications):

$$\varepsilon_{db} = \frac{\sum_{\forall(i,j)\,|\,i\neq j} d_{db,\,(i,j)}}{\sum_{\forall(i,j)} d_{db,\,(i,j)}}, \tag{1}$$

being $d_{db,\,(i,j)}$ the number of documents reported in the i-th row and j-th column of the error table, for the database of interest (*db*).

The ε_{db} values resulting from the analysis, also shown in the lower right vertex of the error tables, are $\varepsilon_{\text{Scopus}} \cong 1.7\%$ and $\varepsilon_{\text{WoS}} \cong 1.2\%$. Although aware that these results are difficult to generalize as they relate to a rather small sample of documents, it is interesting to note some general trends. First, the ε_{db} values are very close for the two databases, denoting a slightly higher error rate for Scopus with respect to WoS. In both cases, the most frequent DT-classification errors involve *reviews* misclassified as *articles* and *vice versa*. *Letters* misclassified as *articles* are also fairly frequent for

both databases. In addition, while Scopus has an erroneously categorized some *articles* as *conference papers* (cf. the example in Figure 1), WoS did not incur this error.

Forensic Science, Medicine and Pathology (2019) 15:276–280
https://doi.org/10.1007/s12024-019-00105-6

CASE REPORT

Check for updates

The risk of assault against mental health professionals: a fatal case report and literature review

Lucia Tattoli[1] · Caterina Bosco[2] · Ignazio Grattagliano[3] · Giancarlo Di Vella[2]

Accepted: 24 February 2019 / Published online: 26 April 2019

Abstract
Assaults by patients against healthcare providers are an increasing phenomenon worldwide. Mental health professionals in acute facilities and rehabilitation wards have the highest risk of being attacked at work. Verbal abuse or intimidating behaviors represent the most common types of violence. Fatal assault by psychiatric patients has been rarely reported in the literature. We present a case of a female psychiatrist who was fatally stabbed in her office in a Mental Health Center. At autopsy seventy stab wounds were found: four wounds of the neck, fifty penetrating wounds of the thorax, three wounds of the abdomen, six wounds of the lumbar region, and seven wounds of the upper arms including defense injuries. The cause of death was massive blood loss due to multiple stab wounds. The perpetrator was a 44-year-old male patient who had been referred to the victim after a previous admission to hospital following experiences of suicidal ideation and confusion. The extreme and unmotivated violence in a non-acute setting were notable. A borderline-antisocial personality disorder was later diagnosed by forensic experts. This case emphasizes the significant occupational risk for mental healthcare staff to sustain life threatening injuries or death, with implications for training of clinicians, and strategies for preventing aggressive behaviors.

Keywords Workplace violence · Psychiatry · Violence · Stab wounds

Figure 3 – Example of a paper incorrectly classified by Scopus as an *article* and by WoS as a *review* (https://doi.org/10.1007/s12024-019-00105-6). The article is actually defined by the publisher (Springer Nature) as a *case report* in the field of forensic psychiatry. Manual analysis revealed that the most plausible DT for Scopus would be *note* and for WoS would be *other* (Clarivate, 2024; Elsevier, 2024).

Next, other more detailed database-error statistics can be constructed from the perspective of individual DTs. In this regard, a *null hypothesis* (H_0) is defined as: "*The DT assigned by the database in use is correct*". Next, two DT-related error types are defined, as described below.

1. The *Type-I error* corresponds to the probability of wrongly rejecting H_0 when it is true, i.e. the probability of a document belonging to a specific DT being wrongly classified into another DT (i.e., *false exclusion from the DT of interest*):

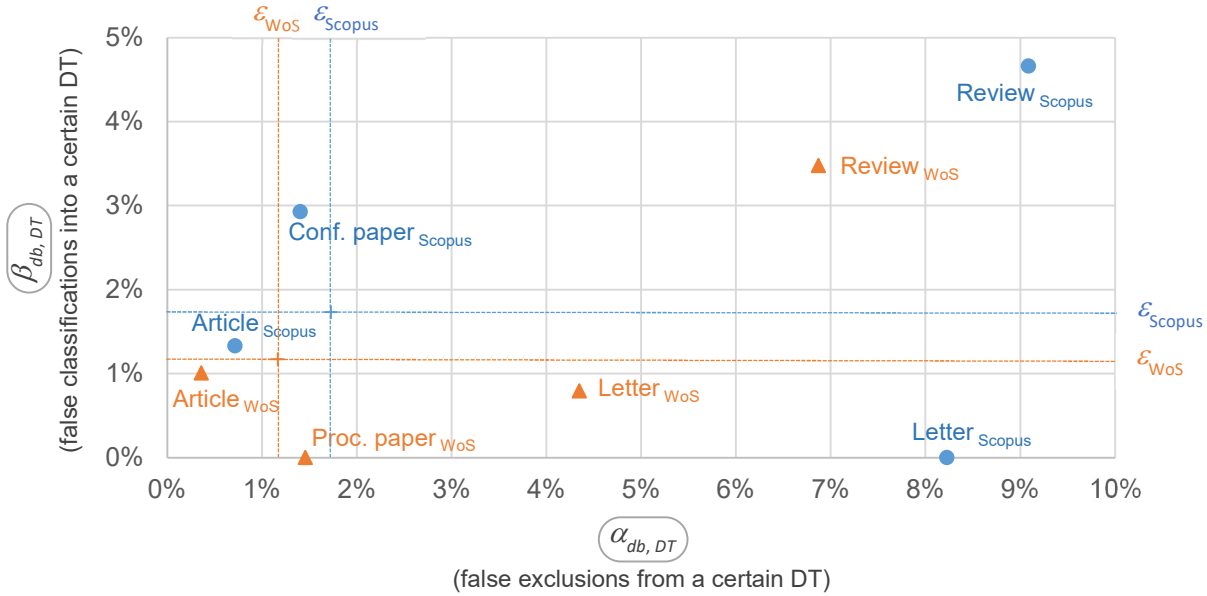
$$\alpha_{db,\,DT} = \frac{\sum_{\forall j \mid j \neq DT} d_{db,\,(DT,j)}}{\sum_{\forall j} d_{db,\,(DT,j)}}, \tag{2}$$

DT being the DT of interest for the database (*db*) in use. $\alpha_{db,\,DT}$ is actually calculated on a row-by-row basis by means of the ratio of misclassifications (i.e., false exclusions from the DT of interest) to the row total.

2. The *Type-II error* corresponds to the probability of failing to reject H_0 when it is false, i.e. the probability of misclassifying a document into a specific DT of interest (i.e., *false classification into the DT of interest*):

$$\beta_{db,\,DT} = \frac{\sum_{\forall i \mid i \neq DT} d_{db,\,(i,\,DT)}}{\sum_{\forall i} d_{db,\,(i,\,DT)}}, \tag{3}$$

$\beta_{db,\,DT}$ is actually calculated on a column-by-column basis by means of the ratio of misclassifications (i.e., false classifications into the DT of interest) to the column total. Interestingly, the same element ($d_{db,\,(i,j)}$) in the error table contributes to both of these two types of error. With reference to the application example, the values of $\alpha_{db,\,DT}$ and $\beta_{db,\,DT}$ are reported respectively in the right and lower parts of the corresponding error tables (Tables 3 and 4). Again, some similarity is observed between the results obtained for the two databases. The map in Figure 4 provides a graphical representation of the above error statistics. The dashed lines represent $\varepsilon_{\text{Scopus}}$ and ε_{WoS} values, which can be read on both



the horizontal and vertical axes. In addition, for each of the DTs covered by a certain database, a corresponding point of coordinates ($\alpha_{db,\,DT}$, $\beta_{db,\,DT}$) is plotted.

Figure 4 – Map of error statistics, referring to the application example.

CONCLUSIONS

This paper proposed a new methodology for semi-automated analysis of a relatively large amount of scientific publications, which is useful in identifying DT-classification errors in the Scopus and WoS databases. Taking advantage of the discordance in DT classification between the two databases, the proposed methodology directly targets a subset of potentially misclassified documents for manual

analysis. In the application example provided, against a total *corpus* of 26,405 documents in both databases, only the 762 off-diagonal documents in the concordance matrix – corresponding to about 2.9% of the total – were manually analysed. Additionally, 0.5% of the 25,643 diagonal documents (i.e., 129) were analysed randomly, finding no DT-assignment error and confirming the plausibility of the working hypothesis that concordant DT classifications are unlikely to be erroneous.

This methodology could be used by individual scientists, institutions and even database managers to detect (and correct) DT-classification errors relatively quickly. Having license to use both databases, the novel methodology can be simply implemented using spreadsheets and/or pivot tables. In the future, the possibility of assisting the time-consuming manual analysis with AI-based techniques will be explored. In addition, a similar investigation based on a larger sample of publications is underway to determine more representative results of DT-classification errors from the two databases under consideration.

AKNOWLEDGEMENTS

This study was carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and was partially funded by the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1551.11-10-2022, PE00000004). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

REFERENCES

Clarivate (2024). Web of Science Help – Document Types, https://webofscience.help.clarivate.com/en-us/Content/document-types.html [accessed May 2024].

Donner, P. (2017). Document type assignment accuracy in the journal citation index data of Web of Science. Scientometrics, 113(1), 219-236.

Donner, P. (2023). Data inaccuracy quantification and uncertainty propagation for bibliometric indicators. arXiv preprint arXiv:2303.16613.

Elsevier (2024). Scopus Content, https://www.elsevier.com/products/scopus/content [accessed May 2024].

Franceschini, F., Maisano, D., Mastrogiacomo, L. (2013). A novel approach for estimating the omitted citation rate of bibliometric databases with an application to the field of bibliometrics. Journal of the American Society for Information Science and Technology, 64/10: 2149–56.

Franceschini, F., Maisano, D., Mastrogiacomo, L. (2015). Errors in DOI indexing by bibliometric databases. Scientometrics, 102/3: 2181–6.

Franceschini, F., Maisano, D., Mastrogiacomo, L. (2016). Empirical analysis and classification of database errors in Scopus and Web of Science. Journal of informetrics, 10(4), 933-953.

Franceschini, F., Maisano, D., Mastrogiacomo, L. (2022) Rankings and Decisions in Engineering: Conceptual and Practical Insights. International Series in Operations Research & Management Science Series, Vol. 319, Springer International Publishing, Cham (Switzerland).

García-Pérez, M. A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in psychology. Journal of the American society for information science and technology, 61/10: 2070–85.

Harzing, A. W. (2013). Document categories in the ISI Web of Knowledge: Misunderstanding the social sciences?. Scientometrics, 94(1), 23-34.

Moed, H. F. (2005). Citation analysis in research evaluation. Springer Science & Business Media. DOI: 10.1007/1-4020-3714-7.

Mokhnacheva, Y. V. (2023). Document Types Indexed in WoS and Scopus: Similarities, Differences, and Their Significance in the Analysis of Publication Activity. Scientific and Technical Information Processing, 50(1), 40-46.

Olensky, M., Schmidt, M., Van Eck, N. J. (2016). Evaluation of the citation matching algorithms of CWTS and iFQ in comparison to the Web of Science. Journal of the Association for Information Science and Technology, 67/10: 2550–64.

Owen, A. B. (2001). Empirical likelihood. CRC Press, Boca Raton, ISBN 978-1-4200-3615-2.

Sigogneau, A. (2000). An analysis of document types published in journals related to physics: Proceeding papers recorded in the Science Citation Index database. Scientometrics, 47(3), 589-604.

Valderrama-Zurián, J.-C., Aguilar-Moya, R., Melero-Fuentes, D., Aleixandre-Benavent, R. (2015). A systematic analysis of duplicate records in Scopus. Journal of Informetrics, 9/3: 570–6.

Yeung, A. W. K. (2019). Comparison between Scopus, Web of Science, PubMed and publishers for mislabelled review papers. Current Science, 116(11), 1909-1914.

Yeung, A. W. K. (2021). Document type assignment by Web of Science, Scopus, PubMed, and publishers to “Top 100” papers. Malaysian Journal of Library & Information Science, 26(3), 97-103.