A new method to extract n-Ary relation instances from scientific documents

Martin Lentschat, Patrice Buche, Juliette Dibie-Barthelemy, Mathieu Roche

Expert Systems With Applications 209 (2022) 118332

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Review

A new method to extract n-Ary relation instances from scientific documents☆

Martin Lentschat [a,b,*], Patrice Buche [a], Juliette Dibie-Barthelemy [c], Mathieu Roche [b,d]

[a] UMR IATE, Univ Montpellier, INRAE, Institut Agro, 2 place Pierre Viala, Montpellier 34060, France
[b] UMR TETIS, Univ. Montpellier, AgroParisTech, CIRAD, CNRS, INRAE, 500 Rue Jean François Breton, Montpellier 34090, France
[c] Univ. Paris-Saclay, INRAE, AgroParisTech, UMR MIA-Paris, 16, rue Claude Bernard, Paris 75231 CEDEX 05, France
[d] CIRAD, UMR TETIS, F-34398 Montpellier, France

ARTICLE INFO

Keywords:
N-ary relations
Natural language processing
Knowledge extraction
Information extraction
Ontological and terminological resource
Smart data

ABSTRACT

A new method to extract knowledge structured as n-Ary relations from scientific articles is presented. We designed and assessed different approaches to reconstruct instances of n-Ary relations extracted from scientific articles in experimental domains, driven by an Ontological and Terminological Resource (OTR) and based on multi-feature representation of relations and their arguments.

The proposed method starts with the identification of partial n-Ary relations in tables of scientific articles and then seeks to reconstruct them with argument instances in the article texts. Based on the so-called Scientific Publication Representation (SciPuRe) of textual arguments and Scientific Table Representation (STaRe) of n-Ary relations representation of an n-Ary relation called STaRe (Scientific Table Representation, originating from partial n-Ary relations extracted from document tables), here we propose and evaluate different approaches for the selection of textual argument instances that could complement partial n-Ary relations: structural, frequentist and word embedding models. The application domain concerns food packaging, especially composition and permeability data. Experiments were conducted on a corpus of 332 relation instances composed of 1547 arguments. Corpora of full and partial relations recognized in document tables and argument instances extracted from texts are available online. Different methods and strategies were measured with an f-score ranging from .34 to .74. These results show that n-Ary relations reconstruction approach depends on the number of selected candidate argument instances.

1. Introduction

Big data research involves approaches that address the management of large, varied, unstructured and potentially redundant data (De Mauro, Greco, & Grimaldi, 2016). Most work in this field is devoted to the extraction of major trends revealed by cross-checking the overall analyzed data. In contrast to these approaches, the smart data research field (Duong, Nguyen, & Jo, 2017; Marcia, 2017) focuses on the analysis of the context in which each data item appears, thereby transforming data into information and knowledge. The smart data approach provides a way to organize and highlight the semantics of various kinds of data by generating information on the data contexts and relevance. It facilitates the manipulation and analysis of the data at different scales, from single specific data to heterogeneous datasets. The smart data approach considers data individually and independently of the extracted dataset. This is why scientific knowledge capitalization and dissemination tasks come within the scope of smart data.

Our goal is to extract experimental knowledge from scientific publications. Automatic knowledge extraction from scientific documents can be done with different objectives: to use the extracted knowledge in decision making systems, for data comparison or to help design data-driven prediction models. Here we tested the approach in the food packaging domain with regard to the packaging permeability characteristics and experimental conditions. It is therefore important to recognize the packaging names, the extent of permeability to different gases, as well as the measurement method used and the control temperature during the experiments. We propose to structure this information in a Ontological and Terminological Resource (OTR) (Buche, Dibie-Barthélemy, Ibanescu, & Soler, 2013) in a format compatible with existing knowledge bases (e.g. @Web[1]).

We represent the knowledge embedded in scientific publications based on the n-Ary relation formalization. W3C (W3C Working Group, 2006) and the literature (Giunti, Sergioli, Vivanet, & Pinna, 2019)

☆ A new framework to reconstruct n-Ary relations from scientific documents.

* Corresponding author at: UMR IATE, Univ Montpellier, INRAE, Institut Agro, 2 place Pierre Viala, Montpellier 34060, France.
E-mail addresses: martin.lentschat@umontpellier.fr (M. Lentschat), patrice.buche@inrae.fr (P. Buche), juliette.dibie_barthelemy@agroparistech.fr (J. Dibie-Barthelemy), mathieu.roche@cirad.fr (M. Roche).

[1] https://www6.inrae.fr/cati-icat-atweb.

https://doi.org/10.1016/j.eswa.2022.118332
Received 7 March 2022; Received in revised form 11 July 2022; Accepted 30 July 2022
Available online 4 August 2022

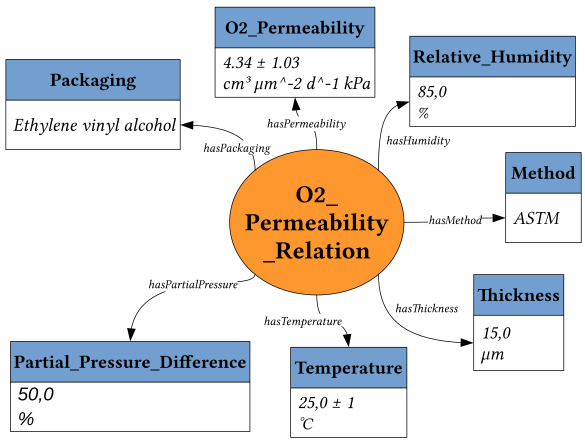


Fig. 1. Instance of n-Ary relation describing a permeability measure.

define an n-Ary relation as *"the natural way of relating one individual to strictly more than one other individual"*. The n-Ary relation term thus encompasses any relation involving at least three arguments. An instance of n-Ary relation (e.g. Fig. 1) consists of an instance of the relation concept and instances of arguments linked to it with binary relations. The relation shown in Fig. 1 formalizes a packaging permeability measure as an n-Ary relation: O2_Permeability_Relation. This relation concept is linked to the arguments that constitute its signature. The O2_Permeability_Relation signature thus contains the name of the packaging (Packaging), its permeability value with regard to a certain gas (O2_Permeability) and the measurement method (Method), along with the control parameters (Temperature, Relative_Humidity, Thickness and Partial_Pressure_Difference). Here, we assume that the n-Ary relation signatures have already been defined in an OTR structured as in Buche et al. (2013).

In order to automatically extract n-Ary relations from an article, the argument must be identified and grouped in n-Ary relations. The extraction process has to take into account the fact that relevant information concerning an experiment may be scattered throughout the documents (e.g. packaging characteristics are usually present in the Material and Methods section, while the permeability values are in Results and Discussion), and that several experiments may be present in the same document. Moreover, it is important to only recognize arguments concerning the n-Ary relations of interest (e.g. all temperatures in a document do not concern the permeability control parameters and results from other publications may also be presented in the article for comparison purposes). The conventional approach to deal with this problem (Zhou, Zhong, & He, 2014) involves: (1) the identification of arguments present in the entire document, and then (2) the reconstruction of the n-Ary relations.

Concerning (1), we previously focused on the identification, representation and relevance evaluation of arguments in scientific publications in experimental domains (Lentschat, Buche, Dibie-Barthelemy, & Roche, 2020; Lentschat, Buche, Dibie-Barthelemy and Roche, 2022), as summarized in Section 2. The method presented here addresses point (2), i.e. the reconstruction of n-Ary relation instances. Our main challenges concerned the multiplicity of relations present in single documents and the dispersion of arguments in various article sections. Indeed, each article in our corpus may present several permeability measurement results, many types of packaging and/or different control parameters in different document sections (e.g. Materials and Methods, Results and Discussion). This increases the number of arguments to extract and the extent of possible combinations. Another main challenge is the dispersion of arguments in one document expressed in different formats in a document.

Here we propose to take advantage of the information presented in document data tables to identify partial n-Ary relations. The extraction of partial n-Ary relations from document tables is based on the method outlined in Buche et al. (2013). The main original aspect of our approach is that it relies on data tables to guide n-Ary relation extraction, i.e. for a given partial n-Ary relation extracted from a table to be completed, all its already instantiated arguments are taken into account and compared to candidate arguments extracted from the text using a multi-criteria approach. We propose three ways to enhance state of the art methods to carry out such comparisons.

It is important to determine which criteria are relevant to complement partial relation instances extracted from tables with the arguments in the text in order to link scattered argument instances in an n-Ary relation and determine which argument instances to link. Section 3 covers the various criteria and their advantages in the n-Ary relation extraction task presented in the state-of-the-art. The method outlined in Section 4 exploits partial n-Ary relations that can be extracted from document tables and links argument instances from the text to reconstitute complete n-Ary relation instances. We addressed this issue by designing a new representation of n-Ary relation instances (STaRe - Scientific Table Representation), presented in 4.2, in the smart data context. Methods to reconstruct n-Ary relations are based on exploitation of the representation of relations (i.e. STaRe) and of arguments (i.e. SciPuRe). Their features are used in structural, frequentist and word embedding approaches. The representation outlines the structure of n-Ary relations (i.e. the arguments constituting each relation) and provides different types of features to describe the relations and their arguments. We also identified three multi-criteria approaches that take advantage of these features to select candidate arguments that may complete each partial n-Ary relation. We assessed these methods on a corpus of 332 relation instances (cf. Section 5), composed of 1547 argument instances in the food packaging domain.

2. Background

The conventional approach to extract n-Ary relations from text involves: (1) extracting argument instances in the text, and then (2) reconstructing the relation instances (Zhou et al., 2014). We previously developed a method that addresses step (1) issues (Lentschat et al., 2020; Lentschat, Buche, Dibie-Barthelemy et al., 2022), i.e. extracting and representing argument instances and proposing relevant scores to filter candidate arguments. Our method uses an Ontological and Terminological Resource (OTR) to define n-Ary relations (cf. Buche et al. (2013) for further details) and drive the extraction of argument instances. n-Ary relation and argument concepts are defined in the *up-core-ontology* of the OTR (e.g. Fig. 2). The *down-core-ontology* represents the main concepts used to describe knowledge in experimental domains, such as the symbolic concepts used for categorization purposes, quantity concepts and associated units for measurements. This core-ontology is generic and can be used in a broad range of experimental domains. The *domain-ontology* contains concepts related to the experimental domain at hand. The domain-ontology formalizes domain knowledge as n-Ary relations. These relations are represented as concepts and linked to their respective arguments, i.e. symbolic or quantitative. An example of an n-Ary relation describing a permeability relation is present in Fig. 1. Symbolic concepts in the food packaging domain are described in depth through more specific sub-concepts (e.g. packaging → composites → Nanocomposites). Quantity concepts are linked to their respective instances of measure unit concepts. Every concept of the domain-ontology is associated with a terminological component in the form of a set of *Labels*.

Example 1 illustrates how the terminological component of the OTR allows us to identify words and phrases in a text indicating the presence of an argument instance of the n-Ary relation instance in Fig. 1. For this, all labels related to an argument of an n-Ary relation are gathered, along with all of its sub-concept labels. This constitutes a vocabulary

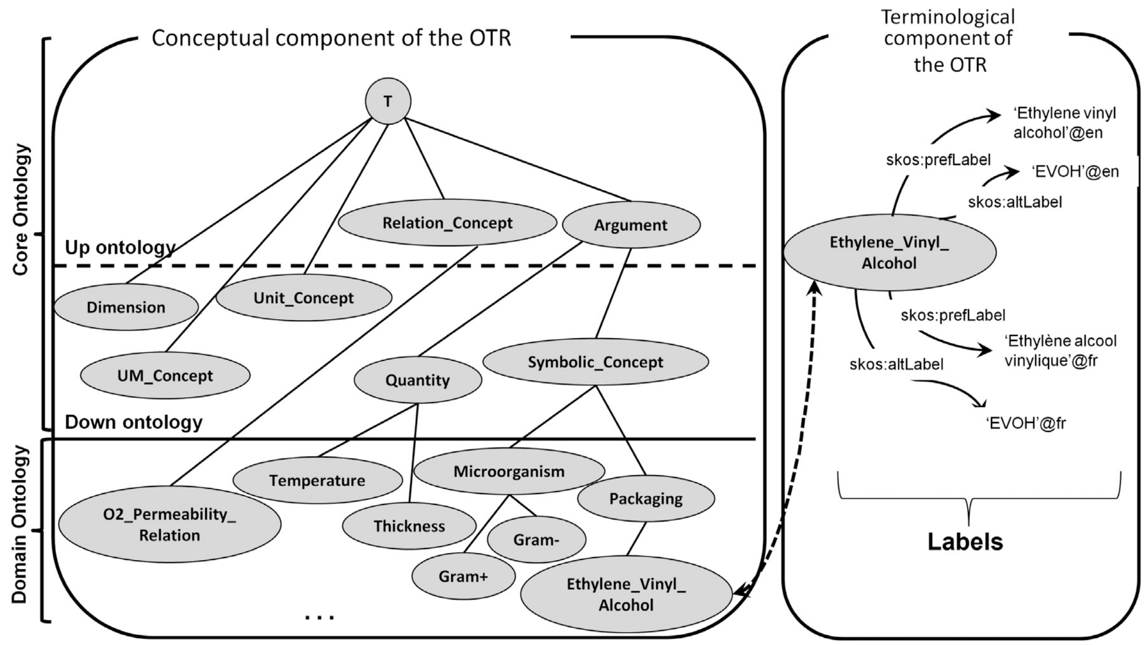


Fig. 2. An excerpt of the *TRANSMAT* OTR used in this study.

that will help identify instances of this argument. Symbolic arguments (e.g. packaging, method) are identified based on the presence of a label, as quantitative arguments (e.g. permeability measure, control parameters) use labels to disambiguate their unit of measure (e.g. by associating an ambiguous unit of measure like % with a label of the quantitative concept relative_humidity like *RH*).

Example 1. Identifying elements of argument instances in a text: *The permeability of* ethylene vinyl alcohol *films (*EVOH*) was measured with the* ASTM D95-96 *method at* 25 ± 1 °C*. The film had a* thickness *of* 15 µm *and showed optimal barrier properties with a* permeability to oxygen *of* $4.34 * 10^{-3}$ cm³ µm m^{-2} d^{-1} kPa*. This measurement was obtained at a constant* RH *of* 85.0 %*.*

Legend: ontological concept Label, measure unit, `numerical value`

Extracted argument instances are described in a multi-feature representation : SciPuRe (Scientific Publication Representation). Table 1 presents the SciPuRe of a quantitative argument instance with its ontological features: Target (the argument) and Node (the specific concept in the ontology), its Lexical Features: Original_Value (the text extracted as instance) and Attached_Value (the text used to categorize the argument), and its Structural features providing the appearance context of the argument instance in the document: Sentence, Window (the previous and next sentence), Segment (the article section and sub-section) and Document.

The argument instance extraction results showed that these instances were well retrieved (i.e. average recall of .85) but that lots of false positive instances were extracted (i.e. precision scores ranging from .08 to .56 depending on the argument). We sorted out some of the false positives using SciPuRe to compute and combine multi-feature scores to assess the relevance of each instance.

Our experiments, as detailed in Lentschat, Buche, Dibie-Barthelemy et al. (2022), showed that a semantic score based on the specificity expressed by ontological features of the argument's SciPuRe (i.e. CD^{node}_{target} in Fig. 6) was effective for measuring the relevance of symbolical arguments. A lexical score, based on the conventional *term frequency* measure (i.e. $TF^{term}_{document}$ in Fig. 6), was also useful for assessing the relevance of these arguments. Both scores can be sequentially combined, by first ranking the argument instances with the semantic score then re-ranking a subset of the top one with the lexical score. The proportion of re-ranked argument instances is expressed by θ in Fig. 6. This figure shows that precision of Packaging instances (i.e. precision of .37) could be improved by ranking the instances according to their relevance scores (R-precision of .63). Conversely, our experiments showed that quantitative instances were better sorted with a lexical score based on the discriminative power of sections in which the argument appeared.

Table 1
SciPuRe of a quantitative argument instances.

	Feature	Value
ONT.	Target Node	`Permeability` `O2_Permeability`
LEX.	Original_Value Attached_Value	['$4.34 * 10^{-3}$', 'cm³ µmm^{-2} d^{-1} kPa'] *'permeability', 'to', 'oxygen'*
STRUCTU.	Sentence Window Segment Document	*'The film had ... d^{-1} kPa'* [*'The permeability ... 25 ± 1 °C'*, *'The film ... d^{-1} kPa'*, Ø] *'Results and Discussion'* *Development of films based on quinoa starch*

3. State-of-the-art

The relation notion designates the connections (e.g. semantic, syntaxic) between different entities, especially in text. An n-Ary relation connects *n* arguments to a relation concept: it is *"the natural way of relating one individual to strictly more than one other individual"* by Giunti et al. (2019) or *"the natural and convenient way to represent certain concepts using relations to link an individual to more than just one individual or value"*[2] using the RDF and OWL languages proposed by W3C. Most approaches to extract n-Ary relations are based on Davidson's hypothesis (Davidson, 1980) that an n-Ary relation can be expressed as a sequence of binary relations between arguments. In this section we present some studies based on this hypothesis and their limitations. We then present our approach to the n-Ary relation reconstruction problem,

[2] https://www.w3.org/TR/swbp-n-aryRelations/.

based on a preliminary instantiation of partial n-Ary relations extracted from document tables.

The relation extraction process first consists of recognizing binary relations present in a sentence through a set of different techniques (Pawar, Palshikar, & Bhattacharyya, 2017), involving human linguistic expertise, external resources and/or learning processes. Binary relations between entity pairs are usually formalized as RDF triple (subject–predicate–object). In the classical paradigm, the relation extraction process relies heavily on the identification of entities (e.g. through named entity recognition) and the classification of relations between entities, by assigning semantic properties to the entities and their relations (Geng, Li, Han, & Zhang, 2022). Recent studies have addressed the relation extraction task with relation-focused methods. In order to avoid the usual two step extraction process (i.e. text entity recognition followed by the determination of the relation linking the entities), they consider the entities as being arguments of the relation. An approach based on the semantics of the relations is thus favored to be able to handle overlapping relations in a document (Takanobu, Zhang, Liu, & Huang, 2018) and minimize the risk of error propagation by simultaneously extracting relations and entities (Geng, Zhang, & Han, 2021).

The research branch focused on the extraction of information in a text and its formalization in triples is called Open Information Extraction (OIE) (Pawar et al., 2017). To extract triples present in a text sentence, OIE research relies mostly on syntactic analysis (Saha et al., 2018) and supervised machine learning (Stanovsky, Michael, Zettlemoyer, & Dagan, 2018). OIE also addresses specific challenges, such as the extraction of numbers and measure units (Saha, Pal, et al., 2017), the extraction of compound noun phrases (Pal et al., 2016) and semantic role labeling of recognized entities (Christensen, Soderland, & Etzioni, 2011). To explore how such systems could be used to meet our objective, we tested Open IE 5.1 (Mausam, 2016) and StanfordOpenIe systems (Manning et al., 2014) (cf. Section 5.3).

The relation extraction context has been extended by considering inter-sentence relations as well as coreferences, to deal with by the growing complexity of tasks to tackle (Grishman & Sundheim, 1996). The complexity of relations to be recognized has also increased by taking a larger number of arguments into account. Event extraction (i.e. the recognition of unstructured information in a text and their extraction into a structured event) is an example of multi-argument relations. Current research on this subject is carried out using statistical methods with two main approaches: a feature-based approach that searches for pre-determined features to extract or the use of neural networks to learn these features automatically (Yang, Feng, Qiao, Kan, & Li, 2019). However, despite commonalities between events and n-Ary relations (e.g. arity greater than two), event extraction is often limited to the context of a sentence or paragraph, while in our case n-Ary relations are extracted at the level of a several page document.

Relative to more complex relations, Davidson's hypothesis (Davidson, 1980) assumes that an n-Ary relation can be reduced to a sum of binary relations. In this framework, the classical (i.e. a binary relation that may be formalized as a triple) extraction of relations is based on different approaches such as syntactic analysis, under the hypothesis that a short dependency path is the main criterion (1) to determine the entities in a relation (Bunescu & Mooney, 2005), and (2) to determine the semantic nature of the relation (Chan & Roth, 2011). Additional criteria can be employed in conjunction with the use of syntactic dependencies, such as the search for frequent associations useful for designing extraction patterns (Greenwood & Stevenson, 2006) and machine learning techniques that largely draw on deep learning and pre-trained language models (Wang, Lu, Yin, & Qin, 2021). In order to extract complex (i.e. n-Ary) relations, such approaches have been used mainly in the biomedical domain (McDonald et al., 2005). An n-Ary relation with n arguments is thus decomposed into $n-1$ binary relations extracted using existing approaches and then reconstructed according to a relation framework. On the other hand, a literature review (Zhou et al., 2014) has shown that Davidson's hypothesis results in a combinatorial explosion, which has a negative impact on the computation time, as well as on the results. Zhou et al. (2014) experimentally estimated that a method with a p precision value in the extraction of binary relations would in turn have a p^{n-1} precision, where n is the number of arguments of the relation in the extraction of n-Ary relations according to Davidson's hypothesis and thus using binary relation extraction techniques.

Once the argument instances have been recognized in the documents, several criteria can then be used to identify relevant candidates for the relation. Linguistic knowledge and text analysis can help identify the links between arguments and can be used to create extraction rules. These rules consist of regular expressions or linguistic patterns based on different textual information levels. However, such rules need to be drawn up manually (Proux, Rechenmann, Julliard, et al., 2000), which requires a huge amount of human time at the expert level and might limit the application to specific cases. They can also be learned (Meng & Morioka, 2015) which requires a learning corpus that may not always exist in the studied domain. Recent promising research on the extraction of n-Ary relations on a document level are based on deep learning, especially LSTM networks (Wang et al., 2021) . This research mainly addresses issues related to the scattering and sparsity of argument instances in full-text document. To alleviate this, Akimoto, Hiraoka, Sadamasa, and Niepert (2019) decomposes cross-sentence ternary relations into unary and binary relations and trains a bidirectional LSTM to re-aggregate them. They obtained a mean average precision of .58 and .84 on Wikipedia and Freebase datasets. On a full-document level, Jia, Wong, and Poon (2019) extends this approach by decomposing n-Ary relations into binary relations based on their co-occurrence on a paragraph level. They then use a bi-LSTM to learn a multi-scale representation of the relation using different representation levels (i.e. paragraph, entity and document levels) and achieved an F1 score of .42 (recall of .43 and precision of .42) at a document level on the Clinical Knowledge Base. The main limitations of such approaches mainly concern the necessity of having access to a learning corpus and also the fact that they still address n-Ary relations of relatively low arity (i.e. 3 and 4 arguments) compared to ours (7 arguments). The constraint of a training corpus can be overcome by using a resource vocabulary and distant supervision (Mintz, Bills, Snow, & Jurafsky, 2009). Yet the latter is known to noticeably deteriorate the results, especially in specialized and complex domains (Quirk & Poon, 2017). Therefore, in this study we implemented techniques that do not require any kind of training.

Beyond machine learning, other criteria can be used to link instances of arguments in a relation. First, a simple assumption is that instances that co-occur frequently in the same context have a higher probability of belonging to the same relation. This can be used in the reconstruction of pre-defined relations, while argument instances to associate are pre-selected using models of the sought-after relations. Several measures are commonly used for this task using association, dependency and sample similarity measures (Lenca, Vaillant, Meyer, & Lallich, 2007). These take into account the occurrences of two argument instances (or more) compared to their number of co-occurrences in a determined context. This context is usually the sentence, but it could be of a different size, such as a set of sentences or a maximum distance between entities in a text. Existing measures do not reflect the same nature of associations, and the results obtained may differ depending on the chosen measure (Lenca et al., 2007).

Other approaches use semantic information to determine links between argument instances. Early studies (Ghersedine, Buche, Dibie-Barthélemy, Hernandez, & Kamel, 2012) used the properties of determined n-Ary relations to identify binary sub-relations in a local context (i.e. sentence) and then to recombine them. A 'pivot' argument instance is thus identified on the basis of its frequent co-occurrence with a result argument instance (i.e. the argument defining the relation). Ghersedine et al. (2012) found that the recombination of n-Ary relations based on binary sub-relations generated better results when done around a single argument, which here is called the pivot argument. This approach (Ghersedine et al., 2012) then seeks instances of arguments that have a binary relation with the pivot argument and then these binary relations are recombined in an n-Ary relation. However, solely relying on associations between argument instances and pivot argument instances limits the number of associations that could be detected, especially when there is a high number of arguments (i.e. the arity of the relation) and instances.

External resources (e.g. thesaurus, ontology, OTR) are often used to formalize the relations to be extracted and describe their arguments through vocabularies. This kind of resources can thus drive the argument instance extraction process and control the reconstruction of n-Ary relations by adding a semantic representation to the textual data. This semantic representation of text can be used to automatically design extraction patterns (Ramadier & Lafourcade, 2016) and be combined with lexical and syntactic patterns in a multi-criteria approach (Berrahou, Buche, Dibie, & Roche, 2017). Resources such as word-embedding models can also be useful for n-Ary relation extraction. These models can indicate syntactic or semantic relations between entities (Shahab, 2017) and thus provide criteria for linking argument instances. However, using a model learned on a domain and then applying it to another domain could result in performance loss, and transfer-learning issues are still a major concern (Peng, Yan, & Lu, 2019). An ontology or knowledge base is sometimes jointly used with a word-embedding model to fine-tune a pre-trained model (Yu et al., 2020) or generate good quality training examples (Yang et al., 2019).

The n-Ary relations we consider here have an arity of seven arguments, which is greater than that of usual n-Ary relations previously considered in the literature at a document level (Zhou et al., 2014), e.g. *drug–gene-variant* (Akimoto et al., 2019; Jia et al., 2019; Peng, Poon, Quirk, Toutanova, & Yih, 2017; Song, Zhang, Wang, & Gildea, 2018), three arguments. In addition to this high arity, the presence of different relation instances in a document increases the risk of confusion during the n-Ary relation reconstruction process by mechanically increasing the number of possible combinations.

The state of the art approaches presented above rely on the decomposition of n-Ary relations into a set of binary relations between arguments. Fig. 3 illustrates Davidson's hypothesis whereby these approaches decompose an n-Ary relation into a sequence of binary relations between arguments or substitute the relation concept for a pivot argument the rest of the argument instances are linked. This transforms an n-Ary relation, with n arguments, into $n-1$ binary relations ($1::1$ relations).

Positioning of our approach

In this paper, we propose a generic method that is applicable to different domains (depending on the OTR domain availability) while not requiring any training dataset. Secondly, the method is able to extract n-Ary relations with high arity (7 with the OTR used to assess the method) at a document level which has not been reported in the literature (i.e. mostly 3-ary relations).

We go beyond conventional binary relation extraction methods to take the specificities of n-Ary relations into account. This is achieved by simultaneously using the properties associated with all arguments of partial n-Ary relation instances extracted from the document tables, and the features associated with arguments in the text. This boosts the number of information sources exploited to reconstruct n-Ary relations through a multi-criteria approach, which is the aspect that we explore here.

The main original contribution of the method, as described in Section 4, is thus based on the search of *all* recognizable connections between an argument instance candidate extracted from a text and the argument instances belonging to an n-Ary relation partial instance extracted from a table.

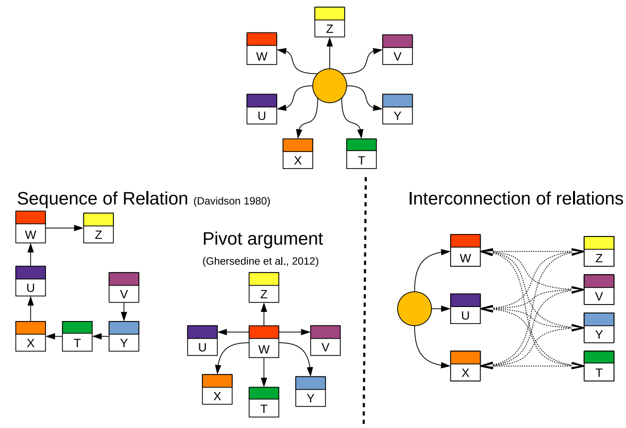


Fig. 3. Scheme of Davidson's hypothesis approaches compared to our method.

Our approach starts by identifying instances of partial n-Ary relations extracted from article tables through an automatic method using the table structure and an OTR (Buche et al., 2013). We then seek to supplement these partial n-Ary relations with argument instances previously extracted from the text. We select argument instances to complement a partial relation by looking for associations between each argument instance and all argument instances already in the relation.

Different multi-criteria methods are thus extended to determine the extent of interconnection between a partial n-ary relation instance and each candidate argument so as to select the best one. This is another original aspect of our approach, which extends existing approaches (structural, frequentist and word-embedding methods for n-Ary relation reconstruction) to take into account multi-feature representations of the relations and arguments, while combining ontological, lexical and structural features. Moreover, we studied the impact of filtering using relevance scores associated with textual arguments. The choice of extended methods is discussed depending on the number of candidate argument instances that must be assessed by an expert for final validation of n-Ary relation instances. A new corpus in the food packaging field has been specifically created to assess the method proposed in the paper and compare it to two state of the art methods.

Combined with our previous research on the identification of argument instances (Lentschat, Buche, Dibie-Barthelemy et al., 2022), this work constitutes a complete and innovative pipeline for the extraction of n-Ary relations driven by an OTR, while taking full text articles as input and generating a knowledge base enriched with n-Ary relation instances found in the articles as output.

4. Methodology

We aimed to extract knowledge from scientific publications and then formalize it in n-Ary relation instances. We thus extracted argument instances contained in the text, described in Section 2. However, the extent to which an argument instance belongs to a relation instance is not explicit in the text. Moreover, if some arguments are only associated with one instance of a relation (e.g. the permeability value), one argument instance could be linked with different relations (e.g. the same packaging in different permeability relations, i.e. to oxygen, carbon dioxide and dihydrogen monoxide) or with different instances of the same relation (e.g. the control temperature is the same in all oxygen permeability relations). The n-Ary relation reconstruction task is thus highly complex.

4.1. General approach to reconstruct n-Ary relations

In each article, argument instances are scattered throughout different sections and subsections. Scientific articles also include figures and tables that contain information. Argument instances in document tables can be linked into n-Ary relations via the table structure. The n-Ary relations formed by information in the tables are partial instances, because some instances of arguments defined in the relation scheme are often missing. We reconstruct complete n-Ary relations based on these partial relation instances from tables. We therefore supplemented this by looking for instances of arguments in the text.

Our work of reconstructing n-Ary relations involves different steps, as illustrated in Fig. 4, based on the SciPuRe representations associated with argument instances extracted from the text and new STaRe representations associated with partial n-Ary relations (cf. Section 4.2).

The three main steps for selecting an argument instance to add to a partial n-Ary relation are:

(A) **Candidates selection.** A set of candidate argument instances is constructed for each missing argument in the partial relation detected in a given table of a given document. The OTR contains the structure of n-Ary relations and each set is only composed of argument instances found in the same document. For example, if an Oxygen_Permeability_Relation partial relation from a table does not have an instance of the Temperature argument, then the candidate set consists of Temperature argument instances from the same document. In Fig. 4, a missing argument instance is shown in green in the partial relation. The set of selected instances is thus $\{D, E, H, E'\}$, i.e. candidates for reconstructing the relation.

(B) **Merging duplicates.** Similar instances are merged in the candidate set. Two argument instances are considered as duplicates if they have the same Original_Value feature values, for quantitative instances, or of Node feature values, for symbolic instances. For example, symbolic instances with Original_Value *'LDPE'* and *'low density polyethylene'* values refer to the same concept, since their Node values are identical: *'Low_Density_Polyethylene*.

(C) **Candidate discrimination.** A score is assigned to each candidate in order to determine the one to be added to the partial relation. This score is based on association indicators between the candidate and existing arguments in the partial relation. We adapted and extended different association indicators using structural, frequentist and word embedding approaches. These approaches use the SciPuRe and STaRe multi-feature representation.

This process is repeated for each missing argument in the partial relation. It should be noted that the order in which the candidate instances are added to the partial relations is not important here. Indeed, scores to discriminate candidates are computed by only considering instances of arguments in the original version of the partial relation, and not any instances of arguments that have been added during a previous iteration. This was done to minimize error propagation risks.

4.2. Representation of partial n-Ary relations from the tables

The n-Ary relations are reconstructed first by extracting the partial n-Ary relations in the document tables. The extraction of n-Ary relation instances is done using a semi-automatic method (Buche et al., 2013; Hignette, Buche, Dibie-Barthélemy, & Haemmerlé, 2009) driven by the same OTR and based mostly on similarity scores between data appearing in the cells and the ontological concepts. We designed STaRe (Scientific Table Representation), a representation associated with partial n-Ary relations found in a given table (cf. Table 2). This representation also factorizes SciPuRe features, i.e. the argument instance. STaRe contains the following features:

- **Ontological Features**: These features describe the relation according to OTR formalism. The *Relation* feature indicates the n-Ary relation belonging to the OTR. Arguments composing the n-Ary relation are described by two ontological features: Result_Argument, which indicates the result argument of the relation in the OTR. The Arguments feature gives the rest of the arguments composing the n-Ary relation. These argument instances are described by three features obtained via the extraction process:
  - Node, the specific concept associated with the argument in the OTR;
  - Original_Value, the cell content;
  - Attached_Value, the column header.
- **Structural Features**: This representation also provides information about the context of the table and its position in the document. The Table feature gives the title of the table and Caption indicates its caption. Segment describes the section and subsection from which the table was extracted. Document gives the title of the original article.

The STaRe representation indicates missing arguments in the relation instance (e.g. Temperature = $\emptyset$). The corpus of STaRe partial n-Ary relation instances semi-automatically extracted from the tables and used as Gold Standard is available online (Lentschat, Buche, Menut and Guari, 2021).

4.3. Methods for reconstructing n-Ary relations

Here we present two endogenous methods and a hybrid, exogenous/endogenous, method to discriminate candidate argument instances. These methods jointly rely on representations of the argument and n-Ary relations instances. This demonstrates how all of the different STaRe and SciPuRe features may be used to generate criteria to drive the reconstruction of n-Ary relation instances.

4.3.1. Structural method

The structural method for measuring the association of a candidate argument instance and argument instances of partial relations is based on scientific article structures (cf. Fig. 5). It has been widely shown that different sections of scientific articles contain different information (Cohen, Johnson, Verspoor, Roeder, & Hunter, 2010; Shah, Perez-Iratxeta, Bork, & Andrade, 2003). The intuitive aspect of this approach is that the best candidates for reconstructing partial relations are located in the same sections as the tables from which they are extracted. Moreover, candidate instances of certain types of arguments are more likely to be relevant in some sections than in others (e.g. the best candidates to reconstruct a Method argument would be in the *'Materials and Methods'* section).

The location of argument instances in the article structures is highlighted in the SciPuRe representation as well as in the STaRe representation based on the structural features. In each set of candidates, the structural association method searches for that closest to the table whose information must be supplemented. The distance is measured in number of tokens between the candidate and the title of the table from which the partial relation instance originates.

A variation of this method, i.e. the so-called guided structural method, is based on associations found between arguments and specific sections. Indeed, different sections in articles hold different types of information. This is illustrated, for example, in the automatic extraction of keywords (Shah et al., 2003), where the assessment of different sections may give different results, but also in the extraction of named entities in the medical domain, where the results may vary significantly according to the sections (Cohen et al., 2010). In our domain, control data will thus be more present in the *Materials and Methods* section while the measured permeability values will be in the *Results and*

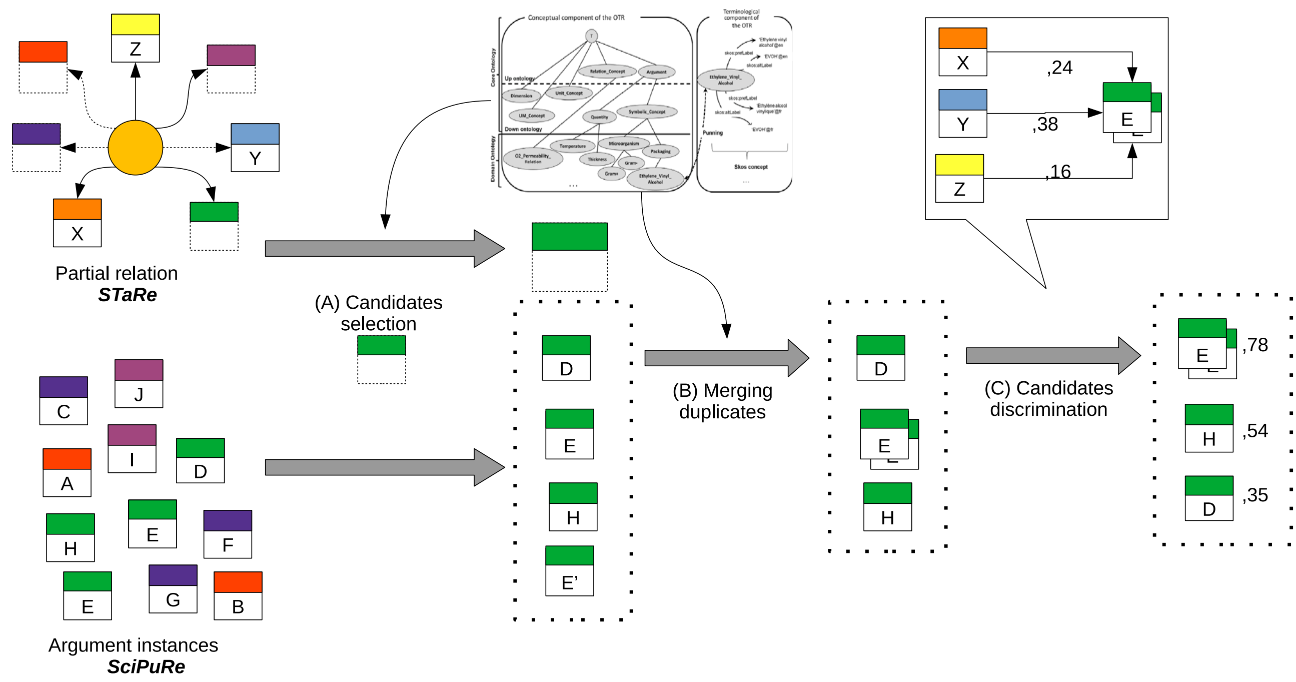


Fig. 4. Reconstruction process of n-Ary relations.

Table 2
Example of STaRe representation of a partial n-Ary relation instance.

	Feature	Value			
ONTOLOGICAL	Relation	H2O_Permeability_Relation			
	Result_Argument	*SciPure*			
		Target	Node	Original_Value	Attached_Value
		H2O_Perm.	H2O_Perm. cm^3 mm^{-2} s^{-1} bar	$1.27 * 10^2$	*Water Perm.*
	Arguments	Target	Node	Original_Value	Attached_Value
		Packaging	Chitosan	*Chitosan films*	*Chitosan films*
		Method	∅	∅	∅
		R_H	∅	∅	∅
		Temperature	Temperature	25 °C	*Temp.* (°C)
		Thickness	∅	∅	∅
STRUCTURAL	Table	Table 3			
	Caption	*Water permeability of tested packaging at* 25 °C			
	Segment	*Results and Discussion*			
	Document	*Barrier properties of chitosan coated polyethylene*			
	DOI	https://doi.org/10.1016/j.memsci.2012.02.037			

Discussion. Specific arguments in sections may be targeted by manually defining priorities relative to an argument-section association table. This table, an example of which is presented in Table 3, indicates sections of scientific articles that should be examined first when searching for specific argument instances. The guided structural method searches among candidates for those whose SciPuRe Segment feature has the highest priority according to the order indicated in Table 3. The priorities were collaboratively determined by the three annotators of the Gold Standard argument instances (Lentschat, Buche, Dibie-Barthelemy, Menut and Roche, 2021; Lentschat, Buche and Menut, 2021). The previously mentioned criteria of the structural method then applies if more than one candidate falls within the argument priority section.

Table 3
Argument search priorities by Segments for the guided structural method.

Argument	Segment				
	Abstract	Intro.	Mat. & Meth.	Res. & Dis.	Conc.
Permeability	2	4		1	3
Packaging	3	1	2		4
Method	3	2	1	4	
Relative_Humidity	3	2	1		4
Temperature	3	2	1		4
Thickness	3	2	1		4

4.3.2. Frequentist method

The frequentist method is based on measures of associations between a candidate, and its duplicates, with partial relation argument manifestations (cf. Fig. 6). An argument manifestation is the reference in the text of an argument instance in a table (e.g. the permeability value in the table is also commented in the text). These manifestations are established by comparing the value representation of the argument instance in a partial relation and argument instances extracted from the text. First, the Argument and Document features are compared to highlight the manifestation of a relation argument in the text. Hereafter we define a manifestation as direct or indirect:

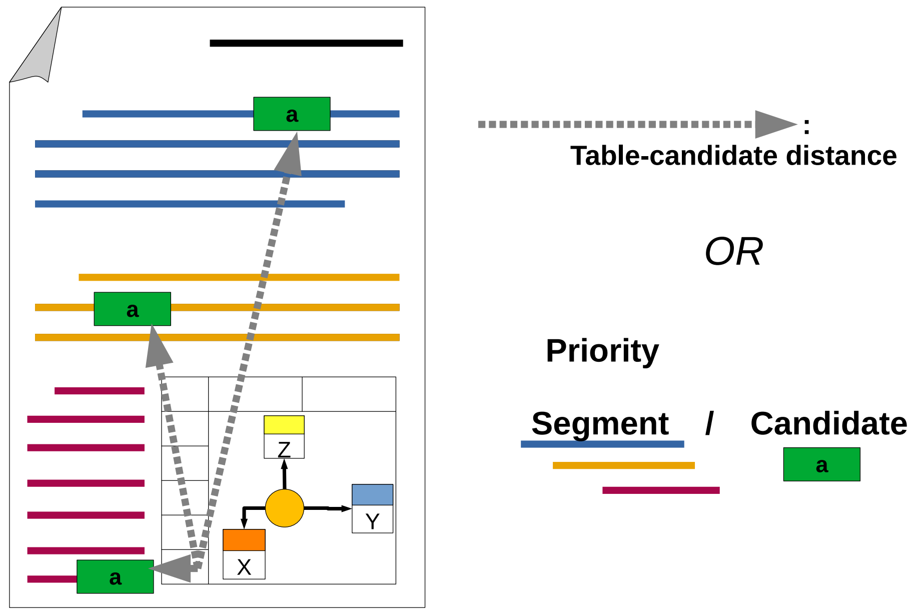


Fig. 5. Structural method.

- a manifestation is considered as *direct* when the Original_Value features values match (e.g. same packaging name, same permeability value).
- a manifestation is considered as *indirect* when the values of Node features for symbolic arguments, or Attached_Value features for quantitative arguments, match (e.g. same OTR packaging concept, same quantity term).

Co-occurrences of candidate instances and manifestations of partial relation arguments are used in the frequentist method to measure associations between an argument candidate and arguments present in the n-Ary relation. Co-occurrences with different manifestation types and different contexts may be considered. Different scores exist to measure entity co-occurrences. We opted to assess three of them: Dice (Dice, 1945), Jaccard (Jaccard, 1901) and Point-wise Mutual Information (PMI) (Church & Hanks, 1990). These measures were chosen for their simplicity and on the basis of the fact that they are commonly used in information extraction (Role & Nadif, 2011; Ru, Tang, Li, Xie, & Wang, 2018). The SciPuRe structural features indicate the different contexts in which two instances may co-occur. The lexical features of an event define its type, i.e. direct or indirect.

4.3.3. Word embedding method

Word embedding refers to learning methods geared towards representing terms of a corpus in real number vectors (Mikolov, Chen, Corrado and Dean, 2013; Mikolov, Sutskever, Chen, Corrado and Dean, 2013). These techniques are commonly used in the natural language processing field to create vector language models. These word-embedding models are learned by deep neural networks on large corpora commonly following a CBOW (continuous bag of words), Skip-Gram (distributed) or, more recently, Transformers approach. The resulting word-embedding models represent the terms of a domain through high dimensional vectors. For example, spaCy and sciSpacy models (Neumann, King, Beltagy, & Ammar, 2019) use 300 dimension vectors and the BERT model (Devlin, Chang, Lee, & Toutanova, 2019) uses vectors with more than 750 dimensions. These models enable computation of semantic similarity scores between two terms by calculating the similarity between vectors of these two terms, commonly through calculation of the cosine of their vector angle. These similarity values can be used as association scores between candidate terms and arguments in a partial relation.

Pre-trained models exist and can be used directly. Most of them are trained on corpora constituted by text derived from media sites, blogs or literature corpora. Some of them are obtained from scientific documents (e.g. medical articles and reports, chemistry and biology publications) (Khattak et al., 2019). The sources chosen to train a word-embedding model have a major impact on the model performance when it is applied to different documents (Peng et al., 2019). Domain adaptation is a widely discussed issue. The fact that a corpus has a specific vocabulary or syntactic forms indeed has an impact on the model, and reusing it in a different domain may then be difficult (Peng et al., 2019; Shahab, 2017).

Models concerning a specific experimental domain sometimes do not exist. This is due to the scarcity of sources, i.e. fewer articles are generally published in these domains than, for example, in the biomedical domain. The models closest to the domains we study are trained on corpora of scientific articles mostly concerning the medical, biological and chemistry fields. We assessed the impact of transferring a model to our domain using eight different models. They are all trained on different English-language corpora from the scientific domain or not (see Table 4). We used the spaCy[3] (v3.0) library because of its high performance and simplicity of use. SpaCy models or models tailored for this library were used.

For all the models used, we applied the same method to determine the semantic similarity scores between candidates and argument instances present in a partial relation (cf. Fig. 7). This score helps identify the best candidates to reconstruct partial relations. For each candidate, we consider the set of terms, i.e. its SciPuRe lexical features Original_Value and Attached_Value, of each of the instances of its duplicates. These lexical features are also retrieved from argument instances in the partial relation. Similarity values between each candidate term and each argument instance term present in the partial relation are computed (using the spaCy *similarity()* function). This function computes the similarity score corresponding to the cosine between two vectors of terms (word or phrase). A value is thus obtained for each possible combination of terms of candidate instances and relation arguments. Terms that do not have a vector in the language model used, such as numerical values, complex measure units or specialized terms are ignored. The arithmetic mean of similarity scores provides the association score between a relation and a candidate argument.

[3] https://spaCy.io/.

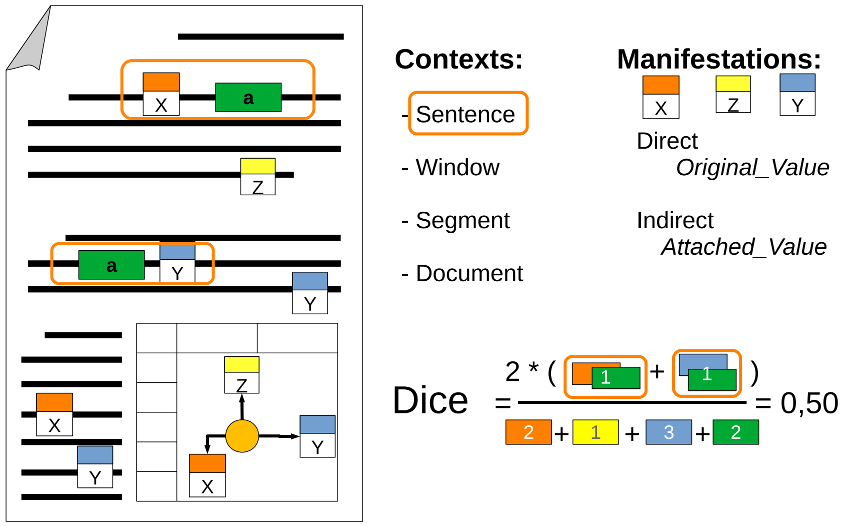


Fig. 6. Example of a co-occurrence measure (Dice).

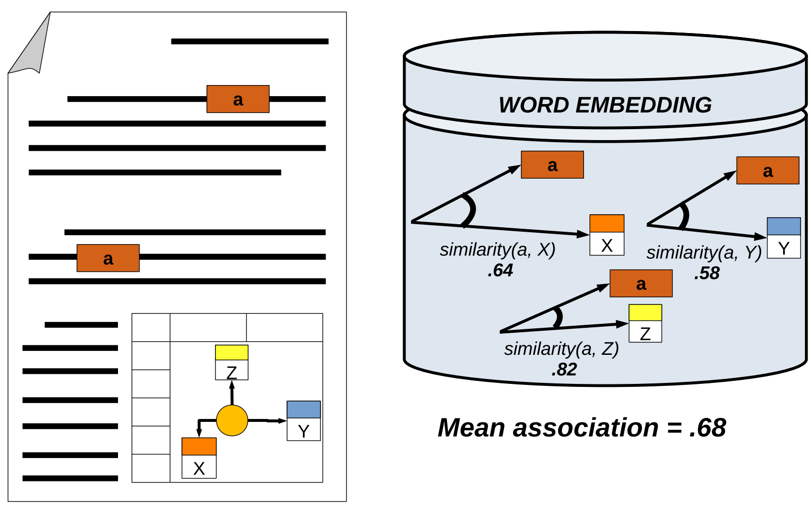


Fig. 7. Semantic similarity computation between a candidate and an argument present in a partial relation.

4.4. Candidate filtering and selection

In addition to methods for reconstructing n-Ary relation instances, we identified two supplementary criteria:

- the **filtering** of candidate argument instances as a prerequisite to the relation reconstruction process. We have previously shown that a significant number of false-positives were present in argument instances extracted from the text, but SciPuRe features could be used to construct relevance scores that may be useful for discarding some of these false-positives (Lentschat et al., 2020; Lentschat, Buche, Dibie-Barthelemy et al., 2022). Filtering candidate argument instances with the lowest relevance scores should thus eliminate some candidates that are not argument instances of n-Ary relations. This filtering is done before starting the search for argument instances to be added to relations via structural, frequentist or word embedding methods. In Section 5, we assess the effects of filtering 0%, 20%, 40% and 60% of the candidate argument instances with the lowest relevance scores.
- the **selection** of multiple candidate instances of a relation instance. Relation instance reconstruction methods provide measures that order candidate instances that can be used to reconstruct partial n-Ary relations, and the candidate with the best score is selected. However, it is also possible to select a set of several candidates. This approach can assist and support experts in selecting information to retain for each relation. The effects of selecting one, three, five, or ten candidate instances per argument to reconstruct a partial relation are assessed in Section 5.

5. Results and discussion

5.1. Results

The methods for reconstructing n-Ary relations are evaluated in this section. These methods are based on relation instances extracted from tables and searches potentially complementary for argument instances using structural, frequentist and word embedding criteria. We assessed the impact of these relation reconstruction methods by comparing the obtained relation instances with relation instances in a Gold Standard (Lentschat, 2022) . It was manually constructed by associating a Gold Standard of argument instances from text (Lentschat, Buche, Dibie-Barthelemy et al., 2021; Lentschat, Buche, Menut, 2021) and a Gold Standard of partial relation instances from tables (Lentschat, Buche, Menut, Guari, 2021; Lentschat, Buche, Menut, Guari and Roche, 2022). This resulted in a Gold Standard of 332 relation instances describing knowledge on oxygen, carbon dioxide and water vapor permeability measures, while also including packaging composition

Table 4
Word-embedding models used.

Model	Sources
en_core_sci_lg	GENIA[a] (biomedical), Pubmed Central Open Access Subset[b] (medical), The MedMentions Entity Linking dataset[c] (medical), Ontonote[d] (blogs, news-sites, comments)
en_core_sci_scibert	SciBERT model[e] (scientific articles[f])
en_ner_craft_md	CRAFT[g] (biomedical, chemistry)
en_ner_jnlpba_md	JNLPBA[h] (biomedical)
en_ner_bc5cdr_md	BC5CDR[i] (biomedical)
en_ner_bionlp13cg_md	BIONLP13CG[j] (biology)
en_core_web_lg	Ontonote[d] (blogs, news-sites, comments)
en_core_web_trf	modèle RoBERTa[k] (Wikipedia, literature)

[a] https://nlp.stanford.edu/∼mcclosky/biomedical.html.
[b] https://evexdb.org/pmresources/vec-space-models/.
[c] https://github.com/chanzuckerberg/MedMentions.
[d] https://catalog.ldc.upenn.edu/LDC2013T19.
[e] https://github.com/allenai/scibert.
[f] https://semanticscholar.org/.
[g] https://bionlp-corpora.sourceforge.net/CRAFT/.
[h] https://github.com/spyysalo/jnlpba.
[i] https://biocreative.bioinformatics.udel.edu/tasks/biocreative-v/track-3-cdr/.
[j] https://2013.bionlp-st.org/.
[k] https://github.com/pytorch/fairseq/tree/master/examples/roberta.

relations, with 1547 argument instances, in 10 documents. By the smart data approach, this corpus size is generally reduced due to the scarcity of articles addressing the expert specifications and the difficulty of manual annotation in these specialized domains. As an example, Brack, D'Souza, Hoppe, Auer, and Ewerth (2020) identified an average of 528 scientific entities in 10 different specialized domains with corpora of 11 documents. The corpora used here required ≈80 man-hours for cross-annotation of the argument instances in the text by three annotators (Lentschat, Buche, Dibie-Barthelemy et al., 2021; Lentschat, Buche, Menut, 2021), ≈60 man-hours for the annotation of tables containing the partial n-Ary relations (Lentschat, Buche, Menut, Guari, 2021) by one annotator and their validation by three annotators and ≈15 man-hours for one annotator to rebuild the complete n-Ary relations (Lentschat, 2022) based on the two corpora. This manual annotation work is important, since automatic annotation processes such as those used in distant supervision (Mintz et al., 2009) would fail to only select relevant information in the text.

The evaluation is based on recall, precision and f-score (micro) values.

$$Precision = \frac{true\ positives}{true\ positives + false\ positives} \tag{1}$$

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives} \tag{2}$$

$$f-score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3}$$

In the assessment of partial relation instance completion approaches, we also considered two points: the filtering of candidate argument instances prior to relation instance completion and the selection of multiple candidates for a partial relation instance argument to be reconstructed. These two criteria allowed us to filter out some false positives present in the candidate argument instances and to position our approach as an expert assistance process.

5.1.1. Structural method

In documents, the structural method uses the proximity between candidate argument instances and the table containing the partial relation to be reconstructed, and searches for the closest ones. A variant, i.e. the guided structural method, uses manually determined associations between document sections and types of arguments to prioritize candidates within certain document sections.

Table 5
Evaluation of the structural method.

Method	Recall	Precision	f-score
Simple	.97	.20	.33
Guided	.98	.25	.40

The results presented in Table 5 show the findings of the evaluations of the structural method presented in Section 4.3.1, with and without guidance. The results here were first measured by selecting only one candidate per missing argument instance without prior filtering of argument instances according to their relevance scores. We obtained high recall values (.97 and .98), with a slight difference noted in the precision values of the two versions of the method, i.e. in favor of the guided structural method (.20 vs .25).

Fig. 8 highlights the variation in the recall scores of the structural methods obtained by changing the percentage of filtered candidate argument instances (different column colors) and the number of candidates selected to reconstruct the partial relations (different histogram bundles). Filtering part of the candidates led to a decrease in recall proportional to the extent of filtering, which meant that some valid candidates were eliminated by this filtering. On the other hand, increasing the number of selected candidates reduced this impact, particularly when there was a high proportion of filtering. These behaviors were noted for both structural method approaches.

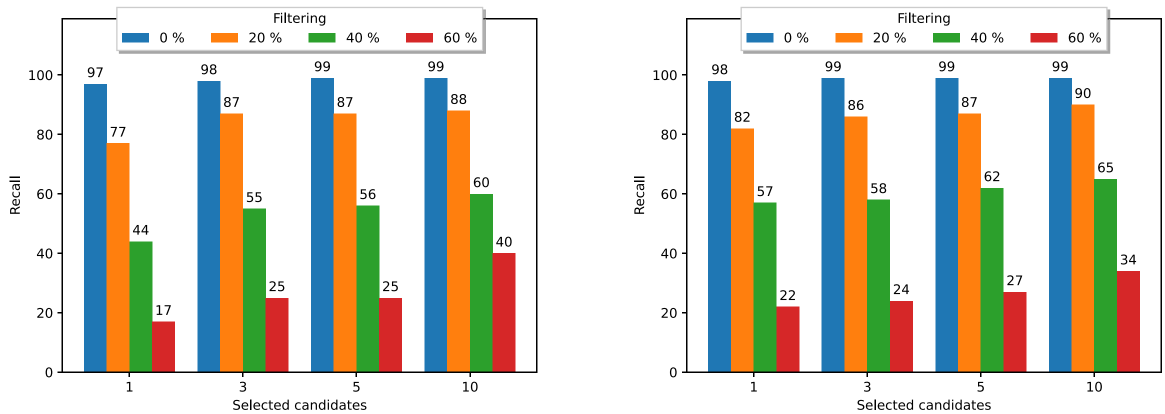
Fig. 9 highlights variations in the precision scores of the structural methods. There was a natural increase in precision with the number of candidates selected per missing argument (i.e. there was a higher probability of selecting a valid candidate as the number of selected candidates increased). This precision increased when three candidates were selected instead of one (e.g. .20 → .36 and .25 → .38) and it increased further when five or ten candidates were selected. This increase was more marked for the guided structural method. These findings tend to indicate that guiding the candidate search in specific sections is relevant, but multiple candidates must be selected in that section. Pre-filtering of irrelevant candidate instances had a positive effect on the accuracy only when it concerned a small proportion of the candidates (i.e. 20%). Beyond that the accuracy decreased, indicating that relevant candidates were discarded in the filtering process.

The recall scores obtained with the structural method were excellent, i.e. around .98, and they only decreased when there was substantial pre-filtering of candidate instances. Conversely, the number of selected candidates per missing argument had to be increased to 5 or 10 to be able to obtain a precision higher than .50. Applying slight candidates filtering (i.e. 20%) also had a slightly favorable effect on the precision without markedly impacting the recall.

5.1.2. Frequentist method

The frequentist method presented in Section 4.3.2 measures associations between a candidate and textual manifestations of argument instances of a partial relation based on the extent of their co-occurrences. Three measures were tested for this purpose: Jaccard, Dice and Point-Wise Mutual Information (PMI). We also considered several co-occurrence contexts based on the structural features of SciPuRe: Sentence, Window, Segment and Document. Different manifestations of partial relation arguments were also considered via the lexical features Original_Value and Attached_Value. As there were many possible configurations, Table 6 shows the highest and lowest f-scores of each measure.

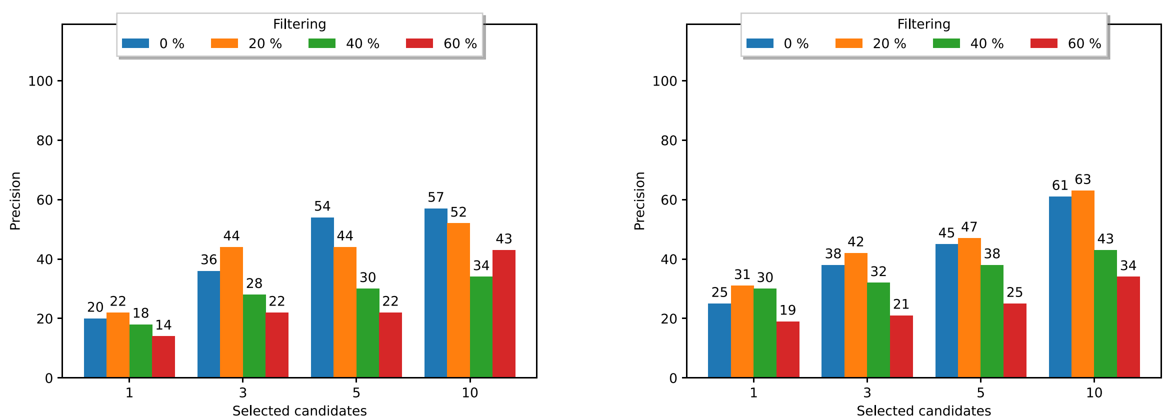
There was little difference between the three tested measures, as shown in Table 6. Although Dice systematically generated higher values than Jaccard and PMI, these differences were not significant (e.g. recall values of .71, .69 and .68, respectively). Yet the Document was clearly the most favorable context for measuring the association between a candidate and manifestations of arguments in a partial relation (e.g. the best f-score was $Dice(Document, AttachedValue)$ =.40) whereas the



(a) Structural method

(b) Guided structural method

Fig. 8. Effects of the filtering and selection of different numbers of candidates on the recall.



(a) Structural method

(b) Guided structural method

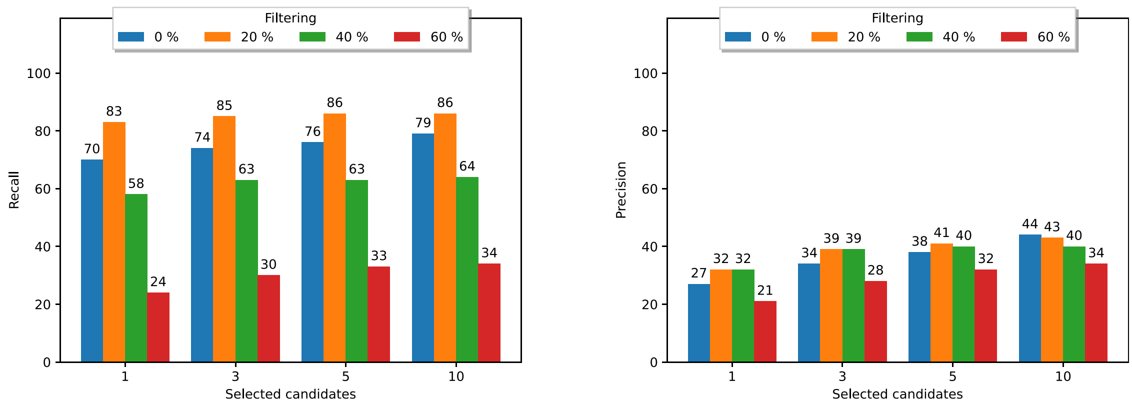
Fig. 9. Effects of the filtering and selection of different numbers of candidates on the precision.

Segment was the least favorable context. Intermediate results were obtained when considering Sentence or Window as context. This means that the overall frequency (i.e. at the Document level) of an argument instance (i.e. candidates or manifestations) was the most likely indicator to detect associations between candidates and relations. The Segment seemed to be the least favorable context due to the presence of different candidate instances of a same argument in a section (e.g. most of the different Relative_Humidity argument instances are presented in the Materials and Methods section) and one candidate was assigned to the different relations, while the others were left out. Our assessment showed that considering the indirect manifestations of the instances of the arguments of partial relations gave the best results. Indirect manifestations, depending on the Attached_Value feature, generally highlight the concept underlying the argument (e.g. *'temperature'*, *'thickness'*) and are therefore relevant for consideration in broad co-occurrence contexts such as Document. Direct manifestations, Original_Value, are more generally associated with reduced contexts, i.e. Sentence or Window, as they allow the association of more specific instances. The best frequentist method scores remained low. A recall of around .70 could be considered acceptable when adopting an expert assistance approach, whereas a maximum precision of .28 is insufficient.

Table 6
Recall, precision and f-score of the frequentist methods.

Measure	Context	Manifestation	Recall	Precision	f-score
Dice	Document	Attached_Value	**.71**	**.28**	**.40**
Jaccard	Document	Attached_Value	.69	.27	.39
PMI	Document	Original_Value	.68	.25	.36
...					
Dice	Segment	Attached_Value	.52	.13	.20
Jaccard	Segment	Attached_Value	.50	.12	.19
PMI	Segment	Attached_Value	.50	.12	.19

Fig. 10 presents variations in recall and precision values obtained through a frequentist method when using a Dice measure at the document level and indirect manifestation, with variations in the percentage of filtered instances and in number of candidates selected to reconstruct the partial relations. Increasing the number of selected candidates produced a slight increase in recall. Filtering part of the candidates increased the recall when this filtering was set at 20%, beyond which the recall decreased. This means that some valid candidates are eliminated by this filtering process as their relevance scores are too low. This suggests that light candidate filtering, i.e. 20%, has a better effect on the recall of frequentist methods than increasing the number of selected candidates. There is a slight increase in the precision according to the number of candidates selected for each missing argument. Pre-filtering of irrelevant candidate instances mainly has a positive effect on the precision when it is just focused on a small proportion of candidates (i.e. 20% and 40%). Fig. 10 also shows that filtering and selection of several candidates can interfere. Indeed, the precision was found to



(a) Recall (b) Precision

Fig. 10. Effets of filtering and number selection of candidates in the frequentist method - $Dice^{Document}_{Attached\_Value}$.

decrease when the filtering was increased to select ten candidates. This clearly shows that too much filtering excludes relevant candidates from the matching process.

Overall, an acceptable recall (i.e. ≥.80) can be obtained with slight candidate filtering. The selection of several candidates has little effect when using frequentist methods. On the other hand, precision is a problem with these methods, i.e. the precision did not exceed .50 with any of the parameters we tested.

5.1.3. Word-embedding method

The word-embedding method presented in Section 4.3.3 uses similarity scores, as computed via word-embedding language models, between terms of argument instances of a partial relation and terms of a candidate argument instance. The set of similarity scores between these terms is reduced to a single value by calculating the arithmetic mean.

The results presented in Table 7 show the recall, precision and f-score for each of the tested language models. We first observed that the majority of the evaluated models generated equivalent recall and precision scores, i.e. around .50 recall and .12 precision. It should be noted that the f-score obtained with the core_web_trf and core_sci_scibert models were nearly .15 higher than the scores obtained with the other models. Both models are based on BERT Transformers, while core_web_trf is trained on web sources and core_sci_scibert on scientific articles from various domains (see Table 4). This difference highlights the quality of the BERT Transformers approach compared to other models based on Skip-Gram and CBOW approaches. Otherwise, there were no significative differences between these two models, although core_sci_scibert is trained on scientific sources. This may be due to the absence of specific articles in the targeted domain in the training corpus of core_sci_scibert. Indeed, it has been shown that learning transfer can only be efficient in very closely related domains (Peng et al., 2019). This could also explain the absence of significant variations in the scores of other models trained on various domains, none of which are sufficiently close to our application domain. The superior performance of models using the BERT method, combined with the lack of differences across training domains, suggests that the general language knowledge gained via the word-embedding model is more important than the domains proximity. This has also been noted in studies comparing the performance of word-embedding models in different domains (Peng et al., 2019; Shahab, 2017; Wang et al., 2018).

We assessed the effects of filtering and the number of selected candidates on the core_web_trf model. Fig. 11 shows variations in the recall scores according to variations in the percentage of filtered instances and the number of candidates selected to reconstruct the partial relations. Increasing the number of selected candidates was found to increase the recall. However, this increase mainly occurred when three candidates were selected, while selecting five or ten did not offer any significant improvement. Filtering part of the candidates increased the recall when this filtering was set at 20%, beyond which the recall decreased. Some valid candidates were thus eliminated by this filtering as their relevance scores were too low. These behaviors were noted for both models. A slight increase (to three) in the number of selected candidates combined with light filtering (20%) of candidates with the lowest relevance scores allowed us to obtain recall scores exceeding .80. Precision scores also increased mechanically with the number of candidates selected. This gain was close to .20 points when the candidate number increased from one to three, with a further increase of about .10 points achieved when there were five and ten candidates. Pre-filtering of irrelevant candidate instances had a positive effect on the precision mainly when performed on a small proportion of candidates (i.e. 20% or 40%). This effect was clearly visible, but never represented a gain of more than .05 points in precision. With more filtering, the accuracy decreased, indicating that some relevant candidates were eliminated by the filtering.

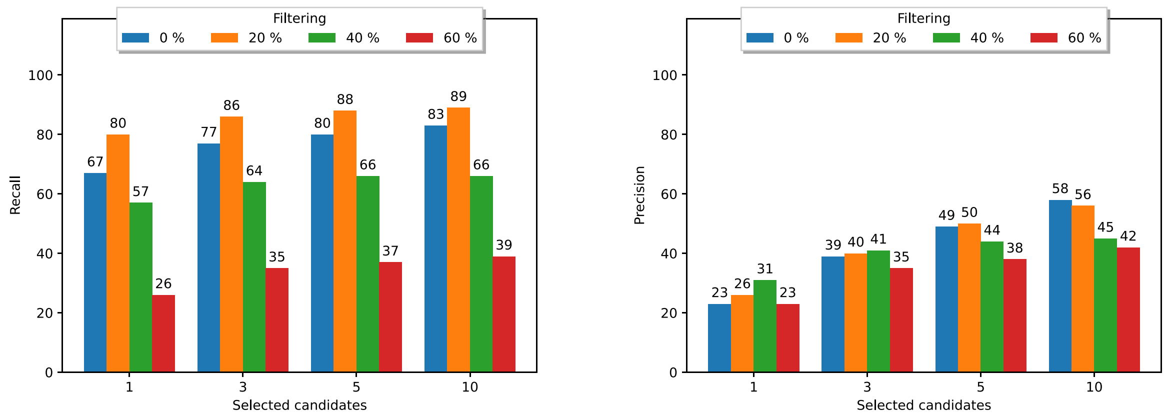
Table 7
Recall, precision and f-score of the word embedding methods.

Model	Recall	Precision	f-score
ner_jnlpba_md	.53	.13	.21
ner_craft_md	.52	.13	.21
ner_bionlp13cg_md	.52	.13	.21
ner_bc5cdr_md	.54	.14	.22
core_web_trf	**.67**	**.23**	**.35**
core_web_lg	.52	.13	.20
core_sci_scibert	**.65**	**.22**	**.33**
core_sci_lg	.52	.13	.20

The word embedding methods approach generated overall low recall and precision scores, while the core_web_trf and core_sci_scibert models, based on the BERT approach, were found to produce the best scores among all the evaluated models. The recall may be boosted to an acceptable level (i.e. ≥.80) via light filtering of the argument instances according to their relevance scores and by selecting three candidates per argument to reconstruct. On the other hand, the number of selected candidates for each missing argument must be increased to 5 or 10 to be able to achieve precision of over .50.

5.2. Comparison of methods

Table 8 should enable operators to choose the best approach to adopt according to the number of selected candidates to be proposed to experts for n-Ary relation extractions. When the aim is to provide results that do not require human intervention, frequentist approaches



(a) Recall (b) Precision

Fig. 11. Effets of filtering and number selection of candidates in the word embedding method—core_web_trf.

Table 8
Best f-score values of the tested methods (filtering = 20%)

Approach	Criteria	f-score selected candidates			
		1	3	5	10
Structural	simple	.35	.58	.58	.65
Structural	guided	.45	.56	.61	**.74**
Frequentist[a]	Jaccard	**.48**$^{d}_{a}$	.54$^{d}_{a}$	.61$^{p}_{o}$	.66$^{p}_{o}$
Frequentist[a]	Dice	.46$^{d}_{a}$	.55$^{d}_{o}$	.60$^{p}_{o}$	.66$^{p}_{o}$
Frequentist[a]	PMI	.44$^{d}_{o}$	.53$^{d}_{a}$	.60$^{p}_{o}$	.68$^{p}_{o}$
Word embedding	core_web_trf	.40	**.59**	.64	.70
Word embedding	core_sci_scibert	.39	.57	**.65**	.70

[a]Contexts: $p = Sentence,\ w = Window,\ s = Segment,\ d = Document$ manifestations: $o = Original\_Value,\ a = Attached\_Value$.

based on co-occurrences of candidate instances with argument manifestations of partial relations at the Document level provide the best results ($Jaccard = .48, Dice = .46$). When an expert is available to assist in picking from three or five argument instance candidates to reconstruct a partial n-Ary relation, word embedding appears to be the best method, with f-scores of .59 and .57 or .64 and .65 obtained for the two BERT based models. When expert intervention is necessary to sort among ten candidates, the guided structural method is the most relevant, with an f-score of .74.

In the following sub-section, we discuss our results compared to the state of the art in n-Ary relation extraction. We conducted a set of additional experiments using two available Open Information Extraction techniques in order to determine if these existing methods would be able to extract triples contained in our n-Ary relation instances. We also carried out two sets of OTR ablation experiments to assess the dependency of our method on the semantic resource.

5.3. Discussion

Comparison to the state of the art. Our experiments on automatic extraction of knowledge in scientific articles, and its formalization as n-Ary relations, were conducted under the smart data paradigm. This led to specific difficulties: since the knowledge targeted for extraction are unique (i.e. application domain, n-Ary relations arity), while adapting existing methods is hard and requires generic approaches that can only be partially compared to other knowledge or domains. Concerning the use of existing methods, hereafter we discus the results of state of the art techniques. We also conducted complementary experiments to determine if open information extraction models could be used in zero-shot information extraction. Finally, we measured the dependence of our approaches to the semantic resource through ablation experiments, where some of the ontology concepts were deleted beforehand.

The comparison of our approach to state of the art techniques is limited by the fact that deep learning is mainly used to extract n-Ary relations in documents, which was not a possibility in our study due to the lack of training corpus. Moreover, as mentioned in Section 3, most research on n-Ary relation extraction in scientific documents has been conducted in the bio-medical domain on 3-Ary (Akimoto et al., 2019; Jia et al., 2019; Peng et al., 2017; Song et al., 2018) relations (e.g. drug–gene-variant). Our study was aimed at extracting 7-Ary relations representing knowledge on food packaging permeability. The higher arity of the relation would surely result in lower results, as observed by Zhou et al. (2014). The results of reported state of the art studies have been quite variable when using accuracy measure (e.g. .71 for Song et al., 2018), mean average precision (e.g. .58 and .84 for Akimoto et al., 2019) or precision–recall (e.g. $p = .43, r = .30 F1 = .35$ for Peng et al., 2017 and $p = .42, r = .43, F1 = .42$ for Jia et al., 2019). The evaluation scores we obtained are in the range of those of Jia et al. (2019) and Peng et al. (2017), with F1 scores ranging from .35 to .48. The major difference concerns the gap between our precision and recall scores, with our approach clearly being in favor of recall over precision, whereas the gap was not as wide in the results of Jia et al. (2019) and Peng et al. (2017). Our adopted method differed markedly from previously reported method since it did not require any training. Yet we relied on an OTR to drive the identification of argument instances and the re-composition of n-Ary relation instances. Furthermore, the presence of structured information in the n-Ary relation document tables obliged us to take our extraction process into account so as to complement it with argument instances previously identified in the text. All of these observations highlight the need for unified evaluation criteria able to reflect, measure and compare the specificities of n-Ary relation extraction tasks such as arity, scattering and sparsity of arguments in given corpora.

Comparison to open information extraction (OIE) approaches. We assessed how OIE approaches could be employed to extract targeted knowledge contained in our corpus. OIE represents information as RDF triples involving two entities, where we seek to extract n-Ary relations. Reconstructing these n-Ary relations based on the extracted triples is the approach sometimes adopted in the literature (Zhou et al., 2014), but we did not take these studies into account here since they were focused on a distinct task. Open IE 5.1[4] (Mausam, 2016) and the OIE

[4] https://github.com/dair-iitd/OpenIE-standalone.

function of Stanford Core NLP 4.4[5] (Manning et al., 2014) are standards in the OIE field and their code is easily accessible. Open IE 5.1 is based on a combination of CALMIE (Saha et al., 2018) for the extraction of conjunctive sentences, BONIE (Saha et al., 2017) for the extraction of numbers and measure units, RelNoun (Pal et al., 2016) to recognize noun phrases and SRLIE (Christensen et al., 2011) which is geared towards attributing a semantic role to the extracted entities. Stanford CoreNLP 4.4 is a Java toolkit that provide numerous functionalities and language models for natural language analysis.

We aimed to determine if these OIE tools could be used in zero-shot information extraction to extract triples (i.e. binary relations) between argument instances participating in our targeted n-Ary relations. In our corpus, Open IE 5.1 and CoreNLP 4.4 detected numerous triples, with 21333 and 86295 results, respectively. Using the labels of our OTR, we filtered out all triples that were not related to an argument of a targeted n-Ary relation. Only triples with at least one entity whose text included an OTR label were kept, which resulted in 2599 triples for Open IE 5.1 and 1325 triples for CoreNLP 4.4. The next step of our comparison focused on finding triples that could be used to reconstruct n-Ary relation instances by measuring the number of triples with both entities present in the same n-Ary relation instance of our Gold Standard. However, no triple meeting these criteria could be found in the CoreNLP 4.4 results. Open IE 5.1 only got 8 triples whose entities jointly appeared in one of the n-Ary relation instances of our Gold Standard. These results and manual observations highlight that these OIE techniques are not tailored for an n-Ary relation extraction pipeline targeting specific knowledge. First, a generic OIE system often fails to recognize quantitative entities containing complex numerical values or measurement units. Moreover, OIE extracts entity pairs based mostly on syntactical analysis at the sentence level and the n-Ary relations we target are highly scattered throughout the document. These systems are thus not able to correctly associate pairs of entities belonging to the same n-Ary relation instance.

Ablation experiment. We conducted a set of ablation experiments on the OTR used in order to estimate the dependency of our approaches on this semantic and terminological resource. The first set of experiments, as presented in Table 9, involved random division of OTR concepts representing the n-Ary relation arguments in three equal parts, with each containing 249 concepts. We then evaluated our approaches independently on each of these OTR subsets (i.e. A-B-C in Table 9). The recall, precision and F1-score values in Table 9 for each approach (i.e. Structural, Frequentist and Word Embedding) are average scores using the best criteria identified in Table 8 with 3 selected candidates. As expected, ablation of a third of the OTR concepts had a negative impact on the n-Ary instance reconstruction process, with the F-score decreasing to around .30. It also reduced the tendency of the Structural approach to favor recall over precision. The *B* subset especially presented a significant loss, and corresponded to the subset in which none of the concepts describing Temperature or Relative_Humidity arguments were present (i.e. since the arguments represent quantities they only have a handful of concepts in the OTR).

The second set of experiments consisted of three random ablations of different proportions of the OTR concepts (i.e. $10-25-50\%$). The results are presented in Table 10 and show that ablation of a proportion of the OTR generally does have a negative impact on the results, with an average F-score loss of .19, .18 and .30, respectively for the 10%, 25% and 50% ablation. Ablation of a random part of the OTR could also have some surprising side effects, e.g. when comparing the 10% with the 25% iteration, we observed a slight increase in recall and precision. We believe that this is due to an ablation of concepts whereby, if they possess some instances in our corpus, they are never considered correct argument instances. Typically, ablation of the most generic concepts representing symbolic arguments (e.g. *'packaging'*, *'multilayer_films'*) reduces noise in the candidate argument instances, i.e. in line with the patterns noted in the semantic relevance scores employed in our previous experiments (Lentschat, Buche, Dibie-Barthelemy et al., 2022).

The finding of these ablation experiments question the adequacy of using a semantic resource that might cover a broad array of concepts when conducting an information and relation extraction task that targets highly specific knowledge. This underlines the need for techniques able to filter out irrelevant results.

Table 9
Performances of the approaches depending on specific OTR ablation.

Approach	33% (A)			33% (B)			33% (C)			Macro avg.		
	r	p	f1	r	p	f1	r	p	f1	r	p	f1
Struct.	.35	.32	.33	.11	.20	.15	.32	.40	.36	.26	.31	.28
Freq.	.35	.28	.30	.15	.18	.16	.35	.27	.31	.28	.24	.26
W. Em.	.35	.29	.32	.12	.17	.14	.36	.29	.32	.28	.25	.26

r: recall, p: precision, f1: F1-score.

Table 10
Average performances of the approaches depending on a random OTR ablation proportion.

Approach	10%			25%			50%		
	r	p	f1	r	p	f1	r	p	f1
Structural	.72	.35	.47	.81	.37	.50	.53	.24	.33
Frequency	.35	.28	.31	.35	.28	.31	.30	.22	.25
Word embedding	.41	.29	.34	.42	.30	.35	.35	.23	.27

r: recall, p: precision, f1: F1-score.

6. Conclusion

The work presented in this paper contributes to the development of approaches for the extraction of n-Ary relations from scientific articles driven by an Ontological and Terminological Resource (OTR). This resource drives the reconstruction of n-Ary relation instances using its formalization of the knowledge of interest in a domain. Since an OTR contains a domain ontology that can be changed (e.g. @Web domain ontologies[6]), the approaches we propose could be applied to corpora of experimental articles in other domains. We adopted a conventional approach to this research issue to extract n-Ary relation instances (Zhou et al., 2014) and proceeded in two steps: (1) recognition of argument instances, and (2) reconstruction of n-Ary relation instances.

This paper focuses on the second step, with the main challenge being to find relevant criteria to reconstruct multiple n-Ary relations in each document. One original aspect of our contribution is that we initiate partial n-Ary relation instances based on the content of tables in the document. This preliminary instantiation process relies on the method proposed in Buche et al. (2013). The specificity of our approach is that, for a given partial n-Ary relation instance extracted from a table, we compare the whole set of arguments already instantiated with a candidate argument extracted in the text using a multi-criteria approach. We achieved this by designing a new multi-feature representation of partial n-Ary relation instances in document tables. This so-called STaRe (Scientific Table Representation) can be implemented jointly with SciPuRe, a previously designed representation of argument instances (Lentschat et al., 2020; Lentschat, Buche, Dibie-Barthelemy et al., 2022). Based on these representations, we extended three state-of-the-art approaches for reconstructing n-Ary relations combined with

[5] https://github.com/stanfordnlp/CoreNLP https://github.com/philipperemy/Stanford-OpenIE-Python.

[6] https://www6.inrae.fr/cati-icat-atweb/Ontologies.

the effect of using relevance scores for argument instances. A complementary expert assistance-based approach could also be adopted depending on the number of candidates to be considered. The structural approach is based on proximity criteria and associations between types of arguments, as well as on the article sections. The frequentist approach seeks co-occurrences with different textual manifestations in different contexts. The word embedding approach uses language models to measure semantic similarity between argument instances in a partial relation and candidate argument instances.

Our experiments were conducted on a corpus (Lentschat, 2022) constituting of 332 relation instances describing knowledge on gas permeability measures and packaging composition, with 1547 argument instances, in 10 documents.

This enabled us to draw general conclusions on the reconstruction of n-Ary relations. First, based on relevance scores of argument instances previously designed (Lentschat et al., 2020; Lentschat, Buche, Dibie-Barthelemy et al., 2022), filtering the least relevant proportion (20%) of the candidate instances improved both the recall and the precision in our n-Ary relation reconstruction process. Conversely, more extensive filtering of candidate argument instances had negative effects. This confirms that the relevance scores we designed previously are effective in filtering out the most obvious false positives, but valid instances may still have average relevance scores and be discarded via more extensive filtering. Selecting several candidates for argument instances to be manually reconstructed increases the recall and precision. Our results allow to place our approach in the assistance to the experts, by offering methods to select candidates in the reconstruction an n-Ary relation along with a complete and multi-features representation of the n-Ary relation as well as of these candidate instances. The most effective method to adopt depends on the approach chosen for knowledge extraction from scientific articles, with various extents of expert involvement.

Our experimental results highlight the difficulties associated with n-Ary relation reconstruction. Additional experiments showed that existing Open Information Extraction approaches are not suited for extracting triples belonging to n-Ary relation instances and that our method is moderately dependent on the coverage of the semantic resource used. Future research in the n-Ary relations extraction area could be focused on combining the different approaches implemented and exploitation of the different types of criteria. Further use of the OTR structure and domain knowledge could also drive and control the reconstruction of n-Ary relations by restricting, or focusing on, the selection of candidate instances through a set of rules. These rules could be drawn up by experts or learned automatically from a knowledge base in the application domain. We also suggest that the STaRe and SciPure representations could be useful in various research domains, as they may help to *'discriminate differences, identify similarities, describe accurately and minimize ambiguity'*. (Boyce et al., 2017) regarding information in a smart data perspective.

CRediT authorship contribution statement

Martin Lentschat: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Patrice Buche:** Conceptualization, Resources, Data curation, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Juliette Dibie-Barthelemy:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition. **Mathieu Roche:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data used are available on Dataverse, references and links to the data are present in the article.

Acknowledgments

The TEXT4LOD project has received funding from the IDEX/I-SITE MUSE[7] Université de Montpellier (France).

Ethics statement

Out of scope

References

Akimoto, K., Hiraoka, T., Sadamasa, K., & Niepert, M. (2019). Cross-sentence n-ary relation extraction using lower-arity universal schemas. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 6226–6232).

Berrahou, S. L., Buche, P., Dibie, J., & Roche, M. (2017). Xart: Discovery of correlated arguments of n-ary relations in text. *Expert Systems with Applications*, *73*, 115–124.

Boyce, B. R., Boyce, B. R., Meadow, C. T., Kraft, D. H., Kraft, D. H., & Meadow, C. T. (2017). *Text information retrieval systems*. Elsevier.

Brack, A., D'Souza, J., Hoppe, A., Auer, S., & Ewerth, R. (2020). Domain-independent extraction of scientific concepts from research articles. In J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, & F. Martins (Eds.), *Advances in information retrieval* (pp. 251–266). Cham: Springer International Publishing.

Buche, P., Dibie-Barthélemy, J., Ibanescu, L., & Soler, L. (2013). Fuzzy web data tables integration guided by an ontological and terminological resource. *IEEE Transactions on Knowledge and Data Engineering*, *25*(4), 805–819.

Bunescu, R., & Mooney, R. (2005). A shortest path dependency kernel for relation extraction. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 724–731).

Chan, Y. S., & Roth, D. (2011). Exploiting syntactico-semantic structures for relation extraction. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies* (pp. 551–560).

Christensen, J., Soderland, S., & Etzioni, O. (2011). An analysis of open information extraction based on semantic role labeling. In *Proceedings of the sixth international conference on knowledge capture* (pp. 113–120).

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22–29.

Cohen, K. B., Johnson, H. L., Verspoor, K., Roeder, C., & Hunter, L. E. (2010). The structural and content aspects of abstracts versus bodies of full text journal articles are different. *BMC Bioinformatics*, *11*(1), 492.

Davidson, D. (1980). *The logical form of action sentences. Essays on actions and events* (pp. 105–121). Oxford: Clarendon Press, [orginally published in 1967].

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library Review*.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171–4186). Minneapolis, Minnesota: Association for Computational Linguistics.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, *26*(3), 297–302.

Duong, T. H., Nguyen, H. Q., & Jo, G. S. (2017). Smart data: Where the big data meets the semantics. *Computational Intelligence and Neuroscience*, *2017*, Article 6925138.

Geng, Z., Li, J., Han, Y., & Zhang, Y. (2022). Novel target attention convolutional neural network for relation classification. *Information Sciences*, *597*, 24–37.

Geng, Z., Zhang, Y., & Han, Y. (2021). Joint entity and relation extraction model based on rich semantics. *Neurocomputing*, *429*, 132–140.

Ghersedine, A., Buche, P., Dibie-Barthélemy, J., Hernandez, N., & Kamel, M. (2012). Extraction de relations n-aires interphrastiques guidée par une RTO. In *CORIA: conférence en recherche d'information et applications* (pp. 179–190).

Giunti, M., Sergioli, G., Vivanet, G., & Pinna, S. (2019). Representing n-ary relations in the semantic web. *Logic Journal of the IGPL*.

Greenwood, M. A., & Stevenson, M. (2006). Improving semi-supervised acquisition of relation extraction patterns. In *Proceedings of the workshop on information extraction beyond the document* (pp. 29–35).

Grishman, R., & Sundheim, B. M. (1996). *Message understanding conference-6: a brief history*. Association for Computational Linguistics.

[7] https://muse.edu.umontpellier.fr/en/muse-i-site/.

Hignette, G., Buche, P., Dibie-Barthélemy, J., & Haemmerlé, O. (2009). Fuzzy annotation of web data tables driven by a domain ontology. In *European semantic web conference* (pp. 638–653). Springer.

Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines. *Bulletin de la Societe Vaudoise des Sciences Naturelles*, *37*, 241–272.

Jia, R., Wong, C., & Poon, H. (2019). Document-level N-ary relation extraction with multiscale representation learning. In *Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 3693–3704). Minneapolis, Minnesota: Association for Computational Linguistics.

Khattak, F. K., Jeblee, S., Pou-Prom, C., Abdalla, M., Meaney, C., & Rudzicz, F. (2019). A survey of word embeddings for clinical text. *Journal of Biomedical Informatics: X*, *4*, Article 100057.

Lenca, P., Vaillant, B., Meyer, P., & Lallich, S. (2007). Association rule interestingness measures: Experimental and theoretical studies. In *Quality measures in data mining* (pp. 51–76). Springer.

Lentschat, M. (2022). TRANSMAT n-ary relations. http://dx.doi.org/10.18167/DVN1/1BBJBQ.

Lentschat, M., Buche, P., Dibie-Barthelemy, J., Menut, L., & Roche, M. (2021). Food packaging permeability and composition dataset dedicated to text-mining. *Data in Brief*, *36*, Article 107135.

Lentschat, M., Buche, P., Dibie-Barthelemy, J., & Roche, M. (2020). SciPuRe: a new Representation of textual data for entity identification from scientific publications. In *Proceedings of the 10th international conference on web intelligence, mining and semantics* (pp. 220–226).

Lentschat, M., Buche, P., Dibie-Barthelemy, J., & Roche, M. (2022). Towards combined semantic and lexical scores based on a new representation of textual data to extract experimental data from scientific publications. *International Journal of Intelligent Information and Database Systems*, *15*(1), 78–103.

Lentschat, M., Buche, P., & Menut, L. (2021). TRANSMAT gold standard. http://dx.doi.org/10.18167/DVN1/U7HK8J.

Lentschat, M., Buche, P., Menut, L., & Guari, R. (2021). TRANSMAT tables data. http://dx.doi.org/10.18167/DVN1/GCZBC9.

Lentschat, M., Buche, P., Menut, L., Guari, R., & Roche, M. (2022). Partial n-ary relation instances on food packaging composition and permeability extracted from scientific publication tables. *Data in Brief*, *41*, Article 108000.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The stanford CoreNLP natural language processing toolkit. In *Association for computational linguistics (ACL) system demonstrations* (pp. 55–60).

Marcia, L. Z. (2017). Smart data for digital humanities. *Journal of Data and Information Science*, *2*(1), 1.

Mausam, M. (2016). Open information extraction systems and downstream applications. In *Proceedings of the twenty-fifth international joint conference on artificial intelligence* (pp. 4074–4077).

McDonald, R., Pereira, F., Kulick, S., Winters, S., Jin, Y., & White, P. (2005). Simple algorithms for complex relation extraction with applications to biomedical IE. In *Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL'05)* (pp. 491–498).

Meng, F., & Morioka, C. (2015). Automating the generation of lexical patterns for processing free text in clinical documents. *Journal of the American Medical Informatics Association*, *22*(5), 980–986.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In Y. Bengio, & Y. LeCun (Eds.), *1st international conference on learning representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, workshop track proceedings*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Mintz, M., Bills, S., Snow, R., & Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In *Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP* (pp. 1003–1011).

Neumann, M., King, D., Beltagy, I., & Ammar, W. (2019). ScispaCy: Fast and robust models for biomedical natural language processing. In *Proceedings of the 18th BioNLP workshop and shared task* (pp. 319–327). Florence, Italy: Association for Computational Linguistics.

Pal, H., et al. (2016). Demonyms and compound relational nouns in nominal open IE. In *Proceedings of the 5th workshop on automated knowledge base construction* (pp. 35–39).

Pawar, S., Palshikar, G. K., & Bhattacharyya, P. (2017). Relation extraction: A survey. arXiv preprint arXiv:1712.05191.

Peng, N., Poon, H., Quirk, C., Toutanova, K., & Yih, W.-t. (2017). Cross-sentence n-ary relation extraction with graph lstms. *Transactions of the Association for Computational Linguistics*, *5*, 101–115.

Peng, Y., Yan, S., & Lu, Z. (2019). Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets. CoRR abs/1906.05474.

Proux, D., Rechenmann, F., Julliard, L., et al. (2000). A pragmatic information extraction strategy for gathering data on genetic interactions. In *Ismb, Vol. 8* (pp. 279–285).

Quirk, C., & Poon, H. (2017). Distant supervision for relation extraction beyond the sentence boundary. In *Proceedings of the 15th conference of the European chapter of the association for computational linguistics: volume 1, long papers* (pp. 1171–1182). Valencia, Spain: Association for Computational Linguistics.

Ramadier, L., & Lafourcade, M. (2016). Patrons sémantiques pour l'extraction de relations entre termes - application aux comptes rendus radiologiques (here the title in english). In *Actes de la conférence conjointe JEP-TALN-RECITAL 2016. Volume 2 : TALN (Posters)* (pp. 514–521). Paris, France: AFCP - ATALA.

Role, F., & Nadif, M. (2011). Handling the impact of low frequency events on co-occurrence based measures of word similarity. In *Proceedings of the international conference on knowledge discovery and information retrieval (KDIR-2011). Scitepress* (pp. 218–223).

Ru, C., Tang, J., Li, S., Xie, S., & Wang, T. (2018). Using semantic similarity to reduce wrong labels in distant supervision for relation extraction. *Information Processing & Management*, *54*(4), 593–608.

Saha, S., Pal, H., et al. (2017). Bootstrapping for numerical open ie. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: short papers)* (pp. 317–323).

Saha, S., et al. (2018). Open information extraction from conjunctive sentences. In *Proceedings of the 27th international conference on computational linguistics* (pp. 2288–2299).

Shah, P. K., Perez-Iratxeta, C., Bork, P., & Andrade, M. A. (2003). Information extraction from full text scientific articles: where are the keywords? *BMC Bioinformatics*, *4*(1), 20.

Shahab, E. (2017). A short survey of biomedical relation extraction techniques. arXiv preprint arXiv:1707.05850.

Song, L., Zhang, Y., Wang, Z., & Gildea, D. (2018). N-ary relation extraction using graph-state LSTM. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 2226–2235). Brussels, Belgium: Association for Computational Linguistics.

Stanovsky, G., Michael, J., Zettlemoyer, L., & Dagan, I. (2018). Supervised open information extraction. In *Proceedings of the 2018 conference of the north American chapter of the association for computational linguistics: human language technologies, volume 1 (long papers)* (pp. 885–895).

Takanobu, R., Zhang, T., Liu, J., & Huang, M. (2018). A hierarchical framework for relation extraction with reinforcement learning. CoRR abs/1811.03925.

W3C Working Group (2006). Defining N-ary relations on the semantic web. https://www.w3.org/TR/2006/NOTE-swbp-n-aryRelations-20060412/. Accessed: 2021-06-10.

Wang, H., Lu, G., Yin, J., & Qin, K. (2021). Relation extraction: A brief survey on deep neural network based methods. In *ICSIM 2021*, *2021 the 4th international conference on software engineering and information management* (pp. 220–228). New York, NY, USA: Association for Computing Machinery.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. (2018). GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of the 2018 EMNLP workshop blackboxNLP: Analyzing and interpreting neural networks for NLP* (pp. 353–355). Brussels, Belgium: Association for Computational Linguistics.

Yang, S., Feng, D., Qiao, L., Kan, Z., & Li, D. (2019). Exploring pre-trained language models for event extraction and generation. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 5284–5294).

Yu, H., Cao, Y., Cheng, G., Xie, P., Yang, Y., & Yu, P. (2020). Relation extraction with BERT-based pre-trained model. In *2020 international wireless communications and mobile computing (IWCMC)* (pp. 1382–1387). IEEE.

Zhou, D., Zhong, D., & He, Y. (2014). Biomedical relation extraction: from binary to complex. *Computational and Mathematical Methods in Medicine*, *2014*.