Identification of Contradictory Knowledge Claims in Uncertain Environments: Postprint

Authors: Guo Jinjing, Huang Qi

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] Scientific research (especially medical research) is fraught with uncertainty. Identifying contradictory knowledge claims in research will help identify "scientific disagreements/inconsistent scientific conclusions" and facilitate the identification of potentially transformative scientific research and the improvement of related studies.

[Method/Process] Taking Alzheimer' s disease as a case study, PubMed abstract data was used as the data source, and the SemRep tool was employed for triple extraction. Identification rules for knowledge claims representing contradictory meanings were formulated. Source statements were categorized based on their level of uncertainty, and two approaches—single-sentence identification and cross-sentence identification—were respectively adopted to identify medical research knowledge claims with contradictory meanings represented in triple form.

[Results/Conclusion] A total of 49 groups (involving 277 triple pairs) of contradictory knowledge claims were identified from 6,574 medical abstracts sourced from PubMed. Research on Alzheimer' s disease in diagnosis and treatment still contains some controversies and contradictions that require further verification. The identification of contradictory knowledge claims provides new ideas for the discovery of potentially transformative frontiers in medical research, and can be applied to knowledge discovery based on knowledge metrics and to provide references for credibility calculation in knowledge graphs.

Full Text

Research on Identifying Contradictory Knowledge Claims Under Uncertain Environments

Guo Jinjing, Huang Qi School of Information Management, Nanjing University, Nanjing 210023

Abstract: [Purpose/Significance] Scientific research, particularly medical research, is fraught with uncertainty. Identifying contradictory knowledge claims within research helps to pinpoint "scientific divergence/inconsistent scientific conclusions," thereby facilitating the identification of potentially transformative scientific research and the improvement of related studies. [Method/Process] Using Alzheimer's disease as a case study, this research employs PubMed abstracts as the data source and utilizes the SemRep tool for triple extraction. Identification rules characterizing contradictory knowledge claims were formulated. Source sentences were categorized based on their degree of uncertainty, and contradictory medical research knowledge claims expressed as triples were identified through both single-sentence and cross-sentence recognition approaches. [Result/Conclusion] A total of 49 groups (involving 277 pairs of triples) of contradictory knowledge claims were identified from 6,574 medical abstracts sourced from PubMed. Research on the diagnosis and treatment of Alzheimer's disease still contains some controversies and contradictions that require further verification. The identification of contradictory knowledge claims provides new ideas for discovering potentially transformative medical research frontiers, which can be applied to knowledge discovery based on knowledge measurement and provide references for credibility calculation in knowledge graphs.

Keywords: knowledge discovery; uncertainty; contradictory knowledge; SPO; Alzheimer's disease

Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2021.20.013

The term "uncertainty" first appeared in James Mill's 1835 publication *Law Reform* [1], which noted that erroneous formulations in legal texts create ambiguity and uncertainty. In quantum mechanics, the A-B effect (Aharonov-Bohm effect) implies that the microscopic world is inherently uncertain [2]. The intrinsic uncertainty of microscopic objects necessarily determines that knowledge describing quantum objects is uncertain. Since knowledge reflects objects or objective processes, scientific knowledge describing the laws of motion in the objective world possesses uncertainty. All clinical data that medical research relies upon—including diagnostic results, patient medical histories, and physical examination findings—are often incomplete and carry uncertainty. Research indicates that approximately 11% of sentences in biomedical abstracts from the commonly used PubMed database employ hedging words or speculative expressions [3], exhibiting high uncertainty. In its inaugural 2021 editorial, *Cell* noted that understanding of scientific progress is not linear; every scientific conclusion is measured by uncertainty, and better experiments that refute earlier hypotheses represent not scientific failure but success. A good scientist should properly confront better evidence that contradicts previous conclusions or hypotheses [4]. Identifying "scientific divergence/inconsistent scientific conclusions" in medical research helps reveal controversies or contradictions in current research and thereby seek better evidence to advance medical research through debate.

Alzheimer's disease (AD) is a form of dementia. According to data from

Alzheimer' s Disease International (ADI) [5], over 50 million people worldwide currently suffer from Alzheimer' s disease or other types of dementia, with this number projected to reach 152 million by 2050. According to the Chinese Geriatrics Association website' s 2019 data [6], China has approximately 10 million Alzheimer' s patients, making it the country with the largest number of AD patients globally, with projections exceeding 40 million by 2050. Building upon the principle of uncertainty and previous research on contradictory/conflict predicate selection, this study formulated identification rules for triples representing contradictory/conflict meanings. Using Alzheimer' s disease as an example, diagnostic and therapeutic clinical knowledge claims expressed as triples and characterized by contradictions and controversies were extracted from massive medical texts, aiming to identify knowledge points with potential breakthrough research tendencies in this field and provide new approaches for identifying medical research frontiers while uncovering potential new diagnostic and treatment methods.

2 Methods and Tools for Identifying Contradictory/Conflicting Knowledge Claims

2.1 Theoretical Foundation and Related Research Progress

In *The Structure of Scientific Revolutions: 50th Anniversary Edition* [7], Thomas Kuhn divides scientific research into normal science and transformative research according to the degree of paradigm shift. Normal science aims to "solve puzzles," working to answer current questions in a knowledge domain and tending to discover what it expects to find—a process of continuously proposing and verifying hypotheses. Transformative research aims to continuously discover new phenomena and anomalies (i.e., "contradictions/conflicts") and attempts to explain or resolve such anomalies, ultimately generating new theories. The relationship between these two types of research in terms of paradigm shift degree and uncertainty level is shown in Figure 1 [Figure 1: see original paper] [8]. Normal science corresponds to the process of hypothesis verification, while transformative research corresponds to the process of resolving controversies and contradictions.

According to the Oxford Dictionary (2nd edition), contradiction refers to the existence of two or more incompatible claims describing the same proposition [9]. C. M. Chen and M. Song consider uncertain scientific knowledge to be knowledge in a state of hypothesis or controversy, noting that scientific progress must address disputed, contradictory, incomplete, and inconsistent information [10]. The principle of uncertainty provides a theoretical foundation for identifying contradictions/conflicts in the scientific research process, thereby offering possible new approaches for identifying potentially transformative scientific research.

Regarding knowledge claims, M. W. McElroy defines them in *The New Knowledge Management: Complexity, Learning, and Sustainable Innovation* as "conjectures, assertions, arguments, or theories about what potential actions may

lead to expected results, aimed at narrowing the gap between current state and target state" [11]. K. Peters et al. note that knowledge claims are diverse, including theories, models, declarations, methods, and designs [12]. On one hand, knowledge claims involve how to perform tasks or change things; on the other hand, knowledge claims may be based on other knowledge claims, concerning how things currently and will function or operate. Knowledge claims in scientific research represent the most important content researchers want to present to readers—the overall conclusions or results that can be understood from research findings or results. Knowledge claims with opposite assertion values are considered potentially contradictory knowledge claims [13].

The emergence of contradictions/conflicts can be divided into syntactic-level conflicts and semantic-level conflicts. Syntactic-level conflicts are primarily caused by various linguistic structures, such as antonyms, negation (including direct and indirect negation), numerical or temporal mismatches, implicit assumptions about the degree of factual truth (including causative verbs, implicit verbs, and non-causative verbs), modal expressions (including epistemic modal verbs and deontic modal verbs), structures (including lexical and syntactic structures), background knowledge, etc. [14]. Semantic-level contradictions mainly arise from differences in interpretation and textual entailment [15]. Correspondingly, methods for identifying contradictions/conflicts can be divided into two categories: for syntactic-level contradictions, analysis of linguistic features such as negation, contrast, and antonyms is required; for semantic-level contradictions, textual entailment between texts can be measured after removing negative propositions—if entailment holds, the text pair is considered contradictory [16].

"Scientific divergence/inconsistent scientific conclusions" refers to situations where two knowledge claims involve the same thing but cannot both be true, thus being considered mutually contradictory. Focusing on measuring this type of uncertain knowledge, some scholars have analyzed citation sentences to evaluate researchers' assessments of uncertainty regarding scientific claims. H. Kilicoglu et al. [17] performed sentiment analysis on citation sentences in clinical medical research literature based on the degree of alignment between citation sentiment and cited clinical research conclusions, where negative sentiment indicates inconsistency with the cited clinical research conclusions. D. Murray et al. [18] constructed cue words representing scientific disagreement (contradict, conflict) and filter words (studies, results) to measure scientific disagreement in citation sentences. Analysis based on citation sentences detects co-occurrence relationships between concepts but cannot identify structured, causal relationships representing inconsistent conclusions. Additionally, some scholars have approached contradictory medical knowledge identification from the perspectives of textual content analysis and knowledge units.

Regarding predicate selection and classification, A. Alamri [13] divides predicates into active/causative categories (e.g., AUGMENTS, CAUSES), passive/inhibitory categories (e.g., DISRUPTS, PREVENTS), and other

categories (e.g., ADMINISTERED_{TO}, OCCURS_{IN}). G. Rosenblat et al. [19] divide predicates into four pairs of opposing predicates with causal meaning and four pairs without causal meaning. J. M. G. Pinto et al. [20] selected seven predicates to identify contradictory and diverse knowledge. X. Y. Li et al. [21] divided predicates into excitatory and inhibitory categories and established identification rules to discover contradictory/conflicting knowledge claims and diverse knowledge claims.

Synthesizing the above research status, this study argues that different knowledge claims express different degrees of knowledge uncertainty. It is necessary to first distinguish the degree of uncertainty based on whether source sentences contain cue words representing uncertainty, then use the extracted triples from source sentences as the judgment object. By extracting predicates to determine relationships between entities and applying rules to judge whether two triples are contradictory/conflicting, the analysis angle shifts from correlation to causation, making it easier to explain the origins of contradictions/conflicts.

2.2 Triple Extraction Tools

Commonly used triple extraction tools include Open Information Extraction (OpenIE) [22], Goldfire [23], and SemRep [24], which target open information, patents, and biomedical domains respectively. This study uses the SemRep tool to extract semantic relations from biomedical texts. The subjects and objects of triples extracted by SemRep are concepts with specific semantic types from the UMLS Metathesaurus, while predicates are sourced from extended UMLS Semantic Network relationship types. SemRep can analyze linguistic features such as negation, contrast, and antonyms at the syntactic level, and reveal textual entailment at the semantic level based on identified semantic relationships. The main types of semantic relationships SemRep can identify include: clinically related (e.g., treats, diagnoses, process_{of}), molecular interactions (e.g., interacts_{with}, inhibits, stimulates), disease etiology (e.g., associated_{with}, causes, predisposes), pharmacogenomics (e.g., affects, augments, disrupts), static relations (e.g., is_a, part_{of}, location_{of}), etc., which can meet the needs of this study for semantic predicate types.

SemRep demonstrates high precision for different types of semantic relationships. For specific predicate domains (e.g., pharmacogenomics) and specific linguistic structures (e.g., comparative predicates), the accuracy reaches 76% [25] and 96% [26] respectively. In sentence-level evaluation based on the standard benchmark corpus CDR (chemical-disease relation), when predicates were limited to causally related semantic predicates such as "causes, affects, augments, stimulates, predisposes, and associated_{with}," the extraction accuracy reached 90% [27], meeting this study's requirements for the accuracy of extracted triples. SemRep offers four usage modes: web interaction, batch processing, API remote call, and local installation. This study adopts batch processing mode for triple extraction, using version 1.8, with output in txt format.

2.3 Methods and Rules for Identifying Contradictory/Conflicting Knowledge Claims

The overall identification approach for contradictory/conflicting knowledge claims is shown in Figure 2 [Figure 2: see original paper]. Before conducting single-sentence and cross-sentence recognition, the uncertainty degree of source sentences must be determined. If source sentences contain cue words such as "may," "could," or "might" indicating high uncertainty, they are classified into the high-uncertainty knowledge claim identification process; otherwise, they are classified into the low-uncertainty knowledge claim identification process.

Contradictory/conflicting knowledge claim identification is divided into two categories: single-sentence recognition and cross-sentence recognition [21].

For single-sentence recognition, sentences can be checked for explicit cue words such as "controversial," "debatable," "contentious," "contradictory," "contrary," "opposing," and "conflicting" to determine whether the sentence contains contradictory/conflicting implications, then extract triples from the sentence.

This study defines the triple representation as T_{in} = (C_{i1}, R_{in}, C_{i2}), extracted from sentence S_i, where C_{i1} is the subject, C_{i2} is the object, and R_{in} is the predicate of triple T_{in}. The cue word set is defined as M = {W1, W2, ···, Wn}, N as the set of all words constituting S_i {N_{i1}, N_{i2}, ···, N_{in}}, and U as the set of triples containing contradictions/conflicts. The single-sentence recognition rules are:

Input S_i // Input sentence S_i
If N M ≠ ϕ // If S_i contains any cue word from set M
Then T_{in} = (C_{i1}, R_{in}, C_{i2}) U // Then the triple T_{in} extracted from S_i is a contradictory/conflicting triple

For cross-sentence recognition, judgments can be made based on the following rules: extract two or more "Subject-Predicate-Object (SPO)" triples from scientific claims that share the same subject and object but have opposite predicate semantic types.

Define another triple T_{jm} = (C_{j1}, R_{jm}, C_{j2}), extracted from sentence S_j. Define the contradictory semantic predicate pair set P = {(R_{i1}, R_{j1}), (R_{i2}, R_{j2})···, (R_{in}, R_{jm})} and V as the set of mutually contradictory/conflicting triple pairs. The cross-sentence recognition rules are:

Input T_{in}, T_{jm} // Input triples T_{in} and T_{jm}
If C_{i1} = C_{j1}, C_{i2} = C_{j2}, (R_{in}, R_{jm}) P // If triples T_i and T_j have the same subject and object, and the predicate pair comes from any element in set P
Then (T_{in}, T_{jm}) V // Then triples T_{in} and T_{jm} extracted from S_i and S_j respectively are contradictory/conflicting triple pairs

Based on requirements, this study selected semantic relationship types closely

related to clinical diagnosis and treatment from the numerous types SemRep can identify. The main predicates include: treats, causes, prevents, predisposes, augments, disrupts, stimulates, inhibits, with their negated forms (e.g., neg_{treats}) also considered. The resulting opposite predicate pairs are shown in Table 1 , including predicate pairs involving causal relationships and those lacking causal relationships. If two triple pairs share the same subject and object, and their predicates are contradictory predicate pairs listed in Table 1, they are determined to be contradictory knowledge claim pairs.

Table 1 shows the semantic predicate pairs representing contradictory relationships used in this study, including both causally related and non-causally related predicate pairs.

An example pair of contradictory knowledge claims identified using the above rules is shown in Table 2 . Sentences 31895707.ab.1 [28] and 32476013.ti.1 [29] express contradictory knowledge claims due to the opposing predicate pair "predisposes" and "neg_{predisposes}" —namely, whether "proton pump inhibitors" do or do not increase the risk of "dementia."

3 Identification Results of Contradictory/Conflicting Knowledge Claims

3.1 Data Acquisition and Processing

Scientific literature is the primary carrier of scientific knowledge and can be understood as records of knowledge claims that address scientific questions and provide evidential support. PubMed (https://pubmed.ncbi.nlm.nih.gov/) is a database providing citations and abstracts for biomedical literature, developed by the U.S. National Library of Medicine (NLM). It currently contains over 32 million biomedical literature citations and abstracts from MEDLINE, life science journals, and online books, serving as a frequently used literature retrieval platform for medical researchers.

Therefore, this study uses PubMed as the data source for contradictory knowledge claim identification. The credibility of evidence/claims differs across literature types [19]. In the real world, compared with case studies, randomized controlled trials (RCT) and systematic reviews have higher levels of clinical evidence [30] and can extract more credible and accurate contradictory/conflicting discoveries from supporting sentences. Consequently, this study uses "Alzheimer' s Disease"as the topic, limits document types to clinical guidelines, meta-analyses, clinical trial articles, comparative studies, and other literature types with higher credibility and accuracy, and retrieves English articles published in the past 10 years. The retrieval date was January 26, 2021. The expanded search query was:

("alzheimer disease"[MeSH Terms] OR ("alzheimer"[All Fields] AND "disease"[All Fields]) OR "alzheimer disease" [All Fields] OR ("alzheimers" [All Fields] AND "disease"[All Fields]) OR "alzheimers disease"[All Fields]) AND ((y_{10}[Filter])

AND (clinicalstudy[Filter] OR clinicaltrial[Filter] OR clinicaltrialphasei[Filter] OR clinicaltrialphaseii[Filter] OR clinicaltrialphaseiii[Filter] OR clinicaltrialphaseiv[Filter] OR comparativestudy[Filter] OR controlledclinicaltrial[Filter] OR guideline[Filter] OR meta-analysis[Filter] OR multicenterstudy[Filter] OR practiceguideline[Filter] OR pragmaticclinicaltrial[Filter] OR randomizedcontrolledtrial[Filter] OR systematicreview[Filter]) AND (english[Filter]))

Using this search query, 6,574 articles related to "Alzheimer's Disease" published in the past 10 years were retrieved from PubMed. Article abstract information (including titles and PMID numbers) was imported into the SemRep tool. After steps including tokenization, lexical/syntactic analysis, phrase mapping, semantic predicate normalization, and concept semantic type constraints, a total of 45,674 triple pairs involving 76,635 sentences were obtained. Based on whether sentences contained cue words such as "may," "could," or "might" indicating high uncertainty [31], the 76,635 sentences were divided into two categories: those with higher uncertainty (containing "may," "could," "might," etc.) and those with lower uncertainty (without such cue words). Single-sentence and cross-sentence identification of contradictory/conflicting knowledge claims was then conducted according to the rules in section 2.3. The specific experimental process and results are shown in Figure 3 [Figure 3: see original paper].

After removing source sentences from which triples could not be extracted, the final number of source sentences and triples included in the analysis is shown in Table 3 . After identifying contradictory/conflicting triple pairs according to the rules, experts reviewed the extraction results to obtain final results.

3.2 Identification of Contradictory/Conflicting Knowledge Claims with Higher Uncertainty

Based on the presence of cue words such as "may," "could," and "might," 3,774 sentences were selected from the 76,635 sentences, yielding 2,890 triple pairs from 1,702 valid sentences (some sentences contained high-uncertainty words but could not yield triples). The analysis proceeds separately for single-sentence and cross-sentence recognition.

3.2.1 Single-Sentence Identification of Contradictory/Conflicting Knowledge Claims with Higher Uncertainty According to the identification rules established in section 2.3, the first step is to check whether the 1,702 sentences contain cue words: "controversial," "debatable," "contentious," "contradictory," "contrary," "opposing," and "conflicting." Sentences containing these cue words are determined to be statements with contradictory/conflicting implications, and triples are then extracted. After screening and expert judgment, 5 triple pairs from 3 sentences were obtained (see Table 4). For example, sentence 30098211.ab.1 [32] shows that the relationship between benzodiazepines and dementia is questionable and requires further verification.

3.2.2 Cross-Sentence Identification of Contradictory/Conflicting Knowledge Claims with Higher Uncertainty After removing the 5 triple pairs identified in section 3.2.1, cross-sentence identification was performed on the remaining 2,885 triple pairs according to the rule "same subject and object, opposite predicate semantic types." Three groups representing contradictory/conflicting knowledge claims were identified, involving 13 triple pairs. After expert evaluation, one group was retained (see Table 5). Sentence 21647938.ab.3 [33] states that brain-derived neurotrophic factor (BDNF) may contribute to the development of AD (Brain-Derived Neurotrophic Factor-predisposes-Alzheimer' s Disease), while sentence 31518516.ab.8 [34] indicates that BDNF may not be an optimal biomarker for AD diagnosis, meaning its relationship with AD is not significant (Brain-Derived Neurotrophic Factor-neg_{predisposes}-Alzheimer' s Disease). The knowledge claims expressed by these two triple pairs are controversial.

3.3 Identification of Contradictory/Conflicting Knowledge Claims with Lower Uncertainty

3.3.1 Single-Sentence Identification of Contradictory/Conflicting Knowledge Claims with Lower Uncertainty Excluding the 3,774 high-uncertainty sentences from section 3.2, there were 72,861 sentences without "may," "could," "might," or similar cue words. After processing and removing sentences from which triples could not be identified, 42,775 triple pairs from 25,394 valid sentences were obtained. The analysis proceeds separately for single-sentence and cross-sentence recognition.

Using the screening method from section 3.2.1, the first step is to check whether the 25,394 sentences contain cue words: "controversial," "debatable," "contentious," "contradictory," "contrary," "opposing," and "conflicting." Sentences containing these words are determined to be statements expressing contradiction/conflict, and triples are then extracted. After initial screening, 76 sentences containing these cue words were found, yielding 125 triple pairs. After secondary screening, 13 triple pairs related to AD diagnosis and treatment from 12 sentences were obtained. After expert review, 10 contradictory/conflicting triple pairs from 9 sentences were finalized (see Table 6).

For example, existing research indicates that the role of antioxidants in AD prevention, the therapeutic effects of Ginkgo biloba extract EGb761 and memantine on AD, whether the vascular endothelial growth factor (VEGF) gene is a risk factor for AD, and the preventive effects of supplementation are all controversial or contradictory.

3.3.2 Cross-Sentence Identification of Contradictory/Conflicting Knowledge Claims with Lower Uncertainty After removing the 125 triple pairs identified in section 3.3.1, cross-sentence identification was performed on the remaining 42,650 triple pairs from 25,318 valid sentences. According to the rules established in section 2.3, 87 groups representing con-

tradictory/conflicting knowledge claims were identified, covering 807 sentences. Since this study focuses on AD-related diagnostic and therapeutic knowledge claims, triples with universal subjects (e.g., diagnosis) or objects (e.g., patients, disease, persons) were excluded from the research scope. Additionally, sentences from titles are often neutral statements (especially in review articles) and cannot yield conclusive knowledge claims, so triples extracted from titles were removed. After processing, 64 groups of contradictory/conflicting triple pairs related to AD diagnosis and treatment were obtained, covering 490 sentences. After expert review, 36 groups comprising 259 triple pairs were finalized.

Based on the extraction results, triple pairs related to Alzheimer' s disease can be divided into diagnostic and therapeutic categories. Table 7 shows that in AD diagnosis, general anesthesia, impaired cognition, neurogranin, and tau proteins are closely related to AD etiology, pathology, or pathogenesis, but the relevant mechanisms remain highly controversial.

The National Health Commission' s "Diagnostic and Treatment Standards for Mental Disorders (2020 Edition) - Alzheimer' s Disease" [35] states that the pathological changes in AD are mainly characterized by senile plaques (SP), neurofibrillary tangles (NFT), and neuronal loss. The center of SP is β-amyloid protein (Aβ), and the imbalance of Aβ production and clearance is the initiating factor for neuronal degeneration and dementia. The main component of NFT is tau protein, and Aβ induces a series of pathological processes including tau hyperphosphorylation, inflammatory responses, and neuronal death. Additionally, neurotransmitter abnormalities (including acetylcholine, monoamine, amino acid, and neuropeptide systems) are also major symptoms in AD patients. The AD diagnosis-related contradictory triples identified in this study are largely consistent with the content of these diagnostic standards, indicating that the identification results are relatively accurate and can serve as a reference point for subsequent research improvement.

Corresponding to the "Diagnostic and Treatment Standards for Mental Disorders (2020 Edition) - Alzheimer' s Disease," this study divides identified AD treatment methods into drug therapy and non-pharmacological interventions. In terms of drug therapy (see Table 8), the roles of agonists, antidiabetics, nonsteroidal anti-inflammatory agents, antioxidants, aspirin, calcium channel blockers, cholinesterase inhibitors, donepezil, hydroxymethylglutaryl-CoA reductase inhibitors, intravenous immunoglobulins, insulin, insulin-like growth factor I, memantine, monastrol, proteins, testosterone, vitamin D, and vitamin E in treating AD or dementia and improving cognitive and neuropsychiatric symptoms (such as depression) remain contradictory or controversial.

In terms of non-pharmacological interventions (see Table 9), the therapeutic effects of physical activity, prophylactic treatment, and other methods on AD, dementia, cognition, or neuropsychiatric symptoms still contain some contradictions or controversies.

4 Discussion

4.1 Triple-Based Contradictory/Conflicting Knowledge Claim Identification Provides a New Approach for Knowledge Discovery

Currently, academic literature measurement research is undergoing a transformation from focusing on external characteristics to content characteristics, and from syntactic features to semantic and even pragmatic features [36]. Frontier identification methods based on citations display relationships between entities through concept co-occurrence, showing only correlation without causal nature. For example, drug-disease co-occurrence cannot determine whether a drug has therapeutic effects or induces adverse events. Knowledge unit identification based on triples deepens the approach from bibliometrics to knowledge metrics. Unlike currently prevalent methods for identifying research hotspots or frontiers based on highly cited papers, triples can reveal what kind of relationship exists between entities and "how" they interact, making it easier for researchers to accurately understand interaction mechanisms and conduct targeted follow-up research. Identifying existing contradictory/controversial knowledge claims can provide references for researchers to identify research frontiers where consensus has not been reached, help solve practical problems, and uncover potential new methods, technologies, and knowledge.

Meanwhile, during the process of identifying contradictory/conflicting knowledge claims, understanding the timing of different knowledge claims in each group (see Figure 4 [Figure 4: see original paper]) can roughly reveal how a research question evolves over time, dynamically understand the development stage of contradictory/controversial knowledge claims, and provide reference points for problem-solving entry points. Using hydroxymethylglutaryl-CoA reductase inhibitors as an example, Figure 4 illustrates the evolution of research on their interaction with AD (for display purposes, predicates were assigned values according to their positive/negative meanings and uncertainty degree: "prevents" = 1, "treats" = 2, "predisposes" = -1). A 2018 study [37] raised questions about the therapeutic or preventive effects of statin hydroxymethylglutaryl-CoA reductase inhibitors on AD, suggesting that "fungus-derived statins" are associated with increased AD risk, providing different research conclusions that help advance existing studies on the mechanism of statins' effects on AD.

4.2 Measuring Knowledge Uncertainty Degree Provides Reference for Knowledge Graph Credibility Calculation

This study uses the presence of high-uncertainty cue words such as "may,""could," and "might" in source sentences as the basis for classification, dividing processed source sentences into high-uncertainty and low-uncertainty sentences, then conducting single-sentence and cross-sentence identification respectively. The triples extracted thereby have different uncertainty degrees. Clinical decision-making is full of uncertainty; doctors' decision-making is a process from "knowledge" to "wisdom," addressing the question of "how to make optimal

decisions under uncertain conditions." In the cross-sentence identification results shown in Tables 7-9, the frequency of predicate occurrence is marked, which can be used to calculate the frequency of a specific triple pair within its knowledge claim group as a basis for quantifying the uncertainty degree of specific triples. This also provides deterministic degree or confidence references for building knowledge graphs supporting clinical decision-making. Complete knowledge units can provide evidence sources for knowledge claims. Accordingly, when constructing knowledge graphs, the probability of interaction between subjects and objects in triples can be marked to provide more convincing references for clinical decision-making based on knowledge graphs.

4.3 Determination of Contradictory/Conflicting Knowledge Claims Requires Professional Involvement

This study identified contradictory/conflicting knowledge claims based on uncertainty degree classification, identifying 15 groups (18 triple pairs) from single sentences and 67 groups (503 triple pairs) from cross sentences, totaling 521 triple pairs. After expert review, 372 triple pairs were determined to be accurately identified, yielding an accuracy rate of 71.4%. After removing triple pairs that could not constitute contradictory/conflicting relationships due to identification errors, 49 groups (covering 277 triple pairs) of contradictory/conflicting knowledge claims were finalized. Analysis reveals the following limitations:

(1) Triple identification contains errors. One type involves named entity recognition errors (see Table 10), and another involves semantic relationship (predicate) extraction errors (see Table 11). Although SemRep's accuracy for identifying different semantic relationship types ranges from 59%-96% [27], basically meeting requirements, expert review and correction of identified triples are still necessary.

(2) Cue words and identification rules representing contradiction/conflict need optimization. Although this study expanded the range of cue words representing contradiction/conflict (from 3 to 7) compared with existing research, further supplementation and optimization are needed. For cross-sentence identification rules, some scholars have listed predicate pairs representing contradiction/conflict, while others divide predicates into excitatory and inhibitory groups representing contradiction/conflict. The first method clearly reveals causal relationships in identified triples but is limited by the number of identification rules, resulting in fewer identified knowledge claims. The second method facilitates identification of more contradictory/conflicting knowledge claims but cannot clearly obtain causal relationships in triples. Both methods have advantages and disadvantages, and selection should be based on research needs.

Conclusion

Research frontiers in various disciplines represent difficulties, hotspots, and development trends, guiding the direction of scientific development in various countries. In 2016, the State Council issued the "National Innovation-Driven Development Strategy Outline" [38], stating that to achieve the grand goal of building a world science and technology innovation powerhouse by 2050, it is necessary to "strengthen original innovation toward scientific frontiers, strive to lead world scientific research directions in more fields,"and "promote transformative research, actively exploring new ideas, new discoveries, new knowledge, new principles, and new methods." Against the background of many countries worldwide regarding innovation-driven development as a core strategy for seeking competitive advantages, identifying scientific research frontiers and conducting transformative research are necessary means to enhance innovation capabilities at scientific frontiers and win development initiative.

This study takes Alzheimer' s disease as an example, adopts the transformative research identification method based on contradiction/conflict detection from the principle of uncertainty, formulates knowledge claim identification rules representing contradictory/conflicting meanings, and identifies contradictory/controversial clinical knowledge claims expressed as triples from massive medical texts. The results were compared and verified with the content of the "Diagnostic and Treatment Standards for Mental Disorders (2020 Edition) - Alzheimer' s Disease" issued by the National Health Commission. The final results show that the scope of identified contradictory/conflicting triple groups is largely consistent with the guideline content regarding AD diagnosis and treatment, and the identification results of contradictory knowledge claims highly align with current research directions. This experiment validates the identification of uncertain knowledge claims based on contradiction/conflict, helps discover potential research frontiers related to Alzheimer' s disease research that remain controversial, and can provide references for guiding related research. Future work will further expand this method in directions including optimizing uncertainty degree measurement methods, extending from declarative sentences to citation sentences, and enriching cue words and identification rules representing contradiction/conflict.

Acknowledgments

We thank Sheng Shu, Yang Yang, Xie Qiwen, Qin Xinguo, and Yin Tianheng for their valuable suggestions during the research process.

References

[1] JAMES M. Law reform [J]. London Review, 1835, 2(3): 1-3.
[2] WU Guolin. On the objectivity of knowledge [J]. Science of Science and Management of S&T, 2000, 21(6): 37-39.
[3] LIGHT M, QIU X Y, SRINIVASAN P. The language of bioscience: facts,

speculations, and statements in between [C] // HLT-NAACL 2004 workshop: linking biological literature, ontologies and databases. Boston: Association for Computational Linguistics, 2004: 17-24.
[4] The Cell editorial team. What happens next [J]. Cell, 2021, 184(2): 291-292.
[5] Alzheimer' s Disease International. Dementia facts & figures [EB/OL]. [2021-02-24]. https://www.alzint.org/about/dementia-facts-figures/.
[6] Chinese Geriatrics Association. Initiative: China should continue to increase attention and investment in Alzheimer' s disease [EB/OL]. [2021-02-24]. https://www.zglnyxh.com/site/content/3036.htm.
[7] KUHN T S, HACKING I. The structure of scientific revolutions: 50th anniversary edition [M]. Chicago: University of Chicago Press, 2012: 72-89.
[8] DU Jian. Progress and prospects of medical knowledge uncertainty measurement [J]. Data Analysis and Knowledge Discovery, 2020, 4(10): 14-27.
[9] Oxford-Dictionary. Oxford english dictionary online, 2nd edition [EB/OL]. [2021-02-24]. http://www.oed.com/.
[10] CHEN C M, SONG M. Representing scientific knowledge: the role of uncertainty [M]. Switzerland: Springer International Publishing, 2017: 1-351.
[11] MCELROY M W. The new knowledge management: complexity, learning, and sustainable innovation [M]. Burlington: Elsevier Butterworth-Heinemann, 2003: 7.
[12] PETERS K, MARUSTER L, JORN A R. Knowledge claim evaluation: a fundamental issue for knowledge management [J]. Journal of Knowledge Management, 2010, 14(2): 243-257.
[13] ALAMRI A. The detection of contradictory claims in biomedical abstracts [D]. Sheffield: University of Sheffield, 2016.
[14] MARNEFEFE M V, RAFFERTY A N, MANNING C D. Finding contradictions in text [C] // Association for Computational Linguistics 2008: Proceedings of the 46th annual meeting of the association for computational linguistics. Stroudsburg: Association for Computational Linguistics, 2008: 1039-1047.
[15] DAGAN I, GLICKMAN O, MAGNINI B. The PASCAL recognising textual entailment challenge [C] // Machine learning challenges. evaluating predictive uncertainty, visual object classification, and recognising tectual entailment. Southampton: Springer, 2005: 177-190.
[16] HARABAGIU S M, HICKL A, LACATUSU V F. Negation, contrast and contradiction in text processing [C] // AAAI' 06: Proceedings of the 21st national conference on artificial intelligence. Menlo Park: The AAAI Press, 2006: 755-762.
[17] KILICOGLU H, PENG Z, TAFRESHI S, et al. Confirm or refute?: a comparative study on citation sentiment classification in clinical research publications [J]. Journal of biomedical informatics, 2019, 91(3): 103123.
[18] MURRAY D, LAMMERS W, BOYACK K, et al. Measuring disagreement in science [C] // Proceedings of the 17th international conference of the international society for scientometrics and informetrics (ISSI). Leuven: ISSI Society, 2019: 2370-2375.
[19] ROSEMBLAT G, FISZMAN M, SHIN D, et al. Towards a characterization

of apparent contradictions in the biomedical literature using context analysis [J]. Journal of biomedical informatics, 2019, 98(10): 103275.
[20] PINTO J M G, WAWRZINEK J, BALKE W. What drives research efforts? find scientific claims that count! [C] // Proceedings of the 2019 ACM/IEEE joint conference on digital libraries. Washington: IEEE Press, 2019: 217-226.
[21] LI X Y, PENG S Y, DU J. Towards medical knowmetrics: representing and computing medical knowledge using semantic predictions as the knowledge unit and the uncertainty as the knowledge context [J]. Scientometrics, 2021, 126(7): 6225-6251.
[22] Turing Center. Open information extraction [EB/OL]. [2021-02-24]. https://openie.allenai.org/.
[23] IHS Markit. Patent solutions for research and discovery [EB/OL]. [2021-02-24]. https://ihsmarkit.com/products/patents-applications.html.
[24] RINDFLESH T C, FISZMAN M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text [J]. Journal of Biomedical Informatics, 2003, 36(6): 462-477.
[25] RINDFLESH T C, LIBBUS B, HRISTOVSKI D, et al. Semantic relations asserting the etiology of genetic diseases [C] // AMIA Annual symposium proceedings/AMIA Symposium. Rockville: American Medical Informatics Association, 2003: 554-558.
[26] FISZMAN M, DEMNER-FUSHMAN D, LANG F M, et al. Interpreting comparative constructions in biomedical text [C] // Biological, translational, and clinical language processing. Prague: Association for Computational Linguistics, 2007: 137-144.
[27] KILICOGLU H, ROSEMBLAT G, FISZMAN M, et al. Broad-coverage biomedical relation extraction with SemRep [J]. BMC Bioinformatics, 2020, 21(1): 1-28.
[28] KHAN M A, YUAN Y, IQBAL U, et al. No association linking short-term proton pump inhibitor use to dementia: systematic review and meta-analysis of observational studies [J]. American journal of gastroenterology, 2020, 115(5): 671-678.
[29] MADHAV D, VENKAT N, SACHIN S, et al. Proton pump inhibitors do not increase the risk of dementia: a systematic review and meta-analysis of prospective studies [J]. Diseases of the esophagus, 2020, 33(10): 1-9.
[30] HAYNES R B. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions [J]. Evidence based medicine, 2006, 38(10): 1010-1020.
[31] SMALL H. What makes some scientific findings more certain than others? a study of citing sentences for low-hedged papers [C] // Proceedings of the 17th international conference of the international society for scientometrics and informetrics (ISSI). Leuven: ISSI Society, 2019: 554-560.
[32] LUCCHETTA R C, MATA B P D, MASTROIANNI P D C. Association between development of dementia and use of benzodiazepines: a systematic review and meta-analysis [J]. Pharmacotherapy: the journal of human pharmacology and drug therapy, 2018, 38(10): 1010-1020.

[33] SU C, UNDERWOOD W, RYBALCHENKO N, et al. ERK1/2 and ERK5 have distinct roles in the regulation of brain-derived neurotrophic factor expression [J]. Journal of neuroscience research, 2011, 89(10): 1542-1550.
[34] XIE B, ZHOU H M, LIU W X, et al. Evaluation of the diagnostic value of peripheral BDNF levels for Alzheimer' s disease and mild cognitive impairment: results of a meta-analysis [J]. International journal of neuroscience, 2019, 130(1): 1-14.
[35] National Health Commission Office. Diagnostic and treatment standards for mental disorders (2020 edition) - Alzheimer' s disease [J]. Clinical Education of General Practice, 2021, 19(1): 4-6.
[36] LU Chao, ZHANG Chengzhi, WANG Yuzhuo, et al. Deepening semantic feature analysis: a review of full-text measurement analysis of academic literature [J]. Journal of Library Science in China, 2021, 47(2): 110-131.
[37] SINYAVSKAYA L, GAUTHIER S, RENOUX C, et al. Comparative effect of statins on the risk of incident Alzheimer disease [J]. Neurology: official journal of the American academy of neurology, 2018, 90(3): 179-187.
[38] Central Committee of the Communist Party of China and State Council issue the "National Innovation-Driven Development Strategy Outline" [EB/OL]. [2021-02-24]. http://www.gov.cn/xinwen/2016-05/19/content_{5074812}.htm.

Author Contributions:
Guo Jinjing: Topic selection, data processing, and initial draft writing;
Huang Qi: Paper structure optimization and revision.

Note: Figure translations are in progress. See original paper for figures.