2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)

Cross-document Validation with Argument Frequency Feature for Event Argument Identification

Qi Li
Beijing Institute of Control Engineering
Beijing, China
rickyleebit@163.com

Songzi Cao*
China Academy of Industrial Internet
Beijing, China
*caosongzi@china-aii.com

Jiehao Chen
China Academy of Industrial Internet
Beijing, China
chenjiehao@china-aii.com

Tao Li
Beijing Institute of Control Engineering
Beijing, China
ietaoli@163.com

Yinuo Zhang
Beijing Institute of Control Engineering
Beijing, China
zhangyn950303@163.com

***Abstract*—This paper introduces a novel cross-document feature called the Argument Frequency Vector (AFV) for event argument identification. Traditionally, this task involves evaluating all named entities as potential arguments to determine their relevance to the document's focal event. However, documents often don't exist in isolation, especially in cases like continuous news reports about one single event (e.g., the 2018 Hualien Earthquake). Our research reveals that while individual documents typically focus on specific aspects of an event, multiple documents often report on the same overall occurrence. This phenomenon creates strong inter-document connections, which we leverage through our proposed AFV feature. The AFV capitalizes on argument frequency across multiple documents related to a single event. By considering this cross-document information, we enhance the typical single-document approach to event argument identification. Our experiments demonstrate that the AFV feature complements traditional single-document features effectively. When AFV feature is integrated into the neural network, it yields significant performance improvements in identifying event arguments.**

Keywords-event argument identification, cross-document feature, argument frequency vector

I. Introduction

In natural language processing (NLP) research, event argument identification plays a fundamental role, which involves recognizing key details (arguments) associated with events mentioned in text. For example, in the event "2018 Hualien earthquake," "magnitude 6.4" would be identified as an argument describing the earthquake magnitude. This task is extremely important for various applications, including question answering [1][2][3] and information extraction [4][5].

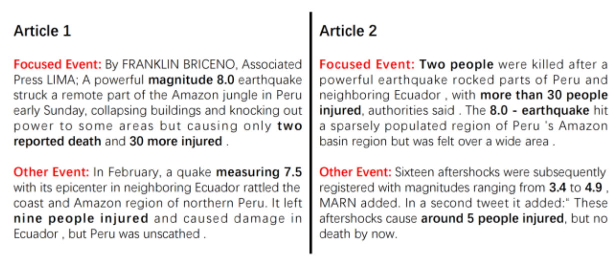
Event argument identification is particularly valuable for analyzing news articles and organizational annual reports, which often focus on detailed accounts of one specific event. In these contexts, the task typically prioritizes identifying arguments related to the reported main event, as these details usually contain the most relevant information for the text and subsequent analysis.

While articles primarily report on a central event, they often include references to other related events for context. These secondary events can complicate the argument identification process, acting as "noise" that challenges the accurate extraction of information about the primary event.

Traditionally, event argument identification has relied on models designed to analyze single documents [6]. Neural networks have been widely adopted for this task due to their ability to automatically extract underlying features. However, single-document approaches have a potential limitation: they overlook relationships across multiple documents. To address this, some researchers have proposed incorporating cross-document information. These cross-document models primarily focus on interactions between different events across documents, not fully leveraging the potential of strong cross-document interactions to identify arguments of the same event across multiple texts. This limitation suggests an opportunity for further research and improvement in the field.

Our observations reveal that many articles, particularly in continuous news coverage, often report on a single common event. In such cases, all related articles often reference the same key arguments of this shared focal event but may introduce different supporting or contextual events.

As a result, we've noticed that the arguments related to the primary event tend to appear with high frequency across multiple articles. In contrast, arguments associated with secondary or supporting events occur less frequently. This pattern of argument frequency provides a valuable signal for distinguishing the primary arguments from others.



Argument	8.0	two people	more than 30	7.5	3.4	4.9	nine people	5 people
Frequency	2	2	2	1	1	1	1	1

Figure 1. Example of arguments frequency in the articles

Fig. 1 illustrates our key observation. Both Article 1 and Article 2 mention the same arguments for the primary event: a magnitude of 8.0, two deaths, and more than 30 injuries. Consequently, these arguments appear with a frequency of 2 across the articles. In contrast, each article also reports on different secondary events. Article 1 discusses a previous earthquake in Peru, while Article 2 covers aftershocks. The arguments related to these secondary events appear only once across the two articles, resulting in a frequency of 1.

This example demonstrates a clear pattern: arguments with higher frequency across multiple articles are more likely to be associated with the reported primary event. This frequency-based distinction provides a valuable indicator for identifying the main event arguments in a collection of related articles.

To capture this characteristic, we introduce a novel feature called the Argument Frequency Vector (AFV). By utilizing the AFV, the neural network model achieves significant performance improvements in event argument identification in continuous news reports.

II. Literature Review

Existing approaches to event argument identification can be categorized into Single-document representation and Cross-document representation.

Single-document representation relies solely on information contained within individual documents. Some studies focus on lexical features [10][11]. Concurrently, many systems employ deep neural networks to automatically generate features for candidate arguments [12]. The Dynamic multi-pooling convolutional neural network (DMCNN) pioneered deep neural network-based approaches [8], while other researchers have explored the task from single-document question and answering [13]. Further advancing this line of research, Sha et al. introduced dependency relations and argument-argument interaction features to the RNN structure, demonstrating the ongoing evolution of these techniques [7].

Several researches have highlighted the significance of cross-document information in event argument identification [14][15]. Li Q et al. focus on the cross-document event and argument role relationship [16]. Ding B et al. considered Rationale-centric Counterfactual Data Augmentation across documents [17], while other researchers explored cross-document timeline relationships in news articles [18]. However, it's worth noting that these cross-document features are typically built on the assumption that events across different documents, while distinct, are interrelated.

III. Methodology

A. Overview

Fig. 2 illustrates the architectural framework of our proposed model incorporating the cross-document AFV feature. The model architecture comprises two primary components. The first component focuses on AFV feature construction, where documents about the same event are aggregated for analysis. This process involves quantifying argument occurrence frequencies within the document cluster of each focused event, followed by normalization and differentiation operations to encode cross-document information into the AFV feature. The second component implements classification using integrated features, where the cross-document AFV feature is concatenated with conventional single-document features. A neural network-based softmax classifier is then trained on these integrated features to determine whether candidate arguments correspond to the news-focused event.

B. Argument Frequency Vector Feature Building

Arguments within texts often appear in different forms despite having the same meaning. For instance, "U.S." and "America" both refer to the same country but use different terminology. To address this variation, we propose encoding all candidate arguments from documents related to the same event, and then employing clustering techniques to aggregate semantically coherent argumentative components. The cluster size corresponds to the frequency of occurrence within that specific argumentative category.

Candidate arguments are classified into two semantic types: word-based and number-based. Word-based arguments use textual descriptors like place names (e.g., "Singapore"), while number-based arguments employ quantitative information such as death tolls (e.g., "more than 100 people"). These distinct argument types require different frequency calculation methods. After calculating frequencies, differential and normalization techniques are applied to create the AFV feature, a novel cross-document feature representation.

Processing word-based arguments begins with removing stop words to reduce noise in the representation. Word embeddings ($w \in \mathbf{R}^{dw}$), with dw representing the embedding space dimensionality, are then applied to the remaining terms. Leveraging pre-trained embeddings enables effective semantic encoding by capturing contextual information from large corpora [19]. The argument representation is ultimately derived by calculating the mean vector of the word embeddings.

$$\mathrm{r}_{\mathrm{A}} = avg_{w_i \in A}(w_i) \quad (1)$$

$\mathbf{r}_{\mathrm{A}}$. represents the argument vector, where A encompasses the word embeddings of semantically related arguments.

Following argument representation, a clustering algorithm is implemented to aggregate semantically similar arguments into cohesive groups for frequency calculation. Specifically, this research utilizes DBSCAN (Density-Based Spatial Clustering of Applications with Noise) for clustering word-based arguments, as it effectively handles clusters of arbitrary shapes without requiring predetermined cluster quantities.

In the DBSCAN approach, Euclidean distance measures the similarity of arguments, enabling the detection of dense regions in the vector space that encompass semantically related arguments.

Upon completion of argument clustering, frequency statistics can be computed for each cluster. The population size of each argument cluster serves as a direct measure of its frequency. These cluster-wise frequencies are subsequently aggregated to form a comprehensive argument frequency representation that enables the systematic analysis of argument prevalence across the document corpus

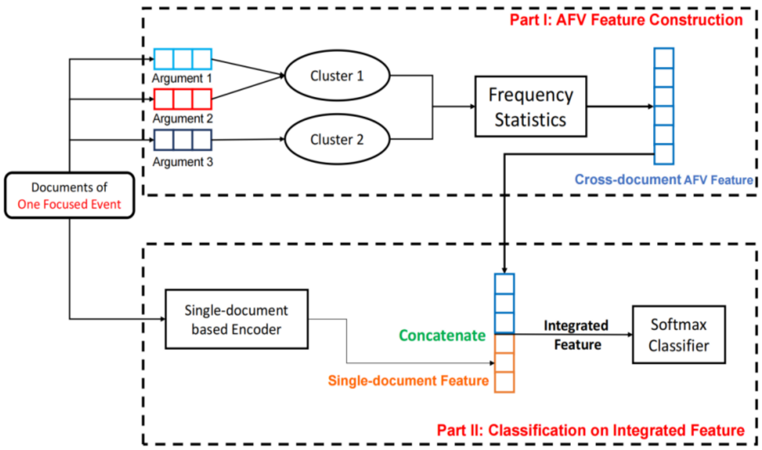


Figure 2. Overall architecture of our proposed AFV feature

The frequency vector is constructed by counting the element number of each clustered group. The frequency vector building mechanism is illustrated in Fig. 3.

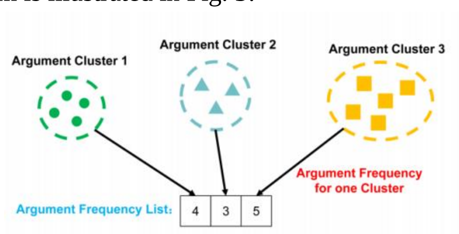


Figure 3. Cross-document argument frequency statistics

Number-based arguments involving numerical textual expressions pose distinct computational challenges in natural language processing. Such arguments often use approximate ranges instead of exact values (e.g., "more than 100 people"), requiring specialized processing techniques. We propose an algorithm that represents these arguments using numerical ranges and measures their semantic similarity through an overlap percentage metric.

Our methodology represents number-based arguments using a dual-range system with a float range and an extended range. The float range defines a numerical interval around the argument's core value N (e.g., 100 in "more than 100 people"). Its resolution is context-dependent, and determined by named entity tags within the dataset. For disaster-related data, we focus on three key numerical entities: magnitude measurements, fatality counts, and injury statistics. Through empirical validation, a resolution of 0.1N provides optimal granularity. Implementing this float range involves a specific mathematical representation that captures the nuanced quantitative characteristics of disaster event arguments.

$$[N-10^{\lfloor \log_{10} 0.1N \rfloor}, N+10^{\lfloor \log_{10} 0.1N \rfloor}] \tag{2}$$

The extended range component incorporates directional properties relative to a reference value N. This directional aspect manifests in two primary forms: greater than N (>N) and less than N (<N). The direction is determined by specific linguistic modifiers that precede the numerical value. We categorize these numerical descriptors into three distinct classifications: (1) expressions indicating values exceeding N (>N), exemplified by phrases such as "more than 100 people"; (2) expressions indicating values below N (<N), as in "up to 10 buildings"; and (3) expressions without directional implications, illustrated by phrases like "around 20 injured." Our methodology systematically catalogs these numerical descriptors and assigns appropriate directional attributes. The extended range is constrained within the magnitude of N, with its mathematical formulation expressed as follows:

$$\begin{aligned} &> N:[N, N+10^{\lfloor \log_{10} N \rfloor}] \\ &< N:[N-10^{\lfloor \log_{10} N \rfloor}, N] \end{aligned} \tag{3}$$

The final numerical range for each candidate argument is derived through the integration of its corresponding float and extended ranges. The composite range is constructed by selecting the minimum value between the two lower range limits as the definitive lower bound, while the maximum value between the two upper range limits establishes the definitive upper bound of the argument's numerical range.

To quantify the similarity between number-based arguments, we employ a metric based on the percentage overlap of their numerical ranges. Given two arguments with numerical ranges denoted as r_{A1} and r_{A2}, the overlap region is defined by establishing boundaries through comparative operations. Establishing the overlap region involves selecting the maximum lower limit and the minimum upper limit between r_{A1} and r_{A2}. In cases where the calculated upper bound is less than the lower bound, the overlap percentage is defined as zero, indicating no intersection between the ranges. Otherwise, the similarity metric is computed as the maximum of the overlap percentages relative to each original range (r_{A1} and r_{A2}). This computation can be formally expressed as:

$$\begin{aligned} &r_{A1} = [a_1, b_1] \\ &r_{A2} = [a_2, b_2] \\ &p_{overlap} = \max(0, \frac{\min(b_1, b_2) - \max(a_1, a_2)}{\min(b_1 - a_1, b_2 - a_2)}) \end{aligned} \tag{4}$$

To adopt the situation that many arguments belong to the same event in multi-documents, we propose a modified clustering algorithm capable of aggregating these temporally evolving arguments into coherent groups. As illustrated in Fig. 4, the overlap percentage similarity metric effectively captures the relationships between sequentially updated arguments. While arguments A_1 and A_3 may exhibit minimal direct overlap, both demonstrate substantial overlap with intermediate argument A_2, establishing a transitive relationship. This transitivity enables the grouping of A_1, A_2, and A_3 into one single cluster. To implement this connection scheme, we enhance the DBSCAN clustering algorithm which uses our overlap percentage metric for computing similarities between number-based arguments.

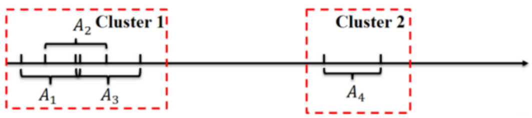


Figure 4. Number-based arguments grouping

The frequency analysis of clustered number-based arguments follows an analogous methodology to that employed for word-based arguments which is presented in Fig. 3 Following the clustering process, we compute individual argument frequencies and generate a comprehensive frequency distribution. This parallel treatment of number-based and word-based arguments enables a unified statistical framework for argument analysis.

Parameter configuration is crucial for optimizing DESCAN's AFV performance, particularly the maximum neighbor distance and minimal points threshold. For word-based arguments, empirical testing showed that a maximum neighbor distance of 0.1 provides optimal results, ensuring semantic coherence and allowing for word usage variations, while maintaining high precision in identifying equivalent arguments. Considering number-based arguments, we set the maximum neighbor distance to a 10% overlap threshold, allowing for proportional variations in numerical values while preserving meaningful relationships across different reporting contexts.

The minimal points parameter is set to 1 for both word-based and number-based arguments, maximizing recall of similar arguments including unique instances.

C. *Frequency Differentiation and Normalization*

Frequency analysis yields two key metrics: individual argument frequencies ($\mathbf{fre}_{A_i}$) and overall frequency list ($\mathbf{fre}_A$) encompassing all arguments. $\mathbf{fre}_{Ai}$ captures argument-specific occurrence patterns, and $\mathbf{fre}_A$ provides a global perspective across all documents. In our methodology, $\mathbf{fre}_{A_i}$ is arranged in descending order to facilitate analysis. For example, considering a case where the "Japan" argument cluster contains five elements ($\mathbf{fre}_{Jap} = 5$), and three distinct argument clusters exist for "Japan," "U.S.," and "Philippines," their respective frequencies form the ordered list $\mathbf{fre}_{\mathbf{A}}$[Jap,U.S,Phil] = [5, 4, 3].

The global frequency distribution $\mathbf{fre}_{\mathbf{A}}$ serves as a reference baseline against which individual frequencies fre_{A_i} can be evaluated for relative prominence. To capture these relative relationships, we compute the differential between fre_{Ai} and $\mathbf{fre}_{\mathbf{A}}$ to construct $\mathbf{diff}_{AFV}$. Subsequently, normalization is applied to mitigate scaling effects that could impact neural network training. The cross-document AFV feature, denoted as $\mathbf{f}_{\mathbf{AFV}}$, is formally constructed using fre_{Ai} and $\mathbf{fre}_{\mathbf{A}}$ according to the following formulation:

$$\begin{aligned} &\text{Differentiation:}\ diff_{AFV} = fre_{Ai} \otimes d_{AFV} - fre_A \\ &Normalization:\ f_{AFV} = \frac{diff_{AFV}}{\sum_i fre_{Ai}} \end{aligned} \tag{5}$$

The feature vectors $\mathbf{f}_{\mathbf{AFV}}$ and $\mathbf{diff}_{\mathbf{AFV}}$ are defined in $R^{d_{AFV}}$, where d_{AFV} represents the predetermined dimensionality of the argument frequency vector. The notation $\otimes$ denotes the vectorization operation that replicates fre_{A_i} d_{AFV} times to generate a vector of the specified dimension. In cases where the length of $\mathbf{fre}_{\mathbf{A}}$ is less than d_{AFV}, the frequency list is augmented through zero-padding to achieve the required dimension.

D. *Classification of AFV Integrated Feature-vector*

Our framework combines cross-document AFV features with single-document representations for argument-based event identification. We integrate neural network-generated single-document features, optimizing them jointly with a softmax classifier.

The concatenation scheme is illustrated in Fig. 5.

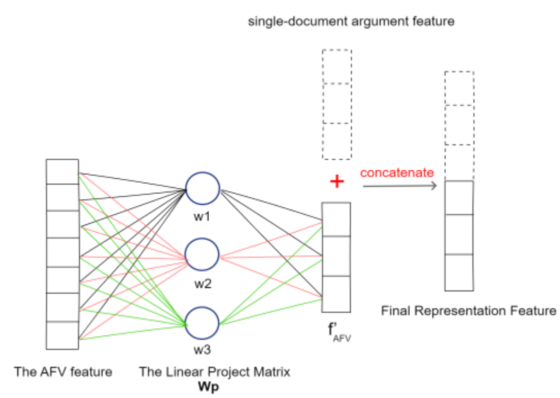


Figure 5. AFV feature and Single-document feature concatenation

The linear projection is applied to the AFV features ($\mathbf{f}_{\mathbf{AFV}}$) to ensure dimensional compatibility and optimal feature representation for neural network training. This linear transformation is implemented through a learnable weight matrix W, which maps the original AFV features to a new feature space while preserving the essential argumentative relationships. The linearly projected AFV feature $\mathbf{f}_{\mathbf{AFV}}'$ and single-document features are then concatenated along the feature dimension to construct a comprehensive argument representation. This integrated feature serves as the input to our classification module, leveraging both the argument-level interactions captured by AFV and the document-specific contextual information. The feature integration process can be formally expressed as follows:

$$\begin{aligned} \mathbf{f}_{\mathbf{AFV}}' &= \mathbf{W}_{\mathbf{p}} \mathbf{f}_{\mathbf{AFV}} \\ \grave{\mathbf{f}} &= [\mathbf{f}_{\mathbf{s}}, \mathbf{f}_{\mathbf{AFV}}'] \end{aligned} \tag{6}$$

$\mathbf{W}_{\mathbf{p}}$ is the linear project matrix in the neural networks. [**a**, **b**] represents the feature concatenation of **a** and **b**.

The feature vector **f**, comprising both cross-document and single-document representations, is subsequently processed through a softmax classifier, which can be formally expressed as:

$$\mathbf{o} = \mathbf{softmax}\left(\mathbf{W}_0\mathbf{f} + \mathbf{b}\right) \tag{7}$$

where **o** denotes the predicted probability distribution over the event categories. In addition, $\mathbf{W_o}$ and **b** serve as the softmax layer linear parameters.

IV. EXPERIMENTS

We evaluate the effectiveness of our proposed cross-document AFV feature using a self-collected dataset of earthquake-related news articles, which comprises 351 news articles covering 41 distinct earthquake events spanning from 2010 to 2019 from various news organizations.The neural network training applies Adam optimizer.

The event arguments are categorized into four semantic types: magnitude, location, death toll, and number of injuries. The dataset contains article-focused event arguments, totaling 1,295 instances ,and the other noise event arguments, accounting for 1,541 instances. In aggregate, the dataset encompasses 2,836 event arguments.

A. Experiment Results

To evaluate our proposed model's effectiveness, we conduct comparative analyses against three baseline models: two utilizing single-document features and one incorporating cross-document information.

a. dbRNN: Dependency bridge recurrent neural network (dbRNN) [7].

b. DMCNN: Dynamic Multi-pooling Convolutional Neural Network (DMCNN) [8].

c. Cross-docInf: Cross-document information features [9].

To assess the efficacy of our proposed cross-document AFV feature, we first evaluate its independent performance by constructing a model that exclusively utilizes the AFV feature for event argument identification in news articles. Subsequently, we develop an integrated feature framework (illustrated in Fig. 2) that enhances candidate argument representation by combining the cross-document AFV feature with established single-document features. The integrated approach is implemented in three distinct configurations: dbRNN+AFV, DMCNN+AFV, and Cross-docInf+AFV, each incorporating our AFV feature in the model. 10-times 10-folds cross-evaluation is applied in our experiments, the detailed results are presented as follows:

TABLE I. PERFORMACNCE ON COMPAIRED MODELS

Models	**Accuracy**	**Precision**	**Recall**	**F1**
dbRNN	0.881	0.835	0.924	0.876
DMCNN	0.865	0.820	0.902	0.859
Cross-docInf	0.875	0.841	0.897	0.868
AFV	0.888	0.874	0.879	0.876
DMCNN+**AFV**	0.923	0.905	0.928	0.917
Cross-docInf+**AFV**	0.916	0.898	0.920	0.909
dbRNN+**AFV**	**0.936**	**0.921**	**0.942**	**0.931**

As demonstrated in TABLE I, the experimental results for event argument identification in news articles reveal significant performance improvements. Models containing our proposed AFV feature consistently outperform all three baseline models, achieving a minimum improvement of 2.6% in test accuracy. Notably, the dbRNN+AFV configuration exhibits superior performance among all evaluated methods, empirically validating the effectiveness of our proposed cross-document AFV feature.

Further analysis of TABLE I, reveals that the AFV feature in isolation does not yield substantial performance improvements. However, detailed examination shows that it achieves higher precision, however, lower recall compared to the baseline models. This pattern indicates inherent limitations in both the cross-document AFV feature and single-document features when applied independently to event argument identification in news articles.

The precision-recall characteristics of our AFV feature suggest strong performance in positive prediction but relatively weaker performance in negative prediction. This behavior can be attributed to two key factors: low-frequency arguments are unlikely to be positive (news-focused event), while high-frequency arguments may still be negative (news other event) in certain cases. Conversely, models utilizing single-document features exhibit a bias toward positive predictions, primarily because candidate arguments from different events often share similar contextual and semantic properties within individual documents.

B. Model Analysis

To conduct a comprehensive analysis of our proposed cross-document AFV feature, we performed two evaluation experiments: complementary test and argument categories analysis. All experimental parameters were maintained consistent with those described in Section IV.A.

Complementary Test

The complementary characteristics provide distinct information sources that address mutual blind spots. To evaluate the complementarity between existing single-document features and our proposed cross-document AFV feature, we conduct statistical analyses using disagreement measure and Q-statistics. TABLE II presents the complementarity analysis results, where the test sets comprise misclassified instances from the three baseline models.

TABLE II. ARGUMENT FEATURES COMPLEMENTARITY

Evaluated Set	**dis**	**Q**
SDMCNN	0.610	-0.707
SCross-docInf	0.513	-0.549
SdbRNN	0.769	-0.874

TABLE II illustrates the complementarity between our proposed AFV feature and three baseline models' features. To evaluate this complementarity, we conducted experiments using the wrong-prediction test-sets from the baseline models (SDMCNN, SCross-docInf, and SdbRNN). We applied the standalone AFV feature model to predict samples from these baseline models and calculated the corresponding disagreement measures and Q-statistics.

The AFV feature shows strong complementarity with single-document features, particularly dbRNN (disagreement measure: 0.769, Q-statistics: -0.874), while displaying lower complementarity with Cross-docInf. Though AFV and Cross-docInf predictions diverge in 50% of cases, this stems from Cross-docInf's reliance on within-sentence information. This highlights how traditional cross-document features fail to fully leverage inter-document relationships, underlining AFV's unique contribution to integrated feature models.

Argument Categories Analysis

This section examines the differential effects of AFV features across distinct semantic categories. Within our self-collected news corpus, four event-specific semantic categories were identified: location, magnitude, fatalities, and casualties. The disagreement measures across these argument classes were systematically evaluated, with results illustrated in Fig. 6.

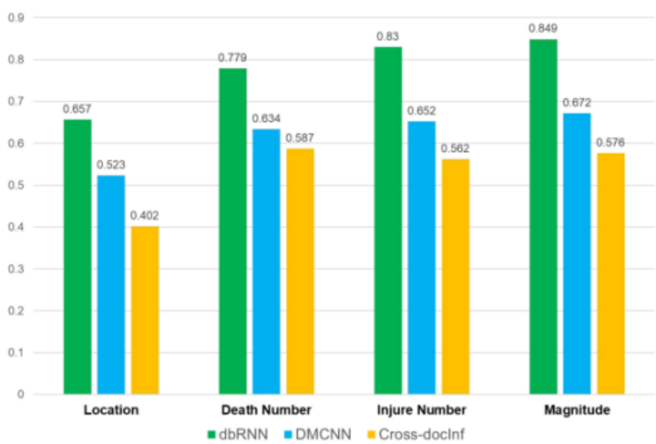


Figure 6. Disagreement on different data categories

Fig. 6 shows the AFV feature's varying performance across semantic classes in the earthquake event argument. It performs particularly well with quantitative arguments like fatalities, casualties, and magnitude measurements, where its ability to cross-validate numerical values across multiple documents enhances extraction accuracy.

The AFV feature shows a limited impact on location-based arguments, as earthquake event triggers in news articles typically embed location information directly (e.g., 'Hualien earthquake', 'Sichuan earthquake'). This allows single-document models to extract locations accurately using local context, without needing cross-document validation.

The performance gap between quantitative and location arguments demonstrates the complementary roles of single and cross-document features. While AFV excels at cross-validating numerical values, single-document features better handle spatial information that remains consistent within individual reports. This suggests optimal extraction systems should selectively apply both feature types based on argument semantics.

V. Conclusion

This paper addresses event argument identification in articles, especially for continuous reports like news. Our research reveals the event-driven nature of news reporting, where multiple publications typically provide sustained coverage of singular events, resulting in significant cross-document interactions. We propose a novel algorithm utilizing cross-document argument frequency statistics, represented as AFV feature. By integrating AFV with single-document features, our classifier substantially outperforms baseline models, leveraging the complementary insights from cross-document and single-document approaches.

Acknowledgment

This work is supported by China Academy of Industrial Internet, Nanyang Technological University.

References

[1] Wang, Yu, Nedim Lipka, Ryan A. Rossi, Alexa Siu, Ruiyi Zhang, and Tyler Derr. "Knowledge graph prompting for multi-document question answering." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 17, pp. 19206-19214. 2024.

[2] Lu, Siyu, Mingzhe Liu, Lirong Yin, Zhengtong Yin, Xuan Liu, and Wenfeng Zheng. "The multi-modal fusion in visual question answering: a review of attention mechanisms." PeerJ Computer Science 9 (2023): e1400.

[3] Omar, Reham, Omij Mangukiya, Panos Kalnis, and Essam Mansour. "Chatgpt versus traditional question answering for knowledge graphs: Current status and future directions towards knowledge graph chatbots." arXiv preprint arXiv:2302.06466 (2023).

[4] Xu, Derong, Wei Chen, Wenjun Peng, Chao Zhang, Tong Xu, Xiangyu Zhao, Xian Wu, Yefeng Zheng, Yang Wang, and Enhong Chen. "Large language models for generative information extraction: A survey." Frontiers of Computer Science 18, no. 6 (2024): 186357.

[5] Landolsi MY, Hlaoua L, Ben Romdhane L. Information extraction from electronic medical documents: state of the art and future research directions. Knowledge and Information Systems. 2023 Feb;65(2):463-516.

[6] Hu, Zhilei, Zixuan Li, Xiaolong Jin, Long Bai, Saiping Guan, Jiafeng Guo, and Xueqi Cheng. "Semantic structure enhanced event causality identification." arXiv preprint arXiv:2305.12792 (2023).

[7] Lei Sha, Feng Qian, Baobao Chang, and Zhifang Sui. Jointly extracting event triggers and arguments by dependency-bridge rnn and tensor-based argument interaction. In Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[8] Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. Event extraction via dynamic multi-pooling convolutional neural networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 167–176, 2015.

[9] Heng Ji and Ralph Grishman. Refining event extraction through cross document inference. In Proceedings of ACL-08: Hlt, pages 254–262, 2008.

[10] Schmidheiny, Kurt, and Sebastian Siegloch. "On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization." Journal of Applied Econometrics 38, no. 5 (2023): 695-713.

[11] Santana, Brenda, Ricardo Campos, Evelin Amorim, Alípio Jorge, Purificação Silvano, and Sérgio Nunes. "A survey on narrative extraction from textual data." Artificial Intelligence Review 56, no. 8 (2023): 8393-8435.

[12] Li, Zhaoning, Qi Li, Xiaotian Zou, and Jiangtao Ren. "Causality extraction based on self-attentive BiLSTM-CRF with transferred embeddings." Neurocomputing 423 (2021): 207-219.

[13] Lu, Di, Shihao Ran, Joel Tetreault, and Alejandro Jaimes. "Event extraction as question generation and answering." arXiv preprint arXiv:2307.05567 (2023).

[14] Liu, Wanlong, Shaohuan Cheng, Dingyi Zeng, and Hong Qu. "Enhancing document-level event argument extraction with contextual clues and role relevance." arXiv preprint arXiv:2310.05991 (2023).

[15] Zhang, Mengxi, and Honghui Chen. "Document-Level Event Argument Extraction with Sparse Representation Attention." Mathematics 12, no. 17 (2024): 2636.

[16] Li, Qian, Shu Guo, Jia Wu, Jianxin Li, Jiawei Sheng, Hao Peng, and Lihong Wang. "Event extraction by associating event types and argument roles." IEEE Transactions on Big Data (2023).

[17] Ding, Bowen, Qingkai Min, Shengkun Ma, Yingjie Li, Linyi Yang, and Yue Zhang. "A Rationale-centric Counterfactual Data Augmentation Method for Cross-Document Event Coreference Resolution." arXiv preprint arXiv:2404.01921 (2024).

[18] Milbauer, Jeremiah, Ziqi Ding, Zhijin Wu, and Tongshuang Wu. "From nuisance to news sense: Augmenting the news with cross-document evidence and context." arXiv preprint arXiv:2310.04592 (2023).

[19] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global vectors for word representation." In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532-1543. 2014.