Earth Science Informatics (2022) 15:1417–1425
https://doi.org/10.1007/s12145-022-00847-y

RESEARCH ARTICLE

Character embedding-based Bi-LSTM for Zircon similarity calculation with clustering

Xiangben Hu[1] · Zhichen Hu[1] · Jielin Jiang[1] · Weiwei Xue[2] · Xiumian Hu[2] · Xiaolong Xu[1]

Received: 21 December 2021 / Accepted: 20 July 2022 / Published online: 4 August 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Similarity calculations for zircons are vital to topical issues in sedimentology, such as provenance analysis, dating of sediment and identification of geotectonic effects. In general, zircon data is stored in a table where each column represents a key-value pair. According to the semantics of the keys, multiple tables are merged to extract data for analyzing the variability of single feature. However, there are conflicts between the different indicators due to sedimentation, which leads to inaccuracy of similarity. Moreover, unknown and semantically ambiguous keys are not recognized by the knowledge base, which results in the inefficiency of aggregating key-value pairs. Therefore, this paper proposed a Fast Much zircon (FM-zircon) framework that combines natural language processing (NLP) and multidimensional scaling (MDS) for calculating the similarity of zircons. First, NLP classifies keys by extracting semantic features. After the key-value pairs with the same key are fused, MDS is implemented to calculate multiple features. Ultimately, the results are represented in a visual representation To evaluate the performance, experiments were performed with zircon tables, that showed the good performance of FM-zircon.

Keywords Similarity of Zircons · Clustering · NLP · Visualization of Zircons

Introduction

Sedimentology is a sub-discipline of geology, providing theoretical support for oil exploration. Mineral sources are analyzed by chemical feature to infer the evolutionary history of the Earth (Bruand et al. 2014). In general, minerals are susceptible to weathering leading to chemical elements decay, which prevents accurate inference of sediment origin (Wang et al. 2016). Considering the wide distribution and stability of zircon. Zircon dating is the standard method for sediment source origin (Watts et al. 2016).

Typically, zircon data is stored in tables where each column is a key-value pair. Thereby, key-value pairs from multiple tables in the study area were integrated into one image based on the semantics of the key, which was visualized the differences in the chemical features distribution with different formations. However, the variability is difficult to identify because a single formation contains hundreds of samples (Wilson et al. 2017). Therefore, it is essential for the quantitative analysis of zircon chemical features.

The chemical features of zircon are defined by sedimentologists where a feature contains multiple isotopes. Machine learning is performed to calculate distance of zircon data, thus analyzing a single chemical features similarities (Bindeman and Melnik 2016). Subsequently, the similarities are ranked to ultimately improve the efficiency of zircon data processing. However, calculating single chemical feature similarity leads to conflict conclusions (Van Lankvelt et al. 2016).

Moreover, due to the different natural language descriptions of a key (e.g., U-pb and U-pbΓ are in semantic agreement), it is tough to recognize the semantics of keys. Therefore, most existing works are unable to recognize

✉ Xiaolong Xu
xlxu@ieee.org

Zhichen Hu
huzhichen@nuist.edu.com

Weiwei Xue
xww@smail.nju.edu.cn

Xiumian Hu
huxm@nju.edu.cn

[1] School of Computer Science, Nanjing University of Information Science and Technology, No.219, Ningliu Road, Nanjing 21000, Jiangsu, China

[2] School of Earth Sciences, Nanjing University, No.163, Xianlin Road, Nanjing 21000, Jiangsu, China

unknown keys that are not retrieved from the knowledge base (Chamiran et al. 2020).

NLP base on word embedding method computes the contextual word frequency of one word in a sentence to represent the semantics, which greatly improves the accuracy of table key recognition (Eslahi et al. 2020). Millions of texts are trained as pre-trained models for accurate semantic representation (Yurin et al. 2021). However, the pre-trained model such as Bidirectional Encoder Representations from Transformers (BERT) performed with low accuracy in the zircon table because of few labels for the table keys resulted in data sparsity (Yan et al. 2020). Furthermore, in the zircon table, some keys are not part of the real-world words (e.g., Ageγ), which results in word embedding failing to accurately present semantics.

To solve these problems, a new framework is proposed named FM-zircon. First, character embedding extracts the semantics of the key to solve the problem of data sparsity. Then, Bi-directional Long Short-Term Memory(Bi-LSTM) and softmax classify the keys according to semantics (Hochreiter and Schmidhuber 1997; Rao et al. 2019). After the key-value pairs are aggregated, a similarity is used to calculate the hybrid features of multiple chemical features to solve conflicting conclusions. The main contributions of this article are as follows:

A hybrid chemical features similarity calculation method is proposed to solve the problem of conflicting conclusions.

Character embedding is used to extract the semantic features of the keys in the zircon table to improve recognition accuracy.

The rest of this paper is organized as follows. Related work is reviewed in Section 2. Section 3 introduces the FM-zircon. Section 4 gives the experimental evaluation and comparative analysis. Conclusions are drawn in Section 5.

Related work

Due to the stability and wide distribution of zircons, zircon dating is the standard method for hot issues in sedimentology (Nemchin and Pidgeon 1997). The similarity of zircons is a variable that measures the variability of the chemical features in different formations. Thousands of zircons tables were aggregated to calculate the similarity. Many existing studies combine zircon and machine learning. Delaigle et al. (2008) introduced U-Pb element age distribution density probability to calculate similarity, which improves the accuracy of zircon data analysis. Saylor et al. (2009) proposed a method based on the Kolmogorov-Smirnov test (K-S test) for calculating age distribution variability by machine learning. However, these methods were unable to analyze zircon data quantitatively.

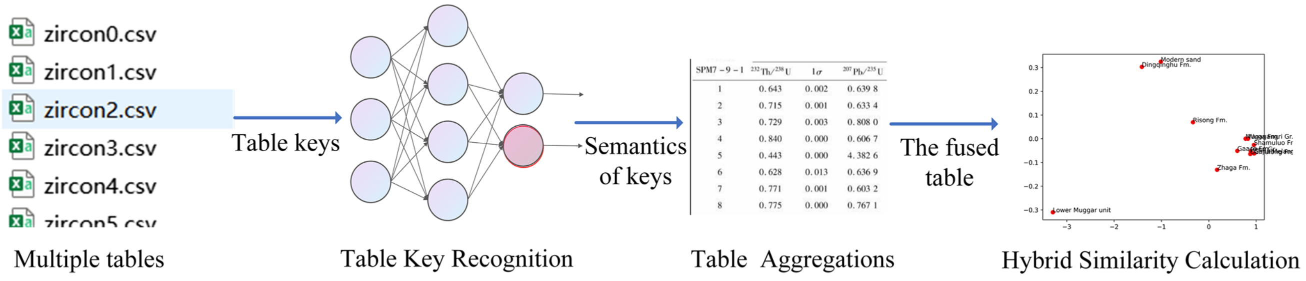
Vermeesch (2012) proposed a method to reduce the dimension of K-S test results with Multidimensional scaling(MDS), which improves the efficiency of calculating the similarity. Sharman and Malkowski (2020) proposed tool with all the chemical characteristics analysis function and the final conclusion is observed by manual. However, there are conflicts between the different chemical features due to sedimentation, which leads to inaccuracy of similarity (Yongvanich et al. 2019). In this paper, a hybrid similarity calculation strategy is proposed to relieve the conflicts between different chemical features by fusing multiple chemical features.

Moreover, zircon data is stored in a table where each column is a key-value pair. Thousands of tables were aggregated according to the semantics of the keys to calculate the similarity (Ahmed 2008). Due to multiple natural language descriptions link to a key in the table, it is tough to recognize keys of tables by dictionary (Xu et al. 2010). Liu et al. (2005) built a knowledge base containing thousands of natural language descriptions groups by semantics to recognize keys. Julthep et al. (Nandakwang and Chongstitvatana 2016) linked Wikipedia data to web tables for classifying the keys with the Term Frequency Inverse Document Frequency(TF-IDF) algorithm. These methods fail to recognize unknown words, which were not included in the knowledge base.

Recently, deep learning is widely applied in the field of table key recognition because it extracts semantics accurately. Shaik et al. (2021) proposed a method base on knowledge graph where graphical neural networks were trained on a large corpus for extracting semantic links between keys and words. Luzuriaga et al. (2021) trained in Wikipedia to obtain links between words, which greatly improves the accuracy of table key recognition. Berant et al. (2019) classified keys by extracting semantics features of table contexts. These works are highly accurate on large corpora but suffer from data sparsity on small corpora. This paper uses character embedding to extract the semantic features of the keys to address this problem.

Design of FM-Zircon

In this section, FM-zircon, a novel framework for zircons similarity calculations, is introduced. Figure 1 illustrates the structure of FM-Zircon. Firstly, according to the character sequence of the key in the knowledge base, NLP classifies input according to key, aiming to extract dense semantic features of input for detecting unknown words. Furthermore, columns with the same semantic are merged. Finally, a similarity computation method is



SPM7 − 9 − 1	$^{232}Th/^{238}U$	1σ	$^{207}Pb/^{235}U$
1	0. 643	0. 002	0. 639 8
2	0. 715	0. 001	0. 633 4
3	0. 729	0. 003	0. 808 0
4	0. 840	0. 000	0. 606 7
5	0. 443	0. 000	4. 382 6
6	0. 628	0. 013	0. 636 9
7	0. 771	0. 001	0. 603 2
8	0. 775	0. 000	0. 767 1

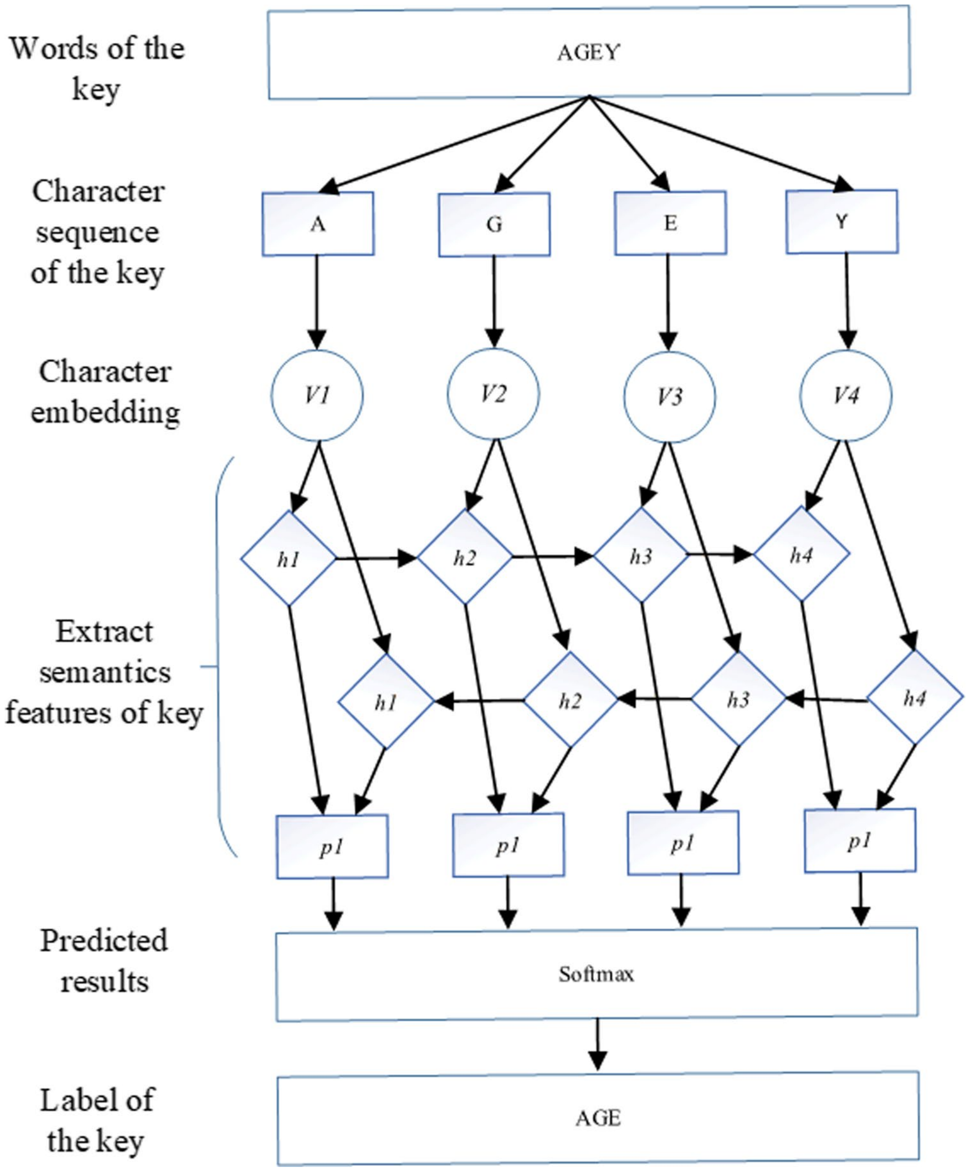
Fig. 1 Structure of FM-zircon

proposed for hybrid zircon chemical features based on MDS and visualize the result.

Table key recognition

The overall operation process of table key recognition is shown in Fig. 2. Firstly, according to the order of the key in

Fig. 2 This is the process of table key i recognition, where ageγ is labeled age



the knowledge base, character embedding is generated for modeling the semantics feature of character. After that, Bi-LSTM is used to help extract semantic features of the character sequence. Finally, softmax classifies input according to key, aiming to extract dense semantic features of input to detect unknown words.

Character embedding of keys

The table key recognition is essentially a classification problem, where according to methods like embedding achieves great results. Word embedding is efficient for indicating the word object coordinates in semantic space but unfit for sparse data. The number of keys in the zircon table is too small (a few hundred), which results in sparse data for word embedding, it is not possible to represent the semantics accurately. The character embedding statistics table key in the character sequence relationship (43 in total), which greatly relieves the sparsity of the data. In light of this, we use character embedding, a lightweight method, which is used for extracting the semantics feature of character. The details of character embedding are described as follows. Let $Ch = \left\{Ch_1, Ch_2, ..., Ch_M\right\}$ represent the one-hot code set of characters, $KeyBase = \left\{KeyBase_1.KeyBase_2.....KeyBase_N\right\}$ represent the key set of zircon tables, $Chem = \left\{Chem_1, Chem_2, ..., Chem_M\right\}$ represent the character embedding set of characters, C represents window size of the context, respectively, let $Ch_i = \left\{Ch_{i1}, Ch_{i2}, ..., Ch_{iL}\right\}$ denote character set of i^{th} key $KeyBase_i$.

First, one-hot code initializies the weights of the model. Then, a full connection layer is used to extract the semantics of characters. Finally, a softmax layer is normalized result and the semantics features of all the characters in $KeyBase_i$ is given by

$$p(ch_{io}\|ch_{ij}) = \frac{e^{Chem_{io}*Chem_{ij}}}{\sum_j e^{Chem_{iL}*Chem_{ij}}}, \tag{1}$$

where L denotes the length of $KeyBase_i$.

In the training stage, entropy loss function calculates distribution difference of the semantics. The formula is given by

$$Loss_{em} = \sum_j p(ch_{io}\|ch_{ij})log_2(chem_o)). \tag{2}$$

Bi-LSTM for extracting sematics

After encoding characters, it is crucial to extract the potential link between character embedding and key. In recent years, Recurrent Neural Networks(RNN) have been widely applied in various tasks of NLP due to the ability to extract correlations between sequences. However, some sequences of table keys with lengths greater than 20 lead to network vanishing gradient and exploding gradient. Compared with RNN, LSTM relieves the gradient elimination due to filtering the useless information in the previous text. Nonetheless, the correlation between language sequences is bidirectional but LSTM extracts the features of unidirectional text (Istiake Sunny et al. 2020). Bi-LSTM is composed of two LSTM blocks with opposite directions which extracts semantic features in the bidirectional. Thus, Bi-LSTM extracts the feature of characters sequence from a key in this paper. For a character embedding, The output of the forward LSTM is defined by

$$outf_i = LSTM(Chem_i, Chem_{i+1}), \tag{3}$$

where $outf_i$ is the semantic features of the previous text. To extract the semantic features of the later text, a backward LSTM is utilized. The formula is defined by

$$outb_i = LSTM(Chem_i, Chem_{i-1}), \tag{4}$$

The final semantic features out_i are defined by

$$out_i = outb_i + outf_i, \tag{5}$$

Softmax for classification keys

Three fully connected layers are used to normalizing the output of Bi-LSTM. Finally, softmax gains the probability of whether the word is a key or not. The formula of softmax is given by

$$P = \frac{e^{out_i}}{\sum_j e^{out_j}}, \tag{6}$$

where P is the list predicted of value.

Table aggregation

The process of table aggregation is shown in Algorithm 1. In Algorithm 1, l represents the number of tables, $table[i][0]$ is the first row of $table[i]$ and r means the number of columns of $table[i][0]$. First, for each table, the first row is searched to obtain a key-list. Then, FM zircon judges whether the semantics of table keys are retrieved from the knowledge base. If so, the key is matched; otherwise, the network classifies keys larger than the threshold, which is set to 0.9. Finally, all key-value pairs are grouped by key with the same semantic.

```
$input : multiple\ tables$
$i \Leftarrow 1$
while $i \leq l$ do
    $Row \Leftarrow table[i][0]$
    $j \Leftarrow 1$
    while $j \leq r$ do
        if $row[j]$ is in knowledge base then
            Match
            Merge
        else
            $a \Leftarrow TableKeyRecognition$
            if $a \leq 0.9$ then
                Match
                Merge
            end if
        end if
    end while
end while
$output : normalization\ table$
```

Hybrid similarity calculation

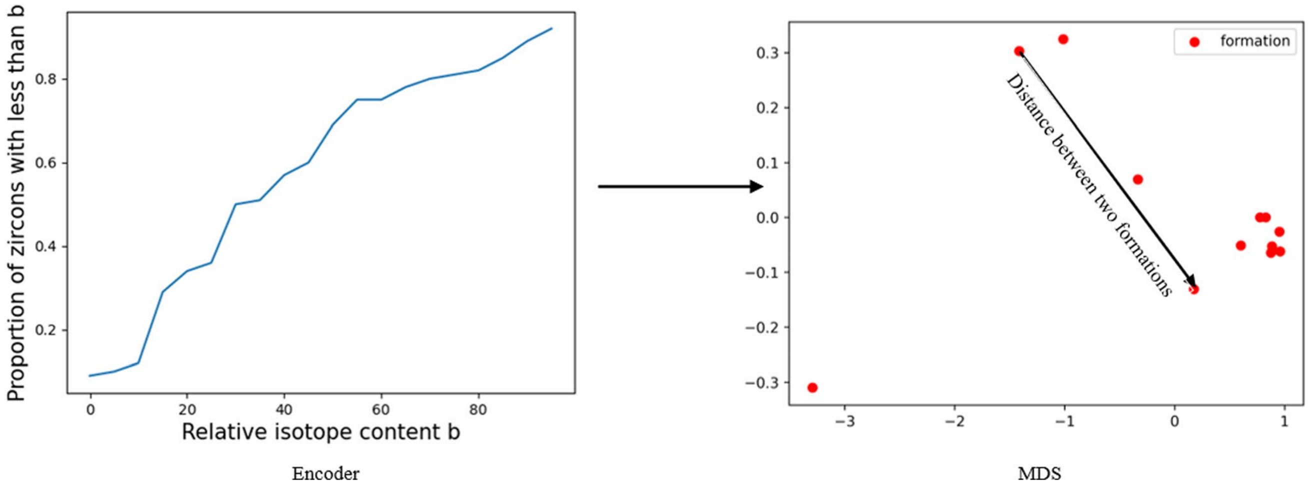
Due to weathering, single chemical features are not adequate for dating sediment. For accurate chemical features, a hybrid similarity calculation is proposed as Fig. 3 shown. First, cumulative distribution is used to encode isotope features. Afterward, all the feature vectors are summed in proportion of 1 :1. Finally, MDS calculates and visualizes the difference between two formations.

Encoder of isotope

First, all isotope data are normalized between 0 to 100. Let $Group = \left\{group_1, group_2, ..., group_M\right\}$ represent the formation set, $groupf_{k,b}$ represent the two dimensional feature matrix of $group_i$, k indicates the isotope type, then $groupf_{k,b}$ is defined by

$$groupf_{k,b} = \frac{Content}{L_{num}}, \tag{7}$$

where $Content$ is the total number of zircon less than b and L_{num} is the total number of zircon.



Encoder

MDS

Fig. 3 The structure of hybrid similarity calculation

Table 1 Labels of table key recognition

id	Label	Semantic
0	SampleId	Id of zircon sample
1	CoreId	Id of zircon
2	FormationId	Id of stratum
3	Position	Study area
4	Era	Era of zircon
5	U-pb Age	U-pb age for zircon
6	Hf Age	Hf age for zircon
7	Position	Study area
8	Era	Era of zircon
9	Pb	Pb content
10	Th	Th content
11	U	U content
12	Hf	Hf content
13	U/Th	Ratio of u to th
14	207Pb/206Pb	Ratio of 207Pb to 206Pb
15	206Pb/238U	Ratio of 206Pb to 238Pb

Afterwards, *groupf* is fused in the k dimension into a final feature vector. The formula is given by

$$V(group_i) = \sum_{k=1}^{K} groupf_k, \tag{8}$$

where K is the sum of isotope.

MDS for calculating Zircons similarity

The similarity calculation is intrinsically a dimensionality reduction problem, where a method based on MDS arrive great results. According to (8), the difference between $group_i$ and $group_j$ is given by

$$dense_{i,j} = ||V(group_i) - V(group_j)||, \tag{9}$$

After normalization, (9) can be further expressed as

$$p_{i,j} = 1 - \frac{dense_{longest}}{dense_{i,j}}, \tag{10}$$

where $p_{i,j}$ is the similarity between $group_i$ and $group_j$. $dense_{longest}$ is the longest distance between two formations in $Group$.

Results

In this section, experiments are performed to prove the effectiveness of FM-zircon. First, the dataset is described and then the performance of table key recognition is studied. What's more, a method to evaluate the performance between hybrid similarity and single similarity is proposed. Finally, zircon similarity and spatial information was visualized.

Datasets

We manually extracted 100 zircon tables from the sedimentological literature. The total number of keys are more than 300 and there are 16 kinds of labels. We choose 270 keys as the train set and 30 as the test set. The labels are shown in Table 1.

Furthermore, we manually extracted three common chemical features in sedimentology from 50 papers. These features are the U-pb feature, the Hf feature and the feature of the Clastic Composition (CC), respectively. Examples of datasets are listed in Table 2.

In Table 2, ID represents the universal unique identifier of the data set, paperid marks the extracted sedimentological literature, and formation1 and formation2 denote the two groups two groups being compared in the literature. HF,

Table 2 Labels of similarity calculation

Id	Paperid	Formation1	Foramtion2	U-pb	CC	Hf	Result
0	0	Duba	Dingqinghu	True	False	True	True
1	0	Mugagangri Gr.	Shamuluo Fm.	True	False	True	True

Table 3 Result for table key recognition

Model	Accuracy	Recall	F1	pre-train
Method base on knowledge base Van Lankvelt et al. (2016)	100 %	65.3 %	79.0	no
Word embbeding+Bi-LSTM+softmax Guo et al. (2020)	85.7 %	78.0	81.6 %	yes
BERT+softmax Guo et al. (2021)	75.3 %	69.8 %	72.4	yes
BERT+Bi-LSTM+softmax Liu and Xie (2021)	72.7 %	64.5 %	68.3	no
Character embedding+Bi-LSTM+softmax	89.4 %	77.8 %	83.1	no

Pre-train represents whether pre-train was performed

UPB and CC represent three chemical characteristics of zircon. Hf,Upb, CC represent the three chemical characteristics of zircon. Result denotes the final conclusion of the paper. True indicates that the two groups are similar in this chemical characteristic dimension, and false is just the opposite. For instance, a record with id 0 means that in paper 0, group Duba and group Dingqinghu are similar in the dimension of U-pb.

Experiment for table key recognition

Contrast test of table key recognition was performed for different models, results are listed in Table 3. We compare our Character embedding+Bi-LSTM+softmax method with the following four table key recognition methods: method base on knowledge base, Word embbeding+Bi-LSTM+softmax, BERT+softmax and BERT+Bi-LSTM+softmax.

From Table 3, it can be seen that FM-zircon greatly improves the recall rate compared to the method based on the knowledge base. The reason is that character embedding extracts accurate semantic features to recognize unknown keys. Furthermore, the recall and accuracy of the method based on the pre-trained model is low because of data sparsity in small sample. Moreover, compared with word embedding, character embedding is more accurate due to alleviating data sparsity. It indicates that FM-zircon mitigates data sparsity.

Table 4 Result for similarity calculation

Group	Chemical feature	Accuracy
1	U-Pb	72.4 %
2	Hf	65 %
3	CC	68.8 %
4	Hf+CC	76.6 %
5	U-Pb+Hf	80 %
6	CC+U-Pb	78.8 %
7	U-Pb+CC+Hf	83.6 %

Experiment for similarity calculation

A contrast test of similarity calculation was performed for different chemical features. The similarity of chemical features between the two groups was calculated and considered similar if it was greater than 0.75. Finally, the results were compared with the descriptions in the sedimentological literature and the accuracy was calculated. The result is listed in Table 4.

As shown in Table 4, compared to individual chemical feature similarity, hybrid similarity improves accuracy. For instance, there was a 5% increase in hybrid similarity in group 6 compared to group 1. In addition, group 7, which merges the three chemical features is the highest in terms of accuracy. The chemical information of zircons is missing due to weathering. Therefore, it leads to inaccurate

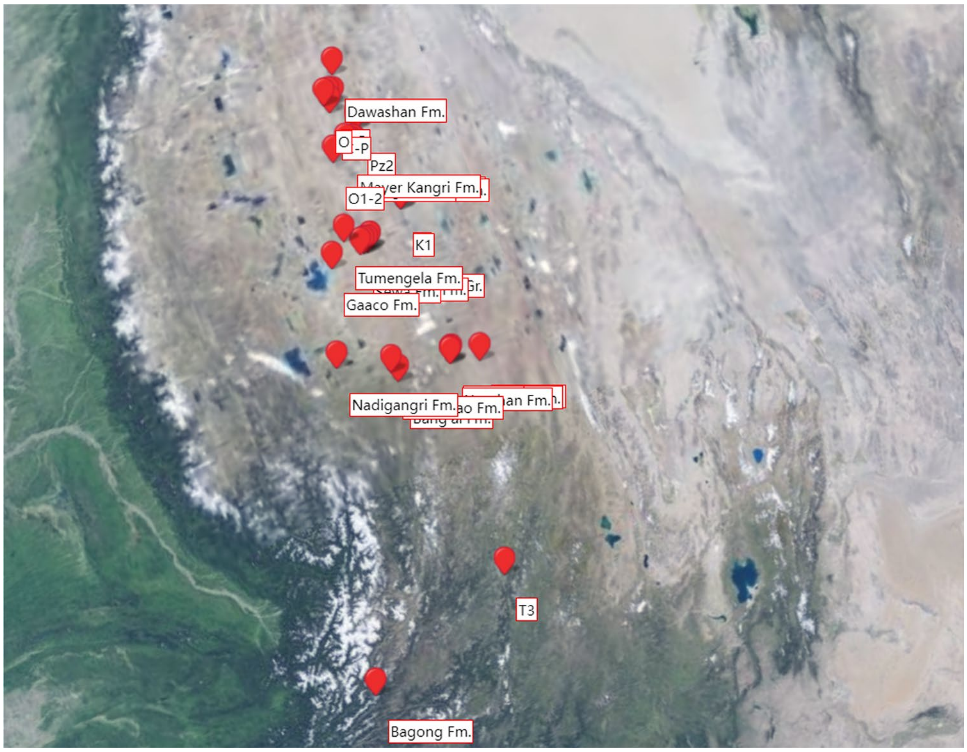


Fig. 4 Zircon data visualization

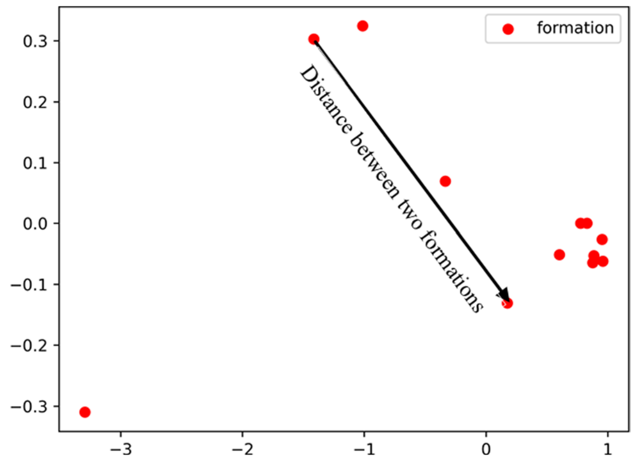


Fig. 5 Zircon similarity visualization

calculation of a single chemical feature. The reason is the conflict between different chemical features. Hybrid similarity extracts multiple chemical features to resolve conflicts and improves the accuracy of zircon similarity.

Visualization

As illustrated in Fig. 4, all zircon data was aggregated and mapped to BaiduMap based on formations locations, where each mark is a formation.

The similarity of zircons is shown in Fig. 5. The distance between two points represents the similarity with closer distance indicating higher similarity.

Conclusion

FM-zircon was designed to rapidly extract zircon age features. NLP extracts semantic features to classify table keys. In addition, the hybrid similarity is calculated to alleviate conflicts between different isotopes, which is an important component of our FM-zircon. In the future, we will work on extracting table structure by table contextual features with few shot learning.

Acknowledgements This research is supported by the National key research and development program of China (International Technology Cooperation Project) under grant no.2021YFE0104400, the National Natural Science Foundation of China under grant no.41975183.

Declarations

Conflicts of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Ahmed E (2008) Resource capability discovery and description management system for bioinformatics data and service integration - an experiment with gene regulatory networks. In: 2008 11th international conference on computer and information technology. pp 56–61. https://doi.org/10.1109/ICCITECHN.2008.4802991

Berant J, Deutch D, Globerson A, Milo T, Wolfson T (2019) Explaining queries over web tables to non-experts. In: 2019 IEEE 35th international conference on data engineering (ICDE). pp 1570–1573. https://doi.org/10.1109/ICDE.2019.00144

Bindeman IN, Melnik OE (2016) Zircon survival, rebirth and recycling during crustal melting, magma crystallization, and mixing based on numerical modelling. Journal of Petrology 57(3):437–460. https://doi.org/10.1093/petrology/egw013

Bruand E, Storey C, Fowler M (2014) Accessory mineral chemistry of high Ba-Sr granites from Northern Scotland: Constraints on petrogenesis and records of whole-rock signature. Journal of Petrology 55(8):1619–1651. https://doi.org/10.1093/petrology/egu037

Chamiran K, Rukshan A, Thayasivam U (2020) Automating web table columns to knowledge base mapping using translation embedding. In: 2020 IEEE 14th international conference on semantic computing (ICSC). pp 150–153. https://doi.org/10.1109/ICSC.2020.00029

Delaigle A, Hall P, Meister A (2008) On deconvolution with repeated measurements. The Annals of Statistics 36:665–685

Eslahi Y, Bhardwaj A, Rosso P, Stockinger K, Cudré-Mauroux PP (2020) Annotating Web tables through knowledge bases: A context-based approach. In: 2020 7th swiss conference on data science (SDS). pp 29–34. https://doi.org/10.1109/SDS49233.2020.00013

Guo S, Fang C, Lin J, Wang Z (2020) A configurable FPGA accelerator of Bi-LSTM inference with structured sparsity. In: 2020 IEEE 33rd international system-on-chip conference (SOCC). pp 174–179. https://doi.org/10.1109/SOCC49529.2020.9524784

Guo H, Liu T, Liu F, Li Y, Hu W (2021) Chinese text classification model based on bert and capsule network structure. In: 2021 7th IEEE Intl conference on big data security on cloud (BigDataSecurity), IEEE Intl conference on high performance and smart computing, (HPSC) and IEEE Intl conference on intelligent data and security (IDS). pp 105–110. https://doi.org/10.1109/BigDataSecurityHPSCIDS52275.2021.00029

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Computation 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Istiake Sunny MA, Maswood MMS, Alharbi AG (2020) Deep learning-based stock price prediction using LSTM and Bi-Directional LSTM model. In: 2020 2nd novel intelligent and leading emerging sciences conference (NILES). pp 87–92. https://doi.org/10.1109/NILES50944.2020.9257950

Liu Q, Gong Y (2005) The application of virtual strategy table base in regional stability control. In: 2005 IEEE/PES transmission and distribution conference and exposition: Asia and pacific. pp 1–5. https://doi.org/10.1109/TDC.2005.1547088

Liu H, Xie L (2021) Research on sarcasm detection of news headlines based on Bert-LSTM. In: 2021 IEEE international conference on emergency science and information technology (ICESIT). pp 89–92. https://doi.org/10.1109/ICESIT53460.2021.9696851

Luzuriaga J, Munoz E, Rosales-Mendez H, Hogan A (2021) Merging web tables for relation extraction with knowledge graphs. In: IEEE transactions on knowledge and data engineering. https://doi.org/10.1109/TKDE.2021.3101479

Nandakwang J, Chongstitvatana P (2016) Extract semantic web knowledge from Wikipedia tables and lists. In: 2016 8th international conference on knowledge and smart technology (KST). pp 108–113. https://doi.org/10.1109/KST.2016.7440520

Nemchin AA, Pidgeon RT (1997) Evolution of the darling range batholith, Yilgarn Craton, Western Australia: a SHRIMP Zircon Study. Journal of Petrology 38(5):625–649. https://doi.org/10.1093/petroj/38.5.625

Rao Q, Yu B, He K, Feng B (2019) Regularization and iterative initialization of softmax for fast training of convolutional neural networks. 2019 international joint conference on neural networks (IJCNN). pp. 1–8. https://doi.org/10.1109/IJCNN.2019.8852459

Saylor JE, Quade J, Dettman DL (2009) The late miocene through present paleoelevation history of Southwestern Tibet. American Journal of Science 309:1–42

Shaik Z, Ilievski F, Morstatter F (2021) Analyzing race and citizenship bias in Wikidata. In: 2021 IEEE 18th international conference on mobile Ad Hoc and smart systems (MASS). pp 665–666. https://doi.org/10.1109/MASS52906.2021.00099

Sharman GR, Malkowski MA (2020) Needles in a haystack: Detrital zircon UPb ages and the maximum depositional age of modern global sediment. Earth-Sci Rev 203. https://doi.org/10.1016/j.earscirev.2020.103109

Van Lankvelt A, Schneider DA, Biczok J, McFarlane CRM, Hattori K (2016) Decoding Zircon geochronology of igneous and alteration events based on chemical and microstructural features: A study from the western superior province. Canada. Journal of Petrology 57(7):1309–1334. https://doi.org/10.1093/petrology/egw041

Vermeesch P (2012) On the Visualisation of detrital age distributions. Chemical Geology 312:190–194. https://doi.org/10.1016/j.chemgeo.2012.04.021

Wang M, Nebel O, Wang CY (2016) The flaw in the crustal 'Zircon archive': Mixed Hf isotope signatures record progressive contamination of late-stage liquid in mafic-ultramafic layered intrusions. Journal of Petrology 57(1):27–52. https://doi.org/10.1093/petrology/egv072

Watts KE, John DA, Colgan JP, Henry CD, Bindeman IN, Schmitt AK (2016) Probing the volcanic-plutonic connection and the genesis of crystal-rich rhyolite in a deeply dissected supervolcano in the Nevada Great Basin: Source of the late eocene caetano tuff. Journal of Petrology 57(8):1599–1644. https://doi.org/10.1093/petrology/egw051

Wilson AH, Zeh A, Gerdes A (2017) In Situ Sr isotopes in plagioclase and trace element systematics in the lowest part of the Eastern Bushveld complex: dynamic processes in an evolving Magma Chamber. Journal of Petrology 58(2):327–360. https://doi.org/10.1093/petrology/egx018

Xu Y, Yu Z, Mao C, Wang Y, Guo J (2010) Entity answer extraction of web table. In: 2010 seventh international conference on fuzzy systems and knowledge discovery. pp 2465–2468. https://doi.org/10.1109/FSKD.2010.5569791

Yan W, Ma H, Yang Z (2020) A general framework of knowledge-based coaching system with application in table tennis training. In: 2020 39th Chinese control conference (CCC). pp 2902–2907. https://doi.org/10.23919/CCC50068.2020.9188412

Yongvanich N, Jitpagdee T, Chukaew B, Papathe S (2019) Yellow ceramic pigments from amorphous nanosized oxides using rice husk and Zircon. In: 2019 IEEE 14th international conference on nano/micro engineered and molecular systems (NEMS). pp 225–228. https://doi.org/10.1109/NEMS.2019.8915653

Yurin AY, Dorodnykh NO, Shigarov AO (2021) Semi-automated formalization and representation of the engineering knowledge extracted from spreadsheet data. IEEE Access 9:157468–157481. https://doi.org/10.1109/ACCESS.2021.3130172

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.