Title: Comparing Empirical Regressions and Machine Learning Methods for Estimating Crop Residue Cover from Sentinel-2 Data

Subtopic: Estimating crop residue cover using remote sensing data

Crop residue cover is an important factor in agricultural management, as it affects soil erosion, water infiltration, and nutrient cycling. Remote sensing data, such as Sentinel-2 imagery, can be used to estimate crop residue cover. In this section, we compare empirical regressions and machine learning methods for estimating crop residue cover from Sentinel-2 data.

One of the machine learning methods that can be used for this task is the artificial neural network (ANN) algorithm. In a study by [REF0], the ANN algorithm was compared with the support vector machine (SVM) algorithm for estimating crop residue cover in detached residential buildings. The study found that the ANN algorithm was more suitable for this task.

Another machine learning method that can be used for estimating crop residue cover is the SVM algorithm. The SVM algorithm has a strict statistical theory and mathematical foundation, which makes it less reliant on the experience and knowledge of designers compared to the ANN algorithm [REF2]. The SVM algorithm can be used for linear classification, and it uses a formulation to find a hyperplane that separates the dataset into two classes [REF3].

Convolutional neural networks (CNN) can also be used for estimating crop residue cover. The CNN is composed of four main steps: convolution, sub-sampling, activation, and complete connection [REF3]. The CNN has been used for various image classification tasks, and it has shown promising results in remote sensing applications.

In addition to machine learning methods, empirical regressions can also be used for estimating crop residue cover. Empirical regressions are statistical models that use observed data to estimate the relationship between variables. Empirical regressions have been used for various remote sensing applications, including estimating crop residue cover [REF5].

In conclusion, both empirical regressions and machine learning methods can be used for estimating crop residue cover from Sentinel-2 data. The choice of method depends on the specific application and the available data. Machine learning methods, such as ANN, SVM, and CNN, have shown promising results in remote sensing applications, but empirical regressions can also be effective in certain cases [REF1][REF4][REF6][REF7][REF8][REF9].

References given to GPT:

[REF0] - paperID: 69b8c111ab8c8cb7dc35ffdd44f166f72ae86b2d Title: Machine Learning-Based Method for Detached Energy-Saving Residential Form Generation Chunk of text: 4. Experiment and Results In this section, the above operation process is practically performed so as to verify the effectiveness of this method, and the ANN and SVM algorithms are compared and studied to derive a machine learning approach more suitable for detached residential buildings for practical reference. 4.1. Sampling Method Comparison Experiment With variables x1, x2, which both take values in the range [0, 1], 10 sample data are sampled in the interval using the LHS operation process proposed in this paper with GHs native random number command, respectively, and the generated data are displayed using scatter plots. The comparison of the sampling points in (Figure 12) easily shows that the sampled data after the operation using the above method exhibits the characteristics of uniform LHS sampling distribution, while the sampled data in the traditional random number command has the disadvantages of overlapping data and uneven sampling. This experiment proves the effectiveness of the sampling method in this paper. Figure 11. Multi-objective optimization operation battery pack.

[REF1] - paperID: 46dd55ee266b08ced91c9423436e601c1ba4fc6d Title: AR Search Engine: Semantic Information Retrieval for Augmented Reality Domain Chunk of text: The second step is to represent AR documents as vectors according to the concepts of AR ontology. The clustering step is to be performed by semantically clustering all web documents into AR domains. Figure 5. Methodology for an ontology-based AR search engine. Preprocessing. This step is necessary to obtain efficient results . The preprocessing step is conducted to convert query and AR documents into a sequence of words using NLP techniques including tokenization and stop-word removal.

[REF2] - paperID: 69b8c111ab8c8cb7dc35ffdd44f166f72ae86b2d Title: Machine Learning-Based Method for Detached Energy-Saving Residential Form Generation Chunk of text: Figure 8. SVR nonlinear fitting schematic diagram. In Octopus, there is a similar operation logic to ANN, which is also composed of an input module (SVM learning) and an output module (SVM Evaluate). Unlike ANN, the input sample value does not need to be mapped. Because SVM has a strict statistical theory and mathematical foundation, unlike ANN, which needs to rely on the experience and knowledge of designers, it needs to adjust fewer parameter values, and its operation difficulty is lower than ANN. (Figure 9). Figure 9.

[REF3] - paperID: d17ffd50ddaa515e971e40a72a1645a0a4ea926a Title: Performance Evaluation of Keyword Extraction Methods and Visualization for Student Online Comments Chunk of text: This research will focus on the linear classification of SVM. Linear SVM needs to input a labelled data with paired, the mechanism is as follows: for dataset D = (x1,y1), (x2,y2), ..., (xn,yn), D include number n couple of elements, where y ∈ {−1, 1}, x is the features. SVM uses the formulation to find out the hyperplane to separate the dataset into two classes. The equation: y = ωx +b with the constraints: −ωxi − b ≤ ε and ωxi + b − yi ≤ ε. SVM algorithms are a cluster of kernel functions that can be used for many types of classification problems; the functions include polynomial, Gaussian, Gaussian radial basis function, Laplace RBF kernel, and sigmoid kernel. 3.4. Convolutional Neural Networks CNN is composed of four main steps: convolution, sub-sampling, activation, and complete connection.

[REF4] - paperID: 1973f251069086d850bde83fabfb05ceaf7b8859 Title: Multi-Stage Conversational Passage Retrieval: An Approach to Fusing Term Importance Estimation and Neural Query Rewriting Chunk of text: The re-ranking effectiveness of NTR (T5) approaches the manual queries, with 0.380 MAP and 0.566 NDCG@3. Second to NTR (T5), NTR (QuReTec) yields 0.367 MAP and 0.522 NDCG@3, slightly higher than scores for HQE (+POS) and Concat (+POS). The re-ranking results indicate that among all CQR methods, NTR (T5) generates the best queries for the BERT re-ranker, perhaps because it generates queries that more resemble natural language queries, and are thus wellsuited to the BERT re-ranker, which was trained on natural language queries. This explains why NTR (QuReTec) shows slightly better R@1000 and MAP using BM25 retrieval, but NTR (T5) yields significantly better early precision in full ranking. Also, note that although HQE (+POS) outperforms Concat (+POS) in early-precision metrics (i.e., NDCG@3, NDCG@1), they show comparable MAP in this condition, which suggests that HQE outperforms Concat in full ranking (see Table 4) mainly due to the gains from first-stage BM25 retrieval. 6.3 Fusion Analysis of Query Variations To better understand how query variations improve the retrieval effectiveness of conversational search, we conduct a thorough examination of query fusion in our multi-stage pipeline. Specifically, we investigate two fusion methods: (1) Late fusion: As shown in Figure 1(a), we directly conduct reciprocal rank fusion on the full ranking results from the two CQR methods.

[REF5] - paperID: 99cb99aed84bc85df3657b5984bb6b31cb8d425f Title: Inductive Entity Representations from Text via Link Prediction Chunk of text: For more specific technical details on datasets and training, we refer the reader to Appendices A and B. 4.1.2 Results. We report the Mean Reciprocal Rank (MRR), and Hits at 1, 3, and 10, on the test set in Table 3. For reference, we also show results reported by Wang et al. for KEPLER. We observe that in both the dynamic evaluation (WN18RR and FB15k-237) and the transfer evaluation (Wikidata5M), BLP-TransE consistently outperforms all the baselines across all metrics. We note that TransE results in higher link prediction performance with BLP, compared to alternatives like DistMult, ComplEx, and SimplE in WN18RR and FB15k-237. ComplEx and SimplE improve on performance in Wikidata5M, which contains around two orders of magnitude more triples for training.

[REF6] - paperID: d45057bdf0008c8c11e50878d3b60b521b09a437 Title: Catplayinginthesnow: Impact of Prior Segmentation on a Model of Visually Grounded Speech Chunk of text: + X i 0 max[0, α + d(u, i) − d(u, i0 )]! (1) This contrastive loss function encourages the network to minimise the cosine distance d by a margin α between an image i and its corresponding utterance u, while maximising the distance between mismatching image/utterance pairs i 0 /u and i/u 0 . In our experiments we set α = 0.2. Hyperparameters For both COCO and Flickr8k we use 1D convolutions with 64 filters of length 6 and a stride of 1 to preserve the original time resolution (and hence, boundary position). We use 512 units per recurrent layer for COCO and 1024 for Flickr8k.

[REF7] - paperID: 06227bc74bcee55471fb37bde0149b317f8a2014 Title: Enhancing Semantic Code Search With Deep Graph Matching Chunk of text: The example demonstrated above shows the original text(query) which is given as an input to the model, some operations are performed on that query, and "to add" and "to the" are masked and afterward, we have got the required output in which one sentence is represented as a combination of mask characters and original text. Now the similarity of sentences is used to find the exact result. The similarity is checked between the code description and the entered query using the attention function after applying GNN, this attention compares each code description with the query (basically a process of mapping) and gets the top relevant results. Afterward, Fmatch is used to compare the similarity score obtained from attention, and then that attention score will act as an input for Fmatch after which the result having the higher relevancy will be ranked using the cosine similarity. Step 3- Graph Generation: In this step, we discussed how to generate graphs for the query and code. Our concern is both contain rich semantic information so treating them as a plain sequence of tokens is inappropriate.

[REF8] - paperID: 7b237ba6d7c1ada14b8c5c3ccc404ce813d81216 Title: Material Translation Based on Neural Style Transfer with Ideal Style Image Retrieval Chunk of text: Unlimited time was given to select the fake image out of three options. Note that there was not an option to indicate that all photos are real. Thus, the participants were forced to carefully find the outlier image. Consequently, if they did not pick the synthesized image, it means that the translated results are real enough to fool human perception. We counted the results when participants do not choose the synthesized images. Given that, the average results of the 3000 votes show that 44.86% of the time, participants took the translated results as representative pictures of their target material. These findings are more significant for some materials, as shown in Figure 12.

[REF9] - paperID: 32ccd0f725fc5f7c81b57cad7787dc15b99151d0 Title: Neural methods for effective, efficient, and exposure-aware information retrieval Chunk of text: John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research, 12(Jul):2121–2159, 2011. Quoc V Le and Tomas Mikolov. Distributed representations of sentences and documents. In ICML, volume 14, pages 1188–1196, 2014. Mihajlo Grbovic, Nemanja Djuric, Vladan Radosavljevic, Fabrizio Silvestri, and Narayan Bhamidipati.

........................................................................................................................................................................................................

Title: Semantic and Geometric Modeling with Neural Message Passing in 3D Scene Graphs for Hierarchical Mechanical Search

Subtopic: Multi-modal retrieval with neural message passing

Semantic and Geometric Modeling with Neural Message Passing in 3D Scene Graphs for Hierarchical Mechanical Search is a challenging task that requires the integration of various techniques from different fields. In recent years, the use of neural networks has become increasingly popular in the field of natural language processing (NLP) and computer vision. Word2vec [41, 40], FastText [12, 30], and Doc2vec are some of the efficient models for learning word embeddings through neural networks. These models have been used to extract semantic information from text data and have shown promising results in various NLP tasks. In addition, CapsTM [REF1] is a neural network architecture that has been used to extract semantic information from text data. It consists of an input layer, a representation layer, an aggregation layer, a capsule layer, and a prediction layer. The representation layer extracts information from each piece of text and the interactive information between the two pieces of text.

Geometric information is also crucial for 3D scene graphs. LBP [REF4] is a simple and computationally efficient method for extracting geometric features from images. However, it has limitations in capturing structural information and is not invariant to rotations. Line segments are another important geometric feature in 3D scene graphs. After extraction, information about the geometry of the line segments is recorded, including the 3D start and end point and the normals of the neighborhood planes [REF5].

In addition to semantic and geometric modeling, the search strategy is also crucial for hierarchical mechanical search. The search strategy should be able to efficiently search through the 3D scene graph to find the target object. One approach is to use a binary threshold function to decide which nodes to search [REF2]. The value of the threshold can be updated during the search to prevent the correct nodes from never being searched.

Finally, the evaluation of the 3D scene graph model is important to ensure its effectiveness. Metrics such as MAE and RMSE are useful for evaluating the model's performance [REF3]. The uncertainties of the neural network models can be incorporated into the total uncertainties in the cost function. The recall rate is also an important metric for evaluating the model's performance [REF9].

In summary, the integration of semantic and geometric modeling with neural message passing in 3D scene graphs is a challenging task that requires the use of various techniques from different fields. The search strategy and evaluation metrics are also crucial for ensuring the effectiveness of the model.

References given to GPT:

[REF0] - paperID: 0fd5b1ef9d3d2e998e5c7f35f66df4de6a79a07f Title: A Comprehensive Study of Learning Approaches for Author Gender Identification Chunk of text: [10, 19, 58, 41]. Word2vec [41, 40] is one of the efficient models for learning word embeddings through CBoW and SG architectures using neural networks. FastText [12, 30] is a simple and efficient model that allows users to learn text representations as embeddings. Doc2vec applies unsupervised learning to infer continuous representations for larger blocks of text. Moreover, these representations are fed into Deep Neural Network (DNN) architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which became popular and have revolutionized the way to address various NLP tasks .

[REF1] - paperID: 05109f905d1015a4fa6f8c361128813bf196535b Title: CapsTM: capsule network for Chinese medical text matching Chunk of text: Figure 1 shows the overview architecture of CapsTM, which consists of an input layer, a representation layer, an aggregation layer, a capsule layer and a prediction layer. All these layers are presented in the following sections in detail. Input layer For the given pair of pieces of text (s1, s2), the input layer frst converts each piece of text into embeddings leant from large-scale unlabeled data by word2vec   or BERT  , denoted by e1 for s1 and e2 for s2, and then further makes a transformation to the embeddings using highway network as follows: where i=1,2, wf and wg are weight vectors, and bf and bg are bias vectors. Representation layer In the representation layer, two types of information are extracted: (1) information of each piece of text; (2) interactive information of the two pieces of text. We utilize BiLSTM to extract the frst type of information (Eq.  4) and attention-based interaction matrix to extract the second type of information (Eqs. 5–8) as follows: e­i = (1) tan h(wf ei + bf ), g = sigmoid (2) ­ wg e‑i + bg , e (3) ′

[REF2] - paperID: 30b7b5db2e096b0eb987852fb3aad39d5aed4ace Title: Semantic and Geometric Modeling with Neural Message Passing in 3D Scene Graphs for Hierarchical Mechanical Search Chunk of text: Once evaluated, the children nodes are again searched in sorted order, until they are all explored and the search reverts to the next node in the layer above. For containers, “search” means deciding which of the containers or objects within that container to search, whereas for objects “search” means moving the object out of the scene. An issue of the above approach is that it would search nodes even if the model gives them a low probability p of containing or occluding the target object. So as to avoid this,for all nodes if p is lower than a threshold T, the node is not searched but rather skipped. The value of T starts the same for all nodes, but can be updated during the search so as to prevent the correct nodes for finding the target object from never being searched. This update happens whenever a node is skipped, at which point the value of T for all subsequent nodes in that layer is set to to the value of p of that node. If the target is not found the first time all top-level nodes are either searched or skipped, the procedure restarts with the updated values of T and removed objects from the previous iteration being preserved.

[REF3] - paperID: f71d1ecea6133e55d10ebbdeb354560b4fea38a8 Title: Efficient multi-angle polarimetric inversion of aerosols and ocean color powered by a deep neural network forward model Chunk of text: (21) Both MAE and RMSE are useful metrics, where MAE is less dependent on outliers compared to RMSE. Analysis shows that the statistics of the differences between the NN prediction and the RT simulations as shown in Fig. 4 can be well modeled by Gaussian distributions and characterized by RMSE. Therefore the RMSE is used to represent the NN uncertainties for both reflectance (σρ ,NN) and DoLP (σρ ,NN) and will be incorporated into the total uncertainties in the cost function. Table 3 summarizes the uncertainties of the NN models. The σρ ,NN at 440 nm is 0.0006, which decreases to 0.0004 at 870 nm. However, due to the smaller reflectance magnitude at 870 nm, the corresponding RMSE for the percentage reflectance difference as shown in Fig. 4 is increased from 0.4 % at 440 nm to 1.0 % at 870 nm. For DoLP, the maximum σP ,NN is 0.003 at 870 nm, which <https://doi.org/10.5194/amt-14-4083-2021> Atmos.

[REF4] - paperID: 43f07db640626b26034d451a7168bec0d63ce5ac Title: Soft Computing based Artificial Neural Network and Multiple Feature set Intelligence system for Image Retrieval Chunk of text: P binary threshold function. Th(v) = 0, ifx ≤ 0 1, ifx ≥ 0 5 Even though LBP is computationally simple, limited in structural information capture, uses the only difference in pixels, ignore magnitude information. Increases computation complexity with respect to time and space as the exponential increase in the size of features with the neighbor number. it is not invariant to rotations

[REF5] - paperID: 9a5c5e207536d1be850da98795c4c8dd40269857 Title: LCD – Line Clustering and Description for Place Recognition Chunk of text: If the planes are co-planar (texture line) or only one plane exists (e.g. due to discontinuity), the line extremities are projected onto the corresponding plane. If the distance between the two plane centers along either of the surface normals is over a certain threshold (0.3 m), the line is projected onto the plane closer to the camera (discontinuity). Otherwise, the extremities are projected onto the intersection line of the two planes (edge). 3.1.2 Geometric information of line segment After extraction, information about the geometry of the line segments is recorded. It consists of the 3D start and end point ps/e, and the normals nl/r of the previously described neighborhood planes (Figure 3a). These planes give further structural information. If one plane does not exist, the corresponding normal is zero-padded.

[REF6] - paperID: 60c3f465402a9c660767fd59b86d708f010212d5 Title: Integrating pathway knowledge with deep neural networks to reduce the dimensionality in single-cell RNA-seq data Chunk of text: Pathways were used to define function-driven curated clusters (i.e. clusters of genes grouped by a biological common functionality using pathways). In order to compare the use of signaling pathways with other types of biological information previously used the genes have also been grouped by PPIs and GRNs (see Table 6). Neural network design The neural network proposed here consists of one input layer, one or two hidden (intermediate) layers and one output layer connected between them by a set of weights. The input layer ciphers the gene expression values, whereas the output layer encodes the probability of each cell type, which is learned as the information is propagated throughout the intermediate layers back and forward, updating the weights at each iteration (the so-called epochs). In the end, the network learns an internal representation of the underlying function of the data which in our case is conditioned by the biological priors used to construct the first hidden layer. The neural network model is formulated as follows: xi ¼ a Wi xð Þ i−1 þ

[REF7] - paperID: c2f8f8198ef9677dcec91f7e926e559da00df4a7 Title: Neural Architecture Search for GNN-based Graph Classification Chunk of text: Combine with the candidate operations shown in Table 3, the search space size of PAS-G variant can be calculated as 62 × 102 × 7 3 × 5 ≈ 6.2 × 106 and the PAS-NE variant is 62 × (10 × 3 × 2) 2 × 7 3 × 5 ≈ 2.2 × 108 . Moreover, more operations can be trivially added to the search space if the computational budget is enough, e.g., diferent score functions fs (·), similarity functions σ (·) in Eq. (5) and (6), etc. It also means that an eicient search method is needed over such a large search space. ACM Trans. Inf.

[REF8] - paperID: f93a5fef29b3a46c4271e6ada5b7d7e8f2e02da3 Title: Global Relation-Aware Attention Network for Image-Text Retrieval Chunk of text: Results of Text retrieval and Image retrieval tasks are shown in Figure6 and Figure7. As shown in Figure 6, we use an image to retrieve text. It can be seen from the results in Table 2-3 and the visualization results in Figure 6 that our method significantly improves Rcall@1 and Rcall@5 in Text retrieval. As shown in Figure 7, we use an text to retrieve image. It can be seen from the results that although the top1 is not the image corresponding to text, the content of the image can reflect the content of the text. 6 CONCLUSION AND FUTURE WORK In this paper, we propose a novel Global Relation-aware Attention Network (GRAN) for image-text retrieval. Instead of using mean pooling feature as global feature, we introduce Global Attention Module (GAM) to obtain global feature which contains sufficient semantics.

[REF9] - paperID: 3bf3ac52a8d50ea3a923465155bf1d818263d4b8 Title: Sketch-Based Retrieval Approach Using Artificial Intelligence Algorithms for Deep Vision Feature Extraction Chunk of text: The outcome is obtained with the introduced InfoGAN, which was trained from scratch compared to the methods based on handcrafted features. It takes about 1087s to train all three InfoGAN models (i.e., generator, discriminator, and auxiliary models). It requires 0.4 s for image indexing and 7.8 s for searching by each query image of the eleven images. Compared with the methods based on handcrafted features, the InfoGAN system training takes 1055 s (i.e., feature learning) over the entire ImagNet-Sketched dataset of images, while it takes 128 s to generate and match feature descriptors using SIFT from the predefined groups of only 80 images. Additionally, for the ORB case, it takes 97s to compute and match the generated descriptors over the sameAxioms 2022, 11, 663 25 of 36 selected images from the same dataset. Consequently, the use of SIFT to extract features across the entire dataset requires 300,000 s, and in the ORB case, 27,500 s are needed. Axioms 2022, 11, x FOR PEER REVIEW 24 of 37 Recall: Figure 23 demonstrates the computed recall based on correct and false matching outcomes and recorded results in Table 2.

........................................................................................................................................................................................................

Title: Unsupervised Cross-Modal Audio Representation Learning

Cross-modal retrieval using neural networks

Cross-modal representation learning has been widely used in various applications, including cross-modality matching/retrieval. Canonical Correlation Analysis (CCA) is a popular technique for identifying a projection pair that maximally captures the correlation between two sets of variables [REF0]. In recent years, deep learning (DL) features have been used in cross-modality matching/retrieval, and several studies have been conducted in this area [REF0]. However, existing preprocessing methods have some weaknesses, and computational efficiency and affordability are also important considerations [REF0]. To address these issues, a new image retrieval method based on OTS deep CNNs has been developed, which is primarily built upon CCA but has several notable differences from related works [REF0].

In evaluating cross-modal matching/retrieval methods, recall has been proposed as a more intuitive metric than directly comparing the square losses of different methods [REF1]. The square loss is highly dependent on hyperparameters and does not reflect any application meaning [REF1]. In a study comparing transfer learning methods, it was found that the proposed methods outperformed those trained without any information of the (item, item) correlation data [REF1].

Content-based image retrieval (CBIR) techniques have been widely used in identifying relevant images in a database [REF3]. However, patent images are binary and texture-less, making it difficult to represent the knowledge within them [REF3]. Local features, such as SIFT features, have been used to represent patent images, but the retrieval results only utilize geometric information of the query image, ignoring domain knowledge [REF3]. CNN enables learning richer features of images due to its deep architecture [REF3].

Gumbel-Top-k sampling technique has been integrated into a gradient descent-based learning-to-rank method to generate samples to approximate Equation (5) [REF4]. This algorithm is run iteratively at each training epoch, generating new samples using the updated model weights [REF4]. The empirical benefits of this weighting have been evaluated, and the results show that it improves the performance of the method [REF4].

Reducing latency by approximation has been proposed to address the problem of extreme tail latency [REF5]. Delayed, dynamic, selective prediction is used to parallelize the processing of long running queries, and dynamic features are collected during the initial processing period to improve the prediction [REF5].

In a study on sentence retrieval task, the incorporation of dialogue context significantly boosted performance by more than 20% in most cases, regardless of the metric used [REF6]. The combination of the D1 metric with the incorporation of the dialogue context led to an accuracy between 60% and 70%, reaching a peak performance of around 72% [REF6].

Dual-VGG network has been trained to extract features from patent images, converting the image space into a latent feature space [REF8]. The loss is calculated based on the categorical cross-entropy, and the backpropagation algorithm is used to change the weights and biases to make the network more predictive [REF8].

To correct metric measurements for pool bias, measuring effectiveness on judged passages has been proposed [REF9]. Additionally, an effect size analysis has been utilized to compare statistical significance across different retrieval collections [REF9].

In summary, cross-modal representation learning has been widely used in various applications, including cross-modality matching/retrieval. Several techniques have been proposed to address the challenges in this area, including the use of DL features, Gumbel-Top-k sampling technique, and delayed, dynamic, selective prediction. The incorporation of dialogue context and the use of Dual-VGG network have also been shown to improve performance.

References given to GPT:

[REF0] - paperID: 9308c9dbe83d67abbdfdf2077871f826fe4906b6 Title: Image Retrieval via Canonical Correlation Analysis and Binary Hypothesis Testing Chunk of text: CCA is a multivariate technique for elucidating the the associations among two sets of variables. It can be used to identify a projection pair of a given dimension that maximally captures the correlation between the two sets. The applications of CCA are too numerous to list. In cross-modality matching/retrieval alone, extensive investigations have been carried out as evidenced by a growing body of literature, from those based on handcrafted features to the more recent ones that make use of DL features [25–27]. There is also some related development on the theoretical front (see, e.g., [28,29]). Motivated by the consideration of computational efficiency and affordability as well as the weaknesses inherent in the existing preprocessing methods, we develop and present in this paper a new image retrieval method based on OTS deep CNNs. Our method is built primarily upon CCA, but has several notable differences from the related works.

[REF1] - paperID: 334bf07262320eb895a22973c948b4111e782daa Title: Zero-Shot Heterogeneous Transfer Learning from Recommender Systems to Cold-Start Search Retrieval Chunk of text: with the highest score(v𝑖 , vℓ), the recall for item 𝑦𝑖 is defined as: recall𝑖 = |S𝑡𝑟𝑢𝑒,𝑖 ∩ S𝑝𝑟𝑒𝑑,𝑖 | |S𝑡𝑟𝑢𝑒,𝑖 | , then we report the average recall over all items. We believe using recall for evaluation is more intuitive than directly comparing the square losses of different methods, because the square loss greatly depends on the hyperparameters, and the value itself does not reflect any application meaning. Results. Table 1 shows the recalls of all methods. We find the two proposed transfer learning methods ZSL\_ME and ZSL\_TE outperform STL and SMC. This is expected, as STL and SMC are trained without any information of the (item, item) correlation data.

[REF2] - paperID: a6b1126e058262c57d36012d0fdedc2417ad04e1 Title: Declarative Experimentation in Information Retrieval using PyTerrier Chunk of text: 8 f u l l \_ p i p e l i n e = p r f >> ( sdm ∗ ∗ b e r t ) >> l t r 9 f u l l \_ p i p e l

[REF3] - paperID: b51f1a2a37e0677cfe959d6d26023fe30550298f Title: A Convolutional Neural Network-Based Patent Image Retrieval Method for Design Ideation Chunk of text: Identifying relevant images to the given image in the database is not a trivial task. Second, all the images found in patent files are binary (black and white). Furthermore, the knowledge within patent images is hard to represent due to their texture-less characteristics. The current image searching methods are mostly based on content-based image retrieval (CBIR) techniques, which exploit the color, shape and textural information of images [10,11]. Some researchers use local features, such as scale-invariant feature transform (SIFT) features, to represent a patent image , but the retrieval results only utilize geometric information of the query image, ignoring domain knowledge. Compared to the classical local descriptors, CNN enables learning richer features of images because of its deep architecture .

[REF4] - paperID: bbf95d2f6e3eca4b3d4aa9906e58755796e8ff04 Title: Stochastic Retrieval-Conditioned Reranking Chunk of text: Algorithm 1 demonstrates how this Gumbel-Top-𝑘 sampling technique can be integrated in a gradient descent-based learningto-rank method to generate samples to approximate Equation (5). This algorithm is run iteratively at each training epoch, such that new samples are generated using the updated model weights 𝜃. Note that Algorithm 1 can be viewed as a generalization of the stochastic reranking approach first introduced by Bruch et al. . In the stochastic reranking setting, we also generate Gumbelperturbed score samples at each training epoch. However, in that approach, the samples can be viewed as uniformly weighted, whereas we introduce an explicit dependence on the retrieval stage by attaching a metric estimate 𝑀(𝑞, 𝐷, 𝑘) to each sample 𝐷. We evaluate the empirical benefits of this weighting in Section 5.3. Robust Estimation of 𝑴(𝒒, 𝑫, 𝒌). One can simply compute this evaluation metric using relevance judgments.

[REF5] - paperID: af1120eec4de18db766bff0d717e247c23b6b990 Title: Managing tail latency in large scale information retrieval systems Chunk of text: revisits this problem for extreme tail latency (at the 99.99th percentile). The authors propose using delayed, dynamic, selective prediction to parallelize the processing of long running queries. In particular, the prediction process is improved by firstly processing the query for a short, fixed period of time. During this time, dynamic features are collected, which help improve the prediction. The advantage of this method is that most of the short running queries will complete their processing during the initial processing period, which improves the precision of the predictor. Finally, predicted long running queries are accelerated using parallel processing. Reducing Latency by Approximation.

Figure 5. Top-3 accuracy using D1 as the distance metric, for k = 125.

[REF7] - paperID: 08e5ce6fc813a15ce8761dbc0462a93c696a444c Title: StruBERT: Structure-aware BERT for Table Search and Matching Chunk of text: after the BERT encoding step to pool over the contextualized tokens for each cell defined by [header\_name type cell\_content]. BERT is composed of L layers of Transformer blocks. The cell-wise average pooling is applied on the contextualized embedding that is obtained from the last layer. The contextualized embedding of the column-based sequence ci is given by: ci = [C LS ]T egj [S E P ]v1i [S E P ] . . .

[REF8] - paperID: b51f1a2a37e0677cfe959d6d26023fe30550298f Title: A Convolutional Neural Network-Based Patent Image Retrieval Method for Design Ideation Chunk of text: Formally, in our model, let the patent image be &. We trained a specialized Dual-VGG network ( = & → &. For every patent image &, i=1 to N, we can use our model to extract its features as a vector &. In this way, we converted the image space I into a latent feature space -. During the training process, the loss is calculated based on the categorical cross-entropy, and the backpropagation algorithm is used to change the weights and biases to make the network more predictive. The objective function is: = − 1 33(6 & log:;6 & <) > 6?@ A &?@ where N represents the number of images in a batch.

[REF9] - paperID: 82e3a63301eeb82090bc79c6009af20a5daea15a Title: Introducing Neural Bag of Whole-Words with ColBERTer: Contextualized Late Interactions using Enhanced Reduction Chunk of text: We follow Sakai to correct our metric measurements for pool bias by observing only measuring effectiveness on judged passages, which means removing all retrieved passages that are not judged and then re-assigning the ranks of the remaining ones. This is in contrast with the default assumption that non-judged passages are not relevant, which naturally favors methods that have been part of the pooling process. Additionally, we follow Soboroff to utilize an effect size analysis that is popular in medicine and social sciences. Soboroff proposed to use this effect size as meta analysis tool to be able to compare statistical significance across different retrieval collections. In this work we combine the 744Introducing Neural Bag of Whole-Words with ColBERTer CIKM ’22, October 17–21, 2022, Atlanta, GA, USA (a) BM25 vs. Uni-ColBERTer (Dim1, BOW2 + CS) Effect Size Weight Mean CI 95% TREC Covid 9.4% 0.53 [0.13, 0.94] TripClick 17.2% 0.17 [0.09, 0.25] NFCorpus 15.4% 0.36

........................................................................................................................................................................................................

Title: Aggregating Deep Features for Retrieval

Subtopic: Deep Learning-based Retrieval Architectures

Document retrieval systems aim to retrieve documents that are relevant to the user's information need (REF1). Traditional document retrieval systems are built on keyword searches, which do not consider the relations among the passages of text within a document (REF1). Models built on neural networks, on the other hand, cluster or classify documents into groups with similar documents belonging to the same group (REF1). In this context, deep features have been used to improve the performance of document retrieval systems.

Aggregating deep features for retrieval has been shown to be effective in various tasks. For instance, EARA, a retrieval system for remote sensing image retrieval, uses a submodule that includes the Main Branch and Residual Attention Branch to make extracted features more discriminative (REF2). In addition, the descriptor ensemble method was adopted by EARA to decrease the high-computation complexity of similar metric algorithms in RSIR (REF2). Similarly, in the context of contextual stance detection, metadata about the tweets have served in discriminating the stance better than the textual information of the tweets themselves (REF3).

To extract deep features, various features have been used, such as BERT, unigram, unigram-hashtag, char-grams, num-hashtag, punctuation marks, and length (REF4). In addition, community-based features have been used as discriminating features, such as network quote community, network reply community, network retweet community, and network friend community (REF4). These features are vectors of numerical attributes that represent the number of retweets, retweets with comments, number of friends, number of followers, count of lists, created at information, and number of emojis in the twitter bio (REF4).

To train models that use deep features, various techniques have been used. For instance, the MLP classifier with 128 hidden layers with 512 nodes each was used to train models for textual stance detection (REF4). The training uses K-fold cross-validation to fine-tune the model parameters with K = 5 folds (REF4). Similarly, GUL.LE.VER, a system built to solve the Italian quiz show "La Ghigliottina," uses web scraped resources, Glove algorithm for word representation, and a custom word correlation measure based on cosine similarity and inverse document frequency (idf) (REF7).

In conclusion, aggregating deep features for retrieval has been shown to be effective in various tasks, such as document retrieval, remote sensing image retrieval, and contextual stance detection. Various features and techniques have been used to extract and train models that use deep features.

References given to GPT:

[REF0] - paperID: bd23ce64a6422c1f73acf51675e53b7a06547da3 Title: UOBIT @ TAG-it: Exploring a Multi-faceted Representation for Profiling Age, Topic and Gender in Italian Texts Chunk of text: 5 Discussion Regarding the first model, we see how the vectorization obtained from the Wikipedia corpus performs well, particularly considering that it represents exclusively the pages’ titles. We also notice that the comparison between the two models is not straighforward since the ensemble model we used was not tested on the vectors obtained from the recurrent neural networks. We did not experiment in this mixed setting, since we believe it would not make sense to deploy a methodology with the power of XGBoost on embeddings solely based on the information present in the pages provided for this task. Indeed, there are high chances that the results for such complex model would still be worse than the one with the pretrained embeddings, since, as we mentioned in Section 4, the knowledge available exclusively in the pages proposed for this task is limited. The other remarkable aspect is that to surpass the performance of the GRU, handcrafted features were helpful, despite them being mostly word occurrences counts. This same information is available to the GRU models, which performs worse. This underlines how the recurrent architecture, though powerful and able to capture long distance relations, can not retain this type of substantial details.

[REF1] - paperID: cd7cc12cb01d87f8eed23911384abcffdfe130f4 Title: Applicability of Deep Neural Networks on the Task of Document Retrieval Chunk of text: Models that are built on neural networks usually cluster or classify documents into groups with similar documents belonging to the same group. Many document retrieval systems are built on keyword searches. However, these systems do not consider the relations among the passages of text within a document (Treeratpituk and Callan, 2006) (Wei and Croft, 2006).286 Computer Science & Information Technology (CS & IT) The task of document retrieval is to find documents of unstructured or semi-structured nature that satisfies the information need of a user from within a large collection. The goal of document retrieval systems is thus to retrieve documents that are relevant to the user's information need. Document retrieval systems must hence be able to retrieve desired information about a subject rather than to retrieve documents that look similar to a given query (Baeza-Yates and RibeiroNeto, 1999). To accomplish that task, document retrieval systems apply specific concepts to represent the query and documents, and to assign relevant documents to the query (Skovajsova, 2010).

[REF2] - paperID: 349746ed41051483e3b89c2cc0d33b2fefbfc616 Title: A Novel Ensemble Architecture of Residual Attention-Based Deep Metric Learning for Remote Sensing Image Retrieval Chunk of text: Discussion EARA shows relatively stable retrieval results despite the variations in time instance and shooting range, as shown in Figures 4–6. The submodule, including the Main Branch and Residual Attention Branch, of EARA makes extracted features more discriminative. In addition, the descriptor ensemble method was adopted by EARA to decrease the highcomputation complexity of similar metric algorithms in RSIR. Three factors affect the EARA performance: the type of the dataset, the number of images in the dataset, and the network structure. When it comes to the dataset type and the number of images in the dataset, EARA can give more play to its advantages on datasets with more categories such as AID and SIRI-WHU. The better performance of EARA on AID and SIRI-WHU than on UCMD demonstrates that EARA is somewhat data-driven with no need for a large number of parameters to constraint learners, and EARA can benefit from large-scale datasets. With a large amount of data, EARA can fully learn the similarities and dissimilarities between the images.

[REF3] - paperID: bd23ce64a6422c1f73acf51675e53b7a06547da3 Title: UOBIT @ TAG-it: Exploring a Multi-faceted Representation for Profiling Age, Topic and Gender in Italian Texts Chunk of text: Our model for Run 1 has outperformed the baseline model in terms of precision of favour and neutral tweets also shown a 11% increase in recall of against tweets over the baseline model. This can be interpreted that the most of the testing instances are identified as relevant tweet against the Sardines movement. For the second task on contextual stance detection, our models for Run 1 and 2 have performed better than the baseline model for the same, whose F1 average is given as 0.6284. The Run 1 for this task has used BERT, numhashtag, network friend community features whereas the run 2 has been built on BERT, network quote community, network friend community features. This can be inferred that the additional information about the Sardine tweets such as the community based contextual features have contributed towards the classification of the tweets. Metadata about the tweets have served in discriminating the stance better than the textual information of the tweets themselves. 5 Conclusion In this paper, we presented the suitable models for stance detection in Italian tweets about Sardine movement.

[REF4] - paperID: bd23ce64a6422c1f73acf51675e53b7a06547da3 Title: UOBIT @ TAG-it: Exploring a Multi-faceted Representation for Profiling Age, Topic and Gender in Italian Texts Chunk of text: ,; and their frequencies as numerical values; Length feature will extract the number of characters, the number of words, the average length of the words in each tweet; Community based features are also used as discriminating features in our work that exhibits the relationship among the tweets, comments such as network quote community, network reply community, network retweet community, network friend community. These features are vectors of numerical attributes that represent the number of retweets, retweets with comments, number of friends, number of followers, count of lists, created at information and number of emojis in the twitter bio. For the textual stance detection, features such as BERT, unigram, unigram-hashtag, char-grams, num-hashtag, punctuation marks and length are extracted from the training instances. These features are given to Multilayer Perceptron (MLP) with 128 hidden layers with 512 nodes each. The training uses K-fold cross validation to fine tune the model parameters with K = 5 folds. For the contextual stance detection, along with the features mentioned for the textual SD, additional features of the tweet such as network quote community, network reply community, network retweet community, network friend community, user info bio, tweet info retweet, tweet info create at were also extracted from the training instances and all are fed to MLP classifier with 512 nodes in each of 128 hidden layers. The second model also undergoes 5 fold cross validation to avoid overfitting and selection bias problems.

[REF5] - paperID: bd23ce64a6422c1f73acf51675e53b7a06547da3 Title: UOBIT @ TAG-it: Exploring a Multi-faceted Representation for Profiling Age, Topic and Gender in Italian Texts Chunk of text: 2020. Analysing lexical semantic change with contextualised word representations. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 3960– 3973, Online, July. Association for Computational Linguistics. Ganesh Jawahar, Benoît Sagot, and Djamé Seddah. 2019. What does BERT learn about the structure of language?

[REF6] - paperID: bd23ce64a6422c1f73acf51675e53b7a06547da3 Title: UOBIT @ TAG-it: Exploring a Multi-faceted Representation for Profiling Age, Topic and Gender in Italian Texts Chunk of text: We empirically found 1e-5 to be a good learning rate value, which is on par with the results of . Lastly, we choose to train all the models for 9 epochs with a batch size of 8 examples. 3.6 Results Table 1 contains the results obtained by our models for the first two tasks of the DANKMEMES competition. The components that were frozen during the training process are varied for the three main conducted experiments (i.e. combining ItalianBERT with VGCN and ResNet50, ResNet152 and VGG-16, respectively) to identify proper adjustments for the weights of the pretrained models. The best results among the four evaluated sets (i.e. validation, test for Task 1 and validation, test for Task 2) are obtained by either freezing only the VGCN-ItalianBERT component or by freezing both textual and image components. The necessity of freezing the text branch of the architecture underlines the fact that the pretrained weights for the ItalianBERT model already properly capture specific traits of Italian and prove to be a viable option, even when analyzing short texts such as memes. Furthermore, the last convolutional block of the image component needs to be unfrozen because training an architecture on potential meme images is a more specific task when compared to analyzing Italian text.

[REF7] - paperID: bd23ce64a6422c1f73acf51675e53b7a06547da3 Title: UOBIT @ TAG-it: Exploring a Multi-faceted Representation for Profiling Age, Topic and Gender in Italian Texts Chunk of text: In this field a language game like Who Wants to be a Millionaire?, in which the player must have a wide background knowledge in order to answer a series of multiple-choice questions, has been shown to be solved mining the web, with the same performance of a human player (Lam et al., 2003). Extract common sense human knowledge from Wikipedia articles is another proposed solution that is able to challenge a human player (Molino et al., 2013). In the same category of open-world language games is set “La Ghigliottina”, an Italian quiz show in which five words are submitted to the player as clues and he has to find the unique word that is correlated with all the clues. In order to find this hidden associations between clues and solution, a human player must possess a wide background knowledge and he has to be able to perform a complex task of reasoning on it in order of finding correlations between different word meanings in different contexts. In literature, a proposed solution to this problem is OTTHO (On the Tip of my THOught) (Semeraro et al., 2009; Semeraro et al., 2012) which achieved performance similar to humans using a network representation of the background knowledge and a spreading algorithm to find the solution. “Il mago della Ghigliottina” (Sangati et al., 2018), based on a co-occurrence matrix obtained from a corpus of patterns mined on web scraped resources and the Pointwise Mutual Information as measure of word correlation, achieved super-human performance. In order to explore a new way to solve this game, GUL.LE.VER is built using similar web scraped resources, Glove algorithm for word representation and a custom word correlation measure based on cosine similarity and inverse document frequency (idf).

[REF8] - paperID: bd23ce64a6422c1f73acf51675e53b7a06547da3 Title: UOBIT @ TAG-it: Exploring a Multi-faceted Representation for Profiling Age, Topic and Gender in Italian Texts Chunk of text: Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. In Yoshua Bengio and Yann LeCun, editors, 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems 26, pages 3111–3119, Lake Tahoe, Nevada, USA. Jeffrey Pennington, Richard Socher, and Christopher Manning.

[REF9] - paperID: bd23ce64a6422c1f73acf51675e53b7a06547da3 Title: UOBIT @ TAG-it: Exploring a Multi-faceted Representation for Profiling Age, Topic and Gender in Italian Texts Chunk of text: “Hate Speech in Pixels: Detection of Offensive Memes towards Automatic Moderation”. In: arXiv preprint arXiv:1910.02334 (2019). Shivangi Singhal et al. “SpotFake: A Multi-modal Framework for Fake News Detection”. In: 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM). IEEE. 2019, pp.

........................................................................................................................................................................................................

Title: Graph-based Feature Extraction and Classification for Document Retrieval

Subtopic: Graph-based document classification

In recent years, the development of automated systems for the classification of scientific publications and the organization of semantic search for scientific information in actual areas of research has become increasingly important [REF2]. One of the challenges in this area is how to select an optimal classification algorithm among a huge number of existing text classification algorithms [REF1]. In this context, graph-based feature extraction and classification have emerged as a promising approach for document retrieval.

One of the advantages of graph-based feature extraction is that it can capture the relationships between words in a document, which can be used to improve the accuracy of document retrieval. For example, in [REF3], a method was proposed to extract a bilingual dictionary by combining the inter-word relationship matrices mapped by cross-language word embedding vectors. The method was implemented for the first time in a minority low-resource language, such as Uyghur, and the overall accuracy was better than that of other previous bilingual dictionary extraction methods.

Graph-based classification methods have also been shown to be effective in document retrieval. For example, in [REF4], a method was proposed to increase the overall ranking of resources by generating a better contextualized representation for queries and resources. The method considered different methods that generate embeddings from different hidden layers, and the results indicated that the last hidden layer without special tokens like [CLS] outperformed the best methods by 1.29% on average.

Another advantage of graph-based feature extraction and classification is that it can be used to improve the accuracy of document retrieval for short queries. According to research from [REF1], the length of the query in 50% of the cases is up to three words. Similar problem is solved in the research described in [REF1]. The paper compares the performances of classification algorithms in classification of texts from social networks. Many of these texts are quite short, although not as short as queries.

In conclusion, graph-based feature extraction and classification have emerged as a promising approach for document retrieval. The use of graph-based methods can capture the relationships between words in a document, which can be used to improve the accuracy of document retrieval. Moreover, graph-based methods can be used to improve the accuracy of document retrieval for short queries.

References given to GPT:

[REF0] - paperID: 9985112cad071a48b498596749f186f5c50cbdf5 Title: Low Dimensional Discriminative Representation of Fully Connected Layer Features Using Extended LargeVis Method for High-Resolution Remote Sensing Image Retrieval Chunk of text: Method UCM WHU-RS 16 32 64 128 256 16 32 64 128 256 Ridge Regression 96.59% 96.45% 96.54% 96.39% 96.33% 97.19% 97.41% 97.20% 97.27% 97.33% Lasso 95.68% 95.54% 95.63% 95.48% 95.42% 96.27% 96.49% 96.28% 96.35% 96.42% SVR 98.27% 98.13% 98.22% 98.07% 98.01% 98.88% 99.11% 98.89% 98.96% 99.03% Method RSSCN7 AID 16 32 64 128 256 16 32 64 128 256 Ridge Regression 91.64% 91.92% 92.10% 92.03% 91.97% 92.95% 92.67% 92.93% 92.92% 92.74% Lasso 90.79% 91.07% 91.24% 91.17% 91.11% 92.09% 91.81% 92.07% 92.06% 91.88% SVR 92.22% 92.50% 92.68% 92.61% 92.55% 93.54% 93.26% 93.52% 93.51% 93.33% It can be seen from Table 3, SVR obtains better performance in all four datasets and dimensions. Ridge Regression and Lasso are classic regression methods, which are widely used for data regression. In this experiment, SVR is at least 0.58% higher than other methods in mAP of image retrieval. Since the results are from the regression of LargeVis, the retrieval performance is affected by the performance of CNN, LargeVis and regression. Therefore, SVR is chosen as the regression method in this paper. 5.5.

[REF1] - paperID: e643ba0a8c7976ef85376e60ba323c677be9c471 Title: Improvement of Information Retrieval Systems by Using Hidden Vertical Search Chunk of text: Domain Index 1 Domain Index N ... (Query, Domain) Classifier Figure 5. IRS with hidden vertical search The next question is how to select an optimal classification algorithm among a huge number of existing text classification algorithms (a comprehensive overview of text classification algorithms can be found in ), and constantly published new ones (for example ). Specificity of the proposed system is that the same method should be used for both document and query classification where queries are extremely short. According to research from , the length of the query in 50 % of the cases is up to three words. Similar problem is solved in the research described in . The paper compares the performances of classification algorithms in classification of texts from social networks. Many of these texts are quite short, although not as short as queries.

[REF2] - paperID: b25a0c97dac320bf2d4dc7a11d982d9ab908b9ca Title: RESEARCH OF APPROACHES TO THE RECOGNITION OF SEMANTIC IMAGES OF SCIENTIFIC PUBLICATIONS BASED ON NEURAL NETWORKS Chunk of text: INTRODUCTION In modern conditions, orientation in the continuously increasing volume of scientific publications without the use of automated tools is becoming more and more difficult. Scientists and specialists in different fields do not always manage to successfully track publications containing new significant results in their area of knowledge. The development of automated systems for the classification of scientific publications and the organization of semantic search for scientific information in actual areas of research will increase the efficiency of research work. Modern approaches to the organization of automated analysis of scientific texts are mainly associated with the use of neural network technologies. The application of machine learning for the classification of scientific publications according to given features will allow creating an automated system for searching scientific publications, as well as increasing the efficiency of searching for the latest publications in a given field of knowledge. When solving this task, a problem arises related to eliminating the contradiction between the contextdependent representation of texts in natural language and Iuliia Bruttan et al. Research of Approaches to the Recognition of Semantic Images of Scientific Publications Based on Neural Networks 39 the context-independent algorithms for their computer processing.

[REF3] - paperID: 4060a19b4cf312f434a66586d33b8c7694a6c441 Title: Chinese-Uyghur Bilingual Lexicon Extraction Based on Weak Supervision Chunk of text: 5. Conclusions In order to automatically construct a Chinese-Uyghur bilingual dictionary, this paper proposed a method to extract a bilingual dictionary by combining the inter-word relationship matrices mapped by cross-language word embedding vectors. Different experimental groups verified the effectiveness of this method, and the overall accuracy was better than that of other previous bilingual dictionary extraction methods. The sizes of the bilingual seed dictionary and the seed parallel corpus were essential for the improved results. This research used vector mapping methods between the languages to build a cross-lingual parallel aligned-term corpora automatically. Information 2022, 13, 175 17 of 18 The method in this paper was implemented for the first time in a minority low-resource language, such as Uyghur. We also analyzed the extracted correct and incorrect bilingual dictionary results, which will help improve the performance of subsequent cross-lingual information retrieval and extraction tasks. At the same time, it was also proven that the construction method of a Chinese-Uyghur dictionary can have better performance in constructing low-resource language dictionaries by borrowing cross-language word embedding vectors.

[REF4] - paperID: 315ac13a819e762da7fc711d1790111dcfdae31c Title: Learning To Rank Resources with GNN Chunk of text: Authors of BERT consider some methods that generate embeddings from different hidden layers. In this experiment, we aim to determine a method that increases an overall ranking of resources by generating a better contextualized representation for queries and resources. The following methods are considered in the experiment: [CLS] token of the last hidden layer (CLS), concatenation of the last 4 hidden layers (C4LH), sum of the last 4 hidden layers (S4LH), last hidden layer (LHCLS), last hidden layer without special tokens like [CLS] (LH-CLS). Table 9 shows the FedGNN results trained on five embedding generation methods. The results are reported on nDCG@10 metric on three datasets. The results indicate that the LH-CLS method outperforms the best methods by 1.29% on average.

[REF5] - paperID: 4eb672f2a74945fa5e5116eb9893e5a32448b104 Title: MusicYOLO: A Vision-Based Framework for Automatic Singing Transcription Chunk of text: MusicYOLO is inspired by the perspective of sound event detection and based on object detection. It includes pre-processing, note detection, and pitch labeling. The pre-processing module transforms one-dimensional (1D) audio sequences into a two-dimensional (2D) spectrogram. The note detection module obtains the note-level onset/offset. The pitch labeling module extracts each note’s pitch. The innovation of this paper is that we use the objection detection method to accomplish the note detection task from a macro perspective, and we obtain the pitch using a search method instead of signal calculation. The following sections will introduce note detection and pitch labeling briefly.

[REF6] - paperID: 49ce5066a85bc9b5c48e193f6a6c80a04ed0b7e2 Title: Effective and practical neural ranking Chunk of text: {sep}; d; {sep}, where “;” represents the concatenation operator.1 The output of the transformer network corresponding to this input is then linearly 1We use the BERT convention of [CLS] and {sep} to represent the classification and separation tokens, respectively. 147combined using a tuned weight matrix Wcombine ∈ R d×1 to compute the final ranking score as follows: R(q, d) = T [CLS]; q; {sep}; d; {sep} Wcombine. (4.1) The processing time of state-of-the-art neural rankers based on transformer networks is very high, e.g., approximately 50 documents ranked per second on a modern GPU, making such rankers impractical for most ad-hoc retrieval tasks. To gain an understanding of where are the most expensive components of a transformer network such as the Vanilla BERT model, we measure the run-times of the main steps of the model. We find that most of the processing is performed in the computations involving the transformer’s layers. In particular, about 50% of the total time is spent performing attention-related tasks.

[REF7] - paperID: c46402c6f6a3179e19a19a20df8b52baf0f6fe35 Title: Improved Acoustic Modeling for Automatic Piano Music Transcription Using Echo State Networks Chunk of text: Introduction Automatic Music Transcription (AMT) is one of the most challenging problems in Music Information Retrieval. The goal of AMT is to generate a score-like representation of a polyphonic audio signal. Due to many concurrently played notes from various instruments, complex overlapping of harmonics occurs in the acoustic signal. In many cases, the polyphony, e.g. the number of simultaneously active notes, is unknown and can vary over time. In recent years, AMT was successfully treated as a multi-label classification problem, in which every possible note is treated as one class. Recurrent Neural Networks (RNNs) define the stateof-the-art for acoustic modeling in piano transcription. In , one of the first approaches for acoustic modeling with recurrent neural networks was presented.

[REF8] - paperID: 0310f2e82ccb474580512cfd17dfc0fb102d1f06 Title: TransHash: Transformer-based Hamming Hashing for Efficient Image Retrieval Chunk of text: The primary reason is that according to , the Cauchy distribution could effectively pull close similar pairs into a small Hamming radius, giving it an edge when the hash code length is short. More importantly, to test the effectiveness of the proposed dualstream feature learning, we also include the performances of the Siamese model with the solo global feature learning module. As depicted in Fig. 2, TransHash w/o P consistently underperform the model with dual-feature learning design. On NUSWIDE and IMAGENET, the average decline is 2.08% and 2.83%, respectively. The above experimental experiments have evidenced the effectiveness of the design of our pure transformer-based hashing framework. Since the hyper-parameter 𝐾, which controls how many groups we will divide our local features into, is rather important in our design, we further provide an ablation study on the sensitivity of 𝐾 for various hash bits on CIFAR-10. Note that if the length of the final hash code vector is 16 and 𝐾 equals 2, then the global feature is responsible for learning the first 8 bit and each local feature vector for the latter 4 bits.

[REF9] - paperID: 49ce5066a85bc9b5c48e193f6a6c80a04ed0b7e2 Title: Effective and practical neural ranking Chunk of text: TREC 2014 web track overview. Technical report, MICHIGAN UNIV ANN ARBOR, 2015. Ronan Collobert, J. Weston, L. Bottou, Michael Karlen, K. Kavukcuoglu, and P. Kuksa. Natural language processing (almost) from scratch. Machine Learning Research, 12, 2011. Alexis Conneau, Guillaume Lample, Marc’Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. Word translation without parallel data.

........................................................................................................................................................................................................

Title: Integrating External Knowledge Retrieval in Fashion-Oriented Image Captioning

Subtopic: External Knowledge Retrieval in Neural Information Retrieval

Fashion-oriented image captioning is a challenging task that requires a deep understanding of fashion-related concepts and their relationships. To improve the performance of image captioning models, external knowledge retrieval techniques have been proposed. In this section, we discuss the integration of external knowledge retrieval in fashion-oriented image captioning.

One approach to integrating external knowledge retrieval is to use pre-trained language models such as BERT, XLNet, RoBERTa, and Transformers-XL (Devlin et al. 2019; Yang et al. 2019; Liu et al. 2019; Dai et al. 2019). These models have shown superior performance in various natural language processing tasks (Mikolov et al. 2013b; Iacobacci et al. 2015; Camacho-Collados et al. 2015; Neelakantan et al. 2014; Ruas et al. 2019; Ruas et al. 2020; Pilehvar et al. 2017). In future work, we plan to include these models in our approach to improve the performance of fashion-oriented image captioning (REF1).

Another approach is to use mathematical embedding techniques to represent complex mathematical expressions as single unit tokens for information retrieval (Krstovski and Blei 2018; Yasunaga and Laferty 2019). These techniques have shown promising results for identifying similar equations and contextual descriptive keywords (Gao et al. 2017; Krstovski and Blei 2018; Yasunaga and Laferty 2019). In this paper, we apply word2vec on the DLMF and on the collection of arXiv.org pre-print archive documents for generating embedding vectors for various math symbols and terms (REF2).

Ensemble methods have also been proposed to improve the performance of retrieval models. However, these methods become computationally expensive as retrieval models grow in size and complexity, and results in a substantial obstacle given the prevalence of large transformer architectures (REF3). To address this issue, we approach capturing uncertainty from a Bayesian perspective where leveraging a convenient property of dropout can be treated as a form of variational inference, referred to as Monte Carlo dropout (REF3).

In summary, integrating external knowledge retrieval techniques such as pre-trained language models, mathematical embedding techniques, and Bayesian approaches can improve the performance of fashion-oriented image captioning models. These techniques have shown promising results in various natural language processing tasks and information retrieval.

References given to GPT:

[REF0] - paperID: edf3b6b3f8a21d97209c1a361cf689d78eaefe0a Title: MMFL-Net: Multi-scale and Multi-granularity Feature Learning for Cross-domain Fashion Retrieval Chunk of text: Compared with CTL-L-R50IBN-CE (the current state-of-the-art method), our model achieves a +4.2% higher mAP, +11.4% higher Acc@1, and +1.0% higherMultimedia Tools and Applications (2022) 17 Acc@20. Our model outperforms L-R50IBN (a suboptimal state-of-the-art method) by a +10.9% higher Acc@1 and +10.4% higher mAP. Additionally, we use the k-reciprocal re-ranking post-processing method by comparing with L-R50IBN(re-ranking), which improves the accuracy of mAP and Acc@1 (+10.6%, +12.6%), respectively. Table 4. Comparison of retrieval accuracy with the state-of-the-art methods on DeepFashion-C2S Methods Backbone Image Size mAP Acc@1 Acc@20 Acc@50 FashionNet VGG-16 224 ×

[REF1] - paperID: 576bf2c6b3ec95b6a0488ee4c64d5e6e90d8cf83 Title: Math-word embedding in math search and semantic extraction Chunk of text: Nonetheless, since some of our results are promising, we plan to include Transformerbased systems, such as BERT (Devlin et al. 2019), XLNet (Yang et al. 2019), RoBERTa (Liu et al. 2019), and Transformers-XL (Dai et al. 2019), in future work. The overall performance of word embedding algorithms has shown superior results in many diferent NLP tasks, such as machine translation (Mikolov et al. 2013b), relation similarity (Iacobacci et al. 2015), word sense disambiguation (Camacho-Collados et al. 2015), word similarity (Neelakantan et al. 2014; Ruas et al. 2019), document classifcation (Ruas et al. 2020), and topic categorization (Pilehvar et al. 2017). In the same direction, we also (2) 퐯king − 퐯man ≈ 퐯queen − 퐯woman 10 Noise means, the data consists of many uninteresting tokens that afect the trained model negatively. 11 <https://github.com/allenai/bilm-tf> [Accessed Feb. 2020].Scientometrics (2020) 125:3017–3046 3023 1 3 explore how well mathematical tokens can be embedded according to their semantic information. However, mathematical formulae are highly ambiguous and, if not properly processed, their representation is jeopardized. Math embedding Recently, Krstovski and Blei (2018) proposed a variation of word embedding for mathematical expressions. Their main idea relies on the construction of a distributed representation of equations, considering the word context vector of an observed word and its word-equation context window.

[REF2] - paperID: 576bf2c6b3ec95b6a0488ee4c64d5e6e90d8cf83 Title: Math-word embedding in math search and semantic extraction Chunk of text: Similarly to this approach, Krstovski and Blei (2018) use a variation of word embedding (briefy discussed in the "Word Embedding" section) to represent complex mathematical expressions as single unit tokens for IR. In 2019, Yasunaga and Laferty (2019) explore an embedding technique based on recurrent neural networks to improve topic models by considering mathematical expressions. They state their approach outperforms topic models that do not consider mathematics in text and report a topic coherence improvement of 0.012 over the LDA8 baseline. Equation embedding, as in Gao et al. (2017); Krstovski and Blei (2018); Yasunaga and Laferty (2019), present promising results for identifying similar equations and contextual descriptive keywords. In the following, we will explore in more detail diferent techniques of word embedding ("Word Embedding" section). Likewise, we will examine diferent styles of adapting the process for math embedding ("Math Embedding" section). Word embedding In this paper, we apply word2vec (Mikolov et al. 2013b) on the DLMF (DLMF 2018) and on the collection of arXiv.org pre-print archive9 documents for generating embedding vectors for various math symbols and terms.

[REF3] - paperID: 1bdd1b60ef6ddf88744fc71d0ac703a34c818802 Title: Not All Relevance Scores are Equal: Efficient Uncertainty and Calibration Modeling for Deep Retrieval Models Chunk of text: Unfortunately, ensemble methods become computationally expensive as retrieval models grow in size and complexity, and results in a substantial obstacle given the prevalence of BERT and other large transformer architectures [26, 28, 40, 41]. As the objective is to maximize performance while ranking as many top 𝑛 documents as possible, running 𝑚 models simultaneously results in 𝑛 𝑚 fewer ranked documents. Lastly, in cases where these large models are used as pre-trained encoders, ensembles do not adequately capture uncertainty over these shared parameters. As such, we approach capturing uncertainty from a Bayesian perspective where leveraging a convenient property of dropout can be treated as a form of variational inference, referred to as Monte Carlo dropout . In this setting, dropout induces a stochastic ranking model which creates a distribution of scores as the dropout mechanism outputs different values each time it is run over the same input candidate document. The characteristics of this distribution then allow us to capture both aleatoric and epistemic uncertainty. While MC sampling still relies on an infeasible 1https://microsoft.github.io/msmarco/ number 𝑚 of forward passes over the ranking model, we expand on this work with a theoretically motivated extension where only the last layer needs to be Bayesian to capture both epistemic and aleatoric uncertainty.

[REF4] - paperID: 79ceaab99aa445f139604478cf0448efa882ec5f Title: HANME: Hierarchical Attention Network for Singing Melody Extraction Chunk of text: [5, 6, 7, 8]. Bittner et al. proposed a fully convolutional neural network to learn the pitch salience and achieved promising results. Hsieh et al. proposed an encoderdecoder architecture and a way to use the bottleneck layer of the network to estimate the presence of a melody line. Unfortunately, other type of network without bottleneck layer cannot enjoy the advantage of it. On the other hand, researchers also attempt to combine the deep learning technique with human auditory

[REF5] - paperID: 6e6e29cdf6ee7c0ca8f6495f6a49dde53d67a2fa Title: Query Resolution for Conversational Search with Limited Supervision Chunk of text: We use a two-step cascade ranking architecture , which we augment with a query resolution module (Section 4). First, the unsupervised initial retrieval step outputs the initial ranked list L1 (Section 3.2.1). 2We follow the TREC CAsT setup and only take into account q1:i−1 but not the passages retrieved for q1:i−1. Second, the re-ranking step outputs the final ranked list L (Section 3.2.2). Below we describe the two steps of the cascade ranking architecture. 3.2.1 Initial retrieval step. In this step we obtain the initial ranked list L1 by scoring each passage p in the passage collection D with respect to the resolved query qˆi using a lexical matching ranking function f1.

[REF6] - paperID: edf3b6b3f8a21d97209c1a361cf689d78eaefe0a Title: MMFL-Net: Multi-scale and Multi-granularity Feature Learning for Cross-domain Fashion Retrieval Chunk of text: 3.5.3 Training of proposed network We have described the primary training procedure used to train and optimize our proposed network in Algorithm 1. The training involves of five steps. (1) Dataset acquisition and data loader construction. We initialize the setup train set data loader, DLtrain , using a triplet sampler; gallery set data loader, DLgallery ; query set data loader, DLquery (lines 1–3). (2) Construct of a MMFL-Net network and initialization of our network with ImageNet pre-trained weights (lines 4–5). (3) Network feed forward propagation scheme. We compute the multi-task classification loss for multi-label attribute prediction and PID classification, metric loss, and overall loss of each mini-batch hard-mining sampling strategy using Eq.

[REF7] - paperID: e7cfda56237ad652cc0b242fd59adf800af0ba86 Title: Deep-Learning-Based Complex Scene Text Detection Algorithm for Architectural Images Chunk of text: In NAS-FPN, a duplicate FPN module can be searched. We obtain a tradeoff between speed and accuracy by controlling the number of repetitions of this module. Then, we output the prediction results at different stages according to the different computing resources. In RetinaNet, the feature fusion strategy of FPN is adopted. NAS-FPN replaces the FPN portion of RetinaNet with the searched fusion architecture. This can find a better FPN architecture of the retinal network framework and improve the accuracy of text detection. The RetinaNet framework with NAS-FPN is depicted in Figure 4.Mathematics 2022, 10, 3914 7 of 22 Figure 4.

[REF8] - paperID: 8dfe6685181b93d0463ff16b8e895f73f1fd4bbc Title: Multi-Head Self-Attention Gated-Dilated Convolutional Neural Network for Word Sense Disambiguation Chunk of text: In this case, it falls into local minimum and optimal solution cannot be found. Therefore, learning rate is assigned with appropriate value for ensuring that the loss reaches minimum value faster. The number of attention heads affects the performance of the proposed network. The second group of experiments aims to investigate the effect of head number on WSD. The number of attention heads is set to 1, 2, 4, and 8. Experimental results are shown in Table 4. TABLE 4. Disambiguation accuracies in the second group of experiments Heads SemEval-2007: Task #5 SemEval-2021: Task #2 1 85.36 72.64 2 86.46 74.91 4 87.66 78.22 8 87.04 76.48 From Table 4, we can see that average accuracy of the proposed network first increases and then decreases with the increase of head number.

[REF9] - paperID: 1bdd1b60ef6ddf88744fc71d0ac703a34c818802 Title: Not All Relevance Scores are Equal: Efficient Uncertainty and Calibration Modeling for Deep Retrieval Models Chunk of text: 𝐸𝑅𝐶𝐸 = Õ 𝑀 𝑚=1 |𝐵𝑚 | 𝑛 1 |𝐵𝑚 | Õ (𝐷𝑖 ,𝐷𝑗) ∈𝐵𝑚 𝑃 (𝐷𝑖 > 𝐷𝑗) − 1 |𝐵𝑚 | Õ (𝐷𝑖 ,𝐷𝑗) ∈𝐵𝑚 1(𝐷𝑖 > 𝐷𝑗) . (12) This allows for a consistent calibration error while still accounting for relevance score distributions being conditioned on queries. The indicator function is defined as 1(𝐷𝑖 > 𝐷𝑗) = ( 1 if ranking 𝐷𝑖 above 𝐷𝑗 increases MAP 0 if ranking 𝐷𝑖 above 𝐷𝑗 decreases MAP, where MAP is mean average precision. This formulation removes all comparisons between pairs of relevant documents, pairs of nonrelevant documents, and documents scored from different queries, which allows for measuring only the calibration between relevant and non-relevant pairs conditioned in the same query. In the case of deterministic models which do not have a probabilistic perspective on relevance, we use the pairwise softmax function to calculate 𝑃 (𝐷𝑖 > 𝐷𝑗). 3.5 Risk Aware Rerankings As each document now has a predictive distribution, we are able to rerank a set of candidates based on a user defined allotted risk.

........................................................................................................................................................................................................