

Multi-modal Preference Modeling for Product Search

- 类型 会议论文
作者 Yangyang Guo
作者 Zhiyong Cheng
作者 Liqiang Nie
作者 Xin-Shun Xu
作者 Mohan Kankanhalli
URL <http://doi.acm.org/10.1145/3240508.3240541>
系列 MM '18
地点 New York, NY, USA
出版社 ACM
页码 1865–1873
ISBN 978-1-4503-5665-7
日期 2018
其它 event-place: Seoul, Republic of Korea
DOI 10.1145/3240508.3240541
访问时间 2019/11/17 上午12:56:47
馆藏目录 ACM Digital Library
摘要 The visual preference of users for products has been largely ignored by the existing product search methods. In this work, we propose a multi-modal personalized product search method, which aims to search products which not only are relevant to the submitted textual query, but also match the user preferences from both textual and visual modalities. To achieve the goal, we first leverage the also_view and buy_after_viewing products to construct the visual and textual latent spaces, which are expected to preserve the visual similarity and semantic similarity of products, respectively. We then propose a translation-based search model (TranSearch) to 1) learn a multi-modal latent space based on the pre-trained visual and textual latent spaces; and 2) map the users, queries and products into this space for direct matching. The TranSearch model is trained based on a comparative learning strategy, such that the multi-modal latent space is oriented to personalized ranking in the training stage. Experiments have been conducted on real-world datasets to validate the effectiveness of our method. The results demonstrate that our method outperforms the state-of-the-art method by a large margin.
投递标题 Proceedings of the 26th ACM International Conference on Multimedia
添加日期 2019/11/17 上午12:56:47
修改日期 2019/11/17 上午12:56:47

标签:

multi-modal fusion, personalization, product search

笔记:

提出了一个基于 结合文字-图片的多模态向量空间 的搜索模型。

proposed a translation-based search model (TranSearch) to 1) learn a multi-modal latent space based on the pre-trained visual and textual latent spaces; and 2) map the users, queries and products into this space for direct matching.

附件

- ACM Full Text PDF

Learning Latent Vector Spaces for Product Search

类型 会议论文
作者 Christophe Van Gysel
作者 Maarten de Rijke
作者 Evangelos Kanoulas
URL <http://doi.acm.org/10.1145/2983323.2983702>
系列 CIKM '16
地点 New York, NY, USA
出版社 ACM
页码 165–174
ISBN 978-1-4503-4073-1
日期 2016
其它 event-place: Indianapolis, Indiana, USA
DOI 10.1145/2983323.2983702
访问时间 2019/11/17 上午12:58:28
馆藏目录 ACM Digital Library
摘要 We introduce a novel latent vector space model that jointly learns the latent representations of words, e-commerce products and a mapping between the two without the need for explicit annotations. The power of the model lies in its ability to directly model the discriminative relation between products and a particular word. We compare our method to existing latent vector space models (LSI, LDA and word2vec) and evaluate it as a feature in a learning to rank setting. Our latent vector space model achieves its enhanced performance as it learns better product representations. Furthermore, the mapping from words to products and the representations of words benefit directly from the errors propagated back from the product representations during parameter estimation. We provide an in-depth analysis of the performance of our model and analyze the structure of the learned representations.
投递标题 Proceedings of the 25th ACM International on Conference on Information and Knowledge Management
添加日期 2019/11/17 上午12:58:29
修改日期 2019/11/17 上午12:58:29

标签:

entity retrieval, latent space models, representation learning

笔记:

提出了一个 基于向量空间的 结合商品和描述模态的 排序模型

附件

- ACM Full Text PDF

Learning a Hierarchical Embedding Model for Personalized Product Search

类型 会议论文

作者 Qingyao Ai

作者 Yongfeng Zhang

作者 Keping Bi

作者 Xu Chen

作者 W. Bruce Croft

URL <http://doi.acm.org/10.1145/3077136.3080813>

系列 SIGIR '17

地点 New York, NY, USA

出版社 ACM

页码 645–654

ISBN 978-1-4503-5022-8

日期 2017

其它 event-place: Shinjuku, Tokyo, Japan

DOI 10.1145/3077136.3080813

访问时间 2019/11/17 上午12:58:50

馆藏目录 ACM Digital Library

摘要 Product search is an important part of online shopping. In contrast to many search tasks, the objectives of product search are not confined to retrieving relevant products. Instead, it focuses on finding items that satisfy the needs of individuals and lead to a user purchase. The unique characteristics of product search make search personalization essential for both customers and e-shopping companies. Purchase behavior is highly personal in online shopping and users often provide rich feedback about their decisions (e.g. product reviews). However, the severe mismatch found in the language of queries, products and users make traditional retrieval models based on bag-of-words assumptions less suitable for personalization in product search. In this paper, we propose a hierarchical embedding model to learn semantic representations for entities (i.e. words, products, users and queries) from different levels with their associated language data. Our contributions are three-fold: (1) our work is one of the initial studies on personalized product search; (2) our hierarchical embedding model is the first latent space model that jointly learns distributed representations for queries, products and users with a deep neural network; (3) each component of our network is designed as a generative model so that the whole structure is explainable and extendable. Following the methodology of previous studies, we constructed personalized product search benchmarks with Amazon product data. Experiments show that our hierarchical embedding model significantly outperforms existing product search baselines on multiple benchmark datasets.

投递标题 Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval

添加日期 2019/11/17 上午12:58:50

修改日期 2019/11/17 上午12:58:50

标签:

latent space model, personalization, product search, representation learning

笔记:

提出了一个神经网络 潜在向量表征 模型， 将 用户， 查询， 商品 三个模态的信息融合
个性化推荐模型

附件

- ACM Full Text PDF

Deep Visual-Semantic Hashing for Cross-Modal Retrieval

类型 会议论文

作者 Yue Cao

作者 Mingsheng Long

作者 Jianmin Wang

作者 Qiang Yang

作者 Philip S. Yu

URL <http://doi.acm.org/10.1145/2939672.2939812>

系列 KDD '16

地点 New York, NY, USA

出版社 ACM

页码 1445–1454

ISBN 978-1-4503-4232-2

日期 2016

其它 event-place: San Francisco, California, USA

DOI 10.1145/2939672.2939812

访问时间 2019/11/17 上午1:04:23

馆藏目录 ACM Digital Library

摘要 Due to the storage and retrieval efficiency, hashing has been widely applied to approximate nearest neighbor search for large-scale multimedia retrieval. Cross-modal hashing, which enables efficient retrieval of images in response to text queries or vice versa, has received increasing attention recently. Most existing work on cross-modal hashing does not capture the spatial dependency of images and temporal dynamics of text sentences for learning powerful feature representations and cross-modal embeddings that mitigate the heterogeneity of different modalities. This paper presents a new Deep Visual-Semantic Hashing (DVSH) model that generates compact hash codes of images and sentences in an end-to-end deep learning architecture, which capture the intrinsic cross-modal correspondences between visual data and natural language. DVSH is a hybrid deep architecture that constitutes a visual-semantic fusion network for learning joint embedding space of images and text sentences, and two modality-specific hashing networks for learning hash functions to generate compact binary codes. Our

architecture effectively unifies joint multimodal embedding and cross-modal hashing, which is based on a novel combination of Convolutional Neural Networks over images, Recurrent Neural Networks over sentences, and a structured max-margin objective that integrates all things together to enable learning of similarity-preserving and high-quality hash codes. Extensive empirical evidence shows that our DVSH approach yields state of the art results in cross-modal retrieval experiments on image-sentences datasets, i.e. standard IAPR TC-12 and large-scale Microsoft COCO.

投递标题 Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

添加日期 2019/11/17 上午1:04:23

修改日期 2019/11/17 上午1:04:23

标签:

cross-modal retrieval, deep hashing, multimodal embedding

笔记:

神经网络多模态哈希模型，结合图片（CNN）和文字（RNN）融合模态

Our architecture effectively unifies joint multimodal embedding and cross-modal hashing, which is based on a novel combination of Convolutional Neural Networks over images, Recurrent Neural Networks over sentences, and a structured max-margin objective that integrates all things together to enable learning of similarity-preserving and high-quality hash codes.

附件

- ACM Full Text PDF

Multi-modal semantic autoencoder for cross-modal retrieval

类型 期刊文章

作者 Yiling Wu

作者 Shuhui Wang

作者 Qingming Huang

URL <http://www.sciencedirect.com/science/article/pii/S0925231218313730>

卷 331

页码 165-175

期刊 Neurocomputing

ISSN 0925-2312

日期 February 28, 2019

刊名缩写 Neurocomputing

DOI 10.1016/j.neucom.2018.11.042

访问时间 2019/11/17 上午1:09:29

馆藏目录 ScienceDirect

语言 en

摘要 Cross-modal retrieval has gained much attention in recent years. As the research mainstream, most of existing approaches learn projections for data from different modalities into a common space where data can be compared directly. However, they neglect the preservation of feature and semantic information, so they are unable to obtain satisfactory results as expected. In this paper, we propose a two-stage learning method to learn multi-modal mappings that project multi-modal data to low dimensional embeddings that preserve both feature and semantic information. In the first stage, we combine both low-level feature and high-level semantic information to learn feature-aware semantic code vectors. In the second stage, we use encoder-decoder paradigm to learn projections. The encoder projects feature vectors to code vectors, and the decoder projects code vectors back to feature vectors. The encoder-decoder paradigm guarantees the embeddings to preserve both feature and semantic information. An alternating minimization procedure is developed to solve the multi-modal semantic autoencoder optimization problem. Extensive experiments on three benchmark datasets demonstrate that the proposed method outperforms state-of-the-art cross-modal retrieval methods.

添加日期 2019/11/17 上午1:09:29

修改日期 2019/11/17 上午1:09:29

标签:

Autoencoder, Cross-modal retrieval, Multi-modal data

笔记:

用autoencoder(encoding + decoder)模型学习双模态表征(图片+文字)

此文章是2019的Neurocomputing的文章，综述部分可以参考

附件

- ScienceDirect Snapshot
- ScienceDirect Full Text PDF

Adversarial Cross-Modal Retrieval

类型 会议论文

作者 Bokun Wang

作者 Yang Yang

作者 Xing Xu

作者 Alan Hanjalic

作者 Heng Tao Shen

URL <http://doi.acm.org/10.1145/3123266.3123326>

系列 MM '17

地点 New York, NY, USA

出版社 ACM

页码 154–162

ISBN 978-1-4503-4906-2

日期 2017

其它 event-place: Mountain View, California, USA

DOI 10.1145/3123266.3123326

访问时间 2019/11/17 上午1:47:12

馆藏目录 ACM Digital Library

摘要 Cross-modal retrieval aims to enable flexible retrieval experience across different modalities (e.g., texts vs. images). The core of cross-modal retrieval research is to learn a common subspace where the items of different modalities can be directly compared to each other. In this paper, we present a novel Adversarial Cross-Modal Retrieval (ACMR) method, which seeks an effective common subspace based on adversarial learning. Adversarial learning is implemented as an interplay between two processes. The first process, a feature projector, tries to generate a modality-invariant representation in the common subspace and to confuse the other process, modality classifier, which tries to discriminate between different modalities based on the generated representation. We further impose triplet constraints on the feature projector in order to minimize the gap among the representations of all items from different modalities with same semantic labels, while maximizing the distances among semantically different images and texts. Through the joint exploitation of the above, the underlying cross-modal semantic structure of multimedia data is better preserved when this data is projected into the common subspace. Comprehensive experimental results on four widely used benchmark datasets show that the proposed ACMR method is superior in learning effective subspace representation and that it significantly outperforms the state-of-the-art cross-modal retrieval methods.

投递标题 Proceedings of the 25th ACM International Conference on Multimedia

添加日期 2019/11/17 上午1:47:12

修改日期 2019/11/17 上午1:47:12

标签:

adversarial learning, cross-modal retrieval, modality gap

笔记:

多模态对抗学习的方法

附件

- ACM Full Text PDF

Cross-Modal Retrieval via Deep and Bidirectional Representation Learning

类型 期刊文章

作者 Y. He

作者 S. Xiang

作者 C. Kang

作者 J. Wang

作者 C. Pan

卷 18

期 7

页码 1363-1377

期刊 IEEE Transactions on Multimedia

日期 2016

DOI 10.1109/TMM.2016.2558463

馆藏目录 Scopus

摘要 Cross-modal retrieval emphasizes understanding inter-modality semantic correlations, which is often achieved by designing a similarity function. Generally, one of the most important things considered by the similarity function is how to make the cross-modal similarity computable. In this paper, a deep and bidirectional representation learning model is proposed to address the issue of image-Text cross-modal retrieval. Owing to the solid progress of deep learning in computer vision and natural language processing, it is reliable to extract semantic representations from both raw image and text data by using deep neural networks. Therefore, in the proposed model, two convolution-based networks are adopted to accomplish representation learning for images and texts. By passing the networks, images and texts are mapped to a common space, in which the cross-modal similarity is measured by cosine distance. Subsequently, a bidirectional network architecture is designed to capture the property of the cross-modal retrieval-the bidirectional search. Such architecture is characterized by simultaneously involving the matched and unmatched image-Text pairs for training. Accordingly, a learning framework with maximum likelihood criterion is finally developed. The network parameters are optimized via backpropagation and stochastic gradient descent. A great deal of experiments are conducted to sufficiently evaluate the proposed method on three publicly released datasets: IAPRTC-12, Flickr30k, and Flickr8k. The overall results definitely show that the proposed architecture is effective and the learned representations have good semantics to achieve superior cross-modal retrieval performance. © 1999-2012 IEEE.

档案 Scopus

添加日期 2019/11/17 上午1:48:38

修改日期 2019/11/17 上午1:48:38

标签:

Bidirectional Modeling, Convolutional Neural Network, Cross-Modal Retrieval, Representation Learning, Word Embedding

笔记:

Cited By :32

CNN + Bi-directional 方法学习图片+文字跨模态表示

附件

- SCOPUS Snapshot

A Multi-View Embedding Space for Modeling Internet Images, Tags, and Their Semantics

类型 期刊文章

作者 Yunchao Gong

作者 Qifa Ke

作者 Michael Isard
作者 Svetlana Lazebnik
URL <https://doi.org/10.1007/s11263-013-0658-4>
卷 106
期 2
页码 210-233
期刊 International Journal of Computer Vision
ISSN 1573-1405
日期 2014-01-01
刊名缩写 Int J Comput Vis
DOI 10.1007/s11263-013-0658-4
访问时间 2019/11/17 上午1:54:29
馆藏目录 Springer Link
语言 en
摘要 This paper investigates the problem of modeling Internet images and associated text or tags for tasks such as image-to-image search, tag-to-image search, and image-to-tag search (image annotation). We start with canonical correlation analysis (CCA), a popular and successful approach for mapping visual and textual features to the same latent space, and incorporate a third view capturing high-level image semantics, represented either by a single category or multiple non-mutually-exclusive concepts. We present two ways to train the three-view embedding: supervised, with the third view coming from ground-truth labels or search keywords; and unsupervised, with semantic themes automatically obtained by clustering the tags. To ensure high accuracy for retrieval tasks while keeping the learning process scalable, we combine multiple strong visual features and use explicit nonlinear kernel mappings to efficiently approximate kernel CCA. To perform retrieval, we use a specially designed similarity function in the embedded space, which substantially outperforms the Euclidean distance. The resulting system produces compelling qualitative results and outperforms a number of two-view baselines on retrieval tasks on three large-scale Internet image datasets.

添加日期 2019/11/17 上午1:54:29
修改日期 2019/11/17 上午1:54:29

标签:

Canonical correlation, Image search, Internet images, Tags

笔记:

CCA（典型相关分析）方法 学三个模态的

附件

- Springer Full Text PDF

Multi-label Cross-Modal Retrieval

类型 会议论文
作者 Viresh Ranjan
作者 Nikhil Rasiwasia

作者 C. V. Jawahar

页码 4094-4102

日期 December 2015

其它 ISSN: 2380-7504

DOI 10.1109/ICCV.2015.466

馆藏目录 IEEE Xplore

学术会议名称 2015 IEEE International Conference on Computer Vision (ICCV)

摘要 In this work, we address the problem of cross-modal retrieval in presence of multi-label annotations. In particular, we introduce multi-label Canonical Correlation Analysis (ml-CCA), an extension of CCA, for learning shared subspaces taking into account high level semantic information in the form of multi-label annotations. Unlike CCA, ml-CCA does not rely on explicit pairing between modalities, instead it uses the multi-label information to establish correspondences. This results in a discriminative subspace which is better suited for cross-modal retrieval tasks. We also present Fast ml-CCA, a computationally efficient version of ml-CCA, which is able to handle large scale datasets. We show the efficacy of our approach by conducting extensive cross-modal retrieval experiments on three standard benchmark datasets. The results show that the proposed approach achieves state of the art retrieval performance on the three datasets.

投递标题 2015 IEEE International Conference on Computer Vision (ICCV)

添加日期 2019/11/17 上午1:55:40

修改日期 2019/11/17 上午1:55:40

标签:

Benchmark testing, Computer vision, Correlation, correlation methods, image processing, image retrieval, learning (artificial intelligence), ml-CCA, multilabel annotation, multilabel canonical correlation analysis, multilabel cross-modal retrieval, Multimedia communication, Portable computers, semantic information, semantic networks, Semantics, Standards, subspace learning

笔记:

CCA（典型相关分析）的神经网络变体版本

附件

- IEEE Xplore Full Text PDF
- IEEE Xplore Abstract Record