MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY
http://mmc.committees.comsoc.org/

MMTC Communications – Review

Vol. 9, No. 5, October 2018

IEEE COMMUNICATIONS SOCIETY

TABLE OF CONTENTS

Message from the Review Board Directors

Welcome to the October 2018 issue of the IEEE ComSoc MMTC Communications – Review.

This issue comprises three reviews that cover multiple facets of multimedia communication research including resource allocation in heterogeneous networks, fine-grained venue discovery from multimedia data, and mobile edge cache placement for adaptive video streaming. These reviews are briefly introduced below.

The first paper, published in the IEEE Transactions on Communications and edited by Dr. Qin Wang designed a novel architecture of two-tier heterogeneous networks in which a massive MIMO technique is applied at a macro-cell base stations overlaid with a second tier of small-cells.

The second paper is published in the IEEE Transactions on Neural Networks and Learning Systems and edited by Dr. Yifang Yin. It describes a method for fine-grained venue discovery with complicated real images taken by users such as venue photos containing objects and geographic categories.

The third paper, published in the IEEE Transactions on Multimedia and edited by Dr. Roger Zimmermann, investigates video caching in the context of the modern and popular dynamic adaptive streaming over HTTP (DASH) in wireless networks while also taking into account the different rate-distortion (R–D) characteristics of videos.

All the authors, nominators, reviewers, editors, and others who contribute to the release of this issue deserve appreciation with thanks.

IEEE ComSoc MMTC Communications – Review Directors

Qing Yang
University of North Texas, USA
Email: qing.yang@unt.edu

Roger Zimmermann
National University of Singapore, Singapore
Email: rogerz@comp.nus.edu.sg

Wei Wang
San Diego State University, USA
Email: wwang@mail.sdsu.edu

Zhou Su
Shanghai University, China
Email: zhousu@ieee.org

Large System Analysis of Resource Allocation in Heterogeneous Networks

A short review for "Large System Analysis of Resource Allocation in Heterogeneous Networks with Wireless Backhaul"

Edited by Qin Wang

W. Xia. J. Zhang, S. Jin, C.-K. Wen, F. Gao and H. Zhu, " Large System Analysis of Resource Allocation in Heterogeneous Networks With Wireless Backhaul," IEEE Transactions on Communications, vol. 65, no. 11, Nov. 2017.

It is envisioned that the Internet of things (IoT) will change the paradigm of modern wireless communications by increasing the number of commonly used Internet-enabled things [1]. IoT applications raise several issues with regard to the next-generation wireless communication. For example, real-time and high-capacity infrastructure required by tele-medicine, highly reliable and low-latency output required by smart grid and critical infrastructure monitoring, and ubiquitous coverage demanded for environmental monitoring. Besides, IoT also brings challenges to energy and cost savings in terms of deploying and managing massive devices. To fulfill the aforementioned requirements of IoT, a novel architecture of two-tier heterogeneous networks (HetNets) was introduced, in which massive multiple-input multiple-output (MIMO) systems [2] and small-cell networks coexist [3] by overlaying macro-cell base stations (BSs) equipped with massive antennas with a large number of small-cell access points (APs). Massive MIMO systems that scale up the number of antennas deployed in each cell site utilize the additional spatial degrees of freedom to serve a large number of devices simultaneously on the same time frequency resource. Meanwhile, small-cell networks rely on a dense deployment of APs and thus reduce the distance between the transmitter and the receiver, as well as the offload traffic from macro-cell BSs.

However, there are still some challenges that need to be addressed by HetNets. In particular, the densification of small-cell APs causes severe inter- and intra-tier interferences and even restricts the system performance. Moreover, forwarding substantial traffic from APs to backbone networks places a heavy burden on the design of backhaul systems. Conventional wired connections through optical fibers are impractical because of their deployment cost. Wireless backhaul is regarded as a more economical and viable alternative and RAN spectrum-based backhauling is considered due to its low sensitivity to propagation conditions, wider coverage, and the reusability of existing equipment [4].

In this paper, authors present a two-tier HetNet, in which a massive MIMO technique is applied at a macro-cell BS overlaid with a second tier of small-cells. The macro-cell BS with a large-scale antenna array is assumed to support mobile macro-cell user equipment units (MUEs), whereas each small-cell AP equipped with one antenna is used to serve, and only serve, the closest static small-cell user equipment unit (SUE). Meanwhile, the macro-cell BS also provides a wireless backhaul for these APs, which shares RAN bandwidth with radio access links. To mitigate interference, authors consider a reverse time-division duplex (R-TDD) transmission protocol [5], where the order of uplink (UL) and downlink (DL) periods in two tiers is reversed. The BS forces its precoding vectors to be orthogonal to the interference subspace so that they avoid interference to the reception at the APs. Regularized zero-forcing (RZF) precoding combined with projection technique is used in DL, whereas joint linear minimum mean square error (LMMSE) detection is used in UL, thereby mitigating inter- and intra-tier interferences significantly [6]. To further improve the system performance, authors optimize the RAN bandwidth allocation between wireless backhaul and access links and time allocation between DL and UL operation intervals. Authors conduct analysis in a large-system regime and derive the deterministic approximation of the system sum rate (SR) which only depends on statistical channel information [7].

The authors' major contribution is to derive the asymptotic results of ergodic DL and UL SRs by considering imperfect CSI, projection technique, and cross-tier interference, based on large system analysis. On the basis of deterministic and asymptotic equivalents, authors formulate an approximate problem as a substitute of the original problem based on ergodic SR, which significantly reduces computational complexity resulting from Monte Carlo averaging. An algorithm is proposed to find the optimal time and bandwidth resource allocation coefficients that maximize system SR. Furthermore, authors investigate important factors that affect bandwidth and time optimization, such as the number of MUEs and the distance between APs and BS.

Authors perform simulations to validate the accuracy of the deterministic equivalent. The results indicate that as the number of MUEs increased, more bandwidth is dedicated to access networks. The results also shed light on the effect of the distance between the APs and BS, which can be optimized given the trade-off between the path loss of the APs and the interference among UE units. Notably, the presented system is more suitable for asymmetric DL/UL traffic considering the influence of the weighting factors on the achievable maximal system SR.

In summary, the authors use large dimensional random matrix theory tools to derive the deterministic equivalents for ergodic system SRs without redundant calculations by considering imperfect CSI, projection technique, and cross-tier interference. The simulation results suggest that the deterministic approximation is accurate and that system performance can be improved via resource allocation, including bandwidth and time optimization, which depends on the number of MUEs, the distance between the APs and BS, and the weighting factors.

References:

[1] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," Comput. Netw., vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," IEEE Trans. Wireless Commun., vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[3] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," IEEE Veh. Technol. Mag., vol. 6, no. 1, pp. 37–43, Mar. 2011.

[4] H. H. Yang, G. Geraci, and T. Q. S. Quek, "Energy-efficient design of MIMO heterogeneous networks with wireless backhaul," IEEE Trans. Wireless Commun., vol. 15, no. 7, pp. 4914–4927, Jul. 2016.

[5] H L. Sanguinetti, A. Moustakas, and M. Debbah, "Interference management in 5G reverse TDD HetNets with wireless backhaul: A large system analysis," IEEE J. Sel. Areas Commun., vol. 33, no. 6, pp. 1187–1200, Jun. 2015.

[6] J. Hoydis, K. Hosseini, S. ten Brink, and M. Debbah, "Making smart use of excess antennas: Massive MIMO, small cells, and TDD," Bell Labs Tech. J., vol. 18, no. 2, pp. 5–21, Sep. 2013.

[7] J. Zhang, C.-K. Wen, C. Yuen, S. Jin, and X. Gao, "Large system analysis of cognitive radio network via partially-projected regularized zero-forcing precoding," IEEE Trans. Wireless Commun., vol. 14, no. 9, pp. 4934–4947, Sep. 2015.

Qin Wang, Ph.D, is an assistant professor with the School of Engineering and Computing Sciences at New York Institute of Technology. She received B.S. and Ph.D degrees in Communication Engineering from Nanjing University of Posts and Telecommunications, China, in 2011 and 2016, respectively. She was a visiting Ph.D. student in the Department of Computer Science, San Diego State University. Her research interests include multimedia pricing, network resource allocation, and Internet of Things. She has published papers in prestigious journals such as IEEE Transactions on Vehicular Technology and IEEE Communications Magazine, in prestigious conferences such as IEEE INFOCOM SDP Workshop.

.

Fine-Grained Venue Discovery from Multimodal Data

A short review for "Category-Based Deep CCA for Fine-Grained Venue Discovery from Multimodal Data"

Edited by Yifang Yin

Y. Yu, S. Tang, K. Aizawa, and A. Aizawa, "Category-Based Deep CCA for Fine-Grained Venue Discovery from Multimodal Data", IEEE Transactions on Neural Networks and Learning Systems, 2018.

Discovering a venue by a user-generated photo is an important yet challenging task for context-aware applications. However, few efforts have been made on fine-grained venue discovery with complicated real images taken by users such as venue photos containing objects and geographic categories. In this work, the authors investigate venue-related multimodal data from Wikipedia and Foursquare, and propose a novel deep learning model, Category-based Deep Canonical Correlation Analysis (C-DCCA), to address the problem of fine-grained venue discovery. Given a user-generated photo, the proposed model is able to perform exact venue search (find the venue where the photo was taken) and group venue search (find relevant venues with the same category as that of the photo), by computing the cross-modal correlation between the input photo and the textual description of venues. Moreover, the authors build a new venue-aware multimodal dataset by integrating Wikipedia featured articles and Foursquare venue photos. Experiments have been conducted on both this new dataset and one publicly available dataset [1] to demonstrate the effectiveness of the proposed approach by comparing to the state-of-the-art cross-modal retrieval methods.

In detail, the authors jointly learn the two tasks, namely the exact venue search and the group venue search, in the same framework. Each venue has an assigned category, text descriptions from Wikipedia, and venue images from Foursquare. During training, textual descriptions and venue images are used to learn the cross-modal correlation to obtain highly correlated textual and visual features in the canonical space. During testing, the task of venue search is per-formed by comparing the visual features of the input photo to the textual features of venues. This work differs from existing methods of visual venue discov-ery, as the authors extend the traditional deep canonical correlation analysis (DCCA) [2] by proposing a new category-based C-DCCA approach for the cross-modal search between images and texts.

In terms of the visual features, the authors utilize the off-the-shelf VGG16 model [3], which is pre-trained on ImageNet, to extract a 4096-dimensional feature for each input photo from the fully connected layer fc15. In terms of the textual features, the authors adopt the Doc2Vec model [4,5], which is an exten-sion to the Word2Vec model, to generate a fixed 300-dimensional feature for each venue article. These features, however, belong to different modalities and cannot be compared directly. Therefore, the authors map the extracted visual features and textual features to a common space, by using two deep convolutional neural networks that consist of three layers. The size of the three hidden layers are 1024, 1024, and 10, respectively. Both the pairwise correlation between photos and texts of the same venue, and the category-based correlation between photos and texts of different venues with the same category, are considered in the objective function.

In the objective function, the covariance matrices of textual feature and visual feature are computed in the same way as in previous work [2]. The cross covariance is computed differently by considering not only the textual and visual features from the same venue, but also the textual and visual features from different venues of the same category. The two factors are next combined linearly using a constant balancing parameter. This balancing parameter plays an important role. A large value improves the pairwise correlation but degrades the category-based correlation. While a small value improves the category-based correlation but degrades the pairwise correlation. The two are conflicting targets that cannot be achieved simultaneously. Thus this balancing parameter should be set empirically to take a tradeoff between the two targets.

Finally, the authors evaluate their proposed method by comparing it with CCA [6], KCCA [7], DCCA [2], C-CCA [8], and C-KCCA [8] on both exact venue search and group venue search. The experimental results show that the proposed C-DCCA method greatly improves the performance of group venue discovery, compared to the state-of-the-art methods. For the exact venue discovery, DCCA outperforms C-DCCA, but the gap can be reduced by leveraging coarse location information associated with the input photo. Additionally, using more images to represent visual aspects of venues also helps to improve the performance of fine-grained venue discovery.

In conclusion, the authors propose an interesting cat-egory-based deep canonical correlation analysis method for fine-grained venue discovery by learning the cross-modal correlation between textual descrip-tions and user generated photos of venues. In addition, the authors build a new venue-aware multimodal dataset, which consists of 19792 photos and 1994 article descriptions for 1994 venues from five cities, for evaluation and perform extensive experiments to compare their method with the state-of-the-art approaches on both exact venue search and group venue search. They also suggest possible solutions for further improvements.

References:

[1] [1] N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G. R. Lanckriet, R. Levy, and N. Vasconcelos, “A new approach to cross-modal multimedia retrieval,” in ACM MM, 2010, pp. 251–260.

[2] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, “Deep canonical correlation analysis,” in ICML, 2013, pp. III–1247–III–1255.

[3] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” CoRR, vol. abs/1409.1556, 2014.

[4] J. H. Lau and T. Baldwin, “An empirical evaluation of doc2vec with practical insights into document embedding generation,” CoRR, vol. abs/1607.05368, 2016.

[5] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, “The stanford corenlp natural language processing toolkit,” in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, 2014, pp. 55–60.

[6] H. Hotelling, “Relations between two sets of variates,” Biometrika, vol. 28, no. 3/4, pp. 321–377, 1936.

[7] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods. Cambridge University Press, 2000.

[8] N. Rasiwasia, D. Mahajan, V. Mahadevan, and G. Aggarwal, “Cluster canonical correlation analysis,” in Proceedings of International Conference on Artificial Intelligence and Statistics, ser. AISTATS’14, vol. 33, pp. 823–831.

Yifang Yin received the B.E. degree from the Department of Computer Science and Technology, Northeastern University, Shenyang, China, in 2011, and received the Ph.D. degree from the National University of Singapore, Singapore, in 2016. She is currently a research fellow with the Interactive and Digital Media Institute at the National University of Singapore. She worked as a Research Intern at the Incubation Center, Research and Technology Group, Fuji Xerox Co., Ltd., Japan, from October, 2014 to March, 2015. Her research interests include geotagged video annotation and retrieval, geo-metadata correction and video summarization.

Mobile Edge Cache Placement with QoE Consideration for Adaptive Video Streaming

A short review for "QoE-Driven Mobile Edge Caching Placement for Adaptive Video Streaming"

Edited by Roger Zimmermann

C. Li, L. Toni, J. Zou, H. Xiong and P. Frossard, "QoE-Driven Mobile Edge Caching Placement for Adaptive Video Streaming," IEEE Transactions on Multimedia, vol. 20, no. 4, April 2018.

Video streaming transmissions are accounting for a steadily increasing fraction of the total Internet traffic. This trend holds for the wired Internet, but even more so in the realm of wireless data communications. Furthermore, video traffic shows distinctive "peak hours" of usage [1], which further exacerbates the stress on the Internet data delivery infrastructure. Caching of video content within a network has long been shown to be an effective solution to reduce the traffic to the origin content server. However, designing a near-optimal caching solution is challenging because many different parameters need to be considered. Many current caching solutions consider videos as generic files, without taking the special characteristics of videos into account. In their work the authors specifically address video caching in the context of the modern and popular dynamic adaptive streaming over HTTP (DASH) technique [2] which stores video content as multiple representations at different qualities and/or resolutions. Additionally, as one of their main contributions, the authors also remove the widely used – but inaccurate – assumption that the encoding bitrate of a video representation is proportional to its quality. Rather, they take into account the different rate-distortion (R–D) characteristics of videos, as is indicative with different content types.

The authors assume an infrastructure setting as it would normally be encountered nowadays, where a base station hosts the DASH server, which is then connected through a number of medium speed links to a set of edge servers. These edge servers provide relatively high speed connections to the mobile clients and they incorporate some limited storage that can be used for video caching purposes. The authors then first formulate the problem of optimally allocating the various representations of a set of videos to the edge caches as an integer linear program (ILP). This formulation considers various constraints such as the storage capacity, the video representations, quality-of-experience (QoE) metrics such as the startup delay and the average video quality and aims to maximize the average video distortion reduction of all users while minimizing the transmission cost of the representation downloads from the base station. The ILP formulation is exponential in its complexity and thus it is not suitable to be executed in an efficient manner in a real world environment.

As another contribution, to alleviate the complexities of solving the ILP, the authors convert the problem into an equivalent set function optimization problem, specifically the authors show that it is a sub-modular maximization problem. Through this transformation the problem can then be solved with a greedy algorithm that exhibits polynomial time complexity. Furthermore, they prove theoretical approximation guarantees of a worst-case performance of the proposed algorithm of 1/2 (1-1/e) $\approx$ 0.316. However, in their simulation results the approximation ratio generally is significantly higher, mostly above 0.95.

The manuscript also includes an extensive experimental section where the authors present simulation results and comparisons to two existing state-of-the-art algorithms and to the optimal solution of the ILP algorithm. An extensive analysis shows the impact of various parameters such as the cache size, the number of users, and the number of edge servers. The authors evaluate a larger system where the number of users, the number of videos, and the video types are scaled up. Overall the proposed solution performs well such that, when the cache size of each edge server is large enough to pre-fetch a large number of representations, the proposed efficient greedy algorithm could almost achieve the same average distortion reduction per user as the optimal solution.

In summary, the authors demonstrate that adaptive videos should be taken into account with different representations and their rate-distortion (R–D) characteristics when designing a video caching infrastructure. The manuscript also shows that an efficient greedy algorithm can achieve near optimal results. A discussion at the end includes some guidelines for system designers which is a useful component of this study.

References:

[1] "Global mobile data traffic forecast update, 2015–2020," Cisco Visual Networking Index Forecast, 2016. [Online]. Available: http://http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visualnetworking-index-vni/mobile-white-paper-c11-520862.pdf.

[2] MPEG-DASH Specification. ISO/IEC 23009-1:2012 Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats, April 2012.

Roger Zimmermann is an associate professor with the Department of Computer Science at the School of Computing with the National University of Singapore (NUS) where he is also an deputy director with the Smart Systems Institute (SSI). His research interests are in both spatio-temporal and multimedia information management, for example, spatio-temporal multimedia, streaming media architectures, georeferenced video management, and mobile location-based services. He has co-authored a book, seven patents and more than two hundred-thirty conference publications, journal articles and book chapters in the areas of multimedia and databases. He has received the best paper award at the ACM IWGS 2016 workshop and the IEEE ISM 2012 conference. He is an investigator with the NUS Research Institute (NUSRI) in Suzhou, China, and he is an Associate Editor of the ACM Transactions on Multimedia journal (TOMM), the Multimedia Tools and Applications (MTAP) journal and the IEEE MultiMedia magazine. He is a Senior Member of IEEE and a Distinguished Member of ACM. For more details, see http://www.comp.nus.edu.sg/~rogerz.

Paper Nomination Policy

Following the direction of MMTC, the Communications – Review platform aims at providing research exchange, which includes examining systems, applications, services and techniques where multiple media are used to deliver results. Multimedia includes, but is not restricted to, voice, video, image, music, data and executable code. The scope covers not only the underlying networking systems, but also visual, gesture, signal and other aspects of communication. Any HIGH QUALITY paper published in Communications Society journals/magazine, MMTC sponsored conferences, IEEE proceedings, or other distinguished journals/conferences within the last two years is eligible for nomination.

Nomination Procedure

Paper nominations have to be emailed to Review Board Directors: Qing Yang (qing.yang@unt.edu), Roger Zimmermann (rogerz@comp.nus.edu.sg), Wei Wang (wwang@mail.sdsu.edu), and Zhou Su (zhousu@ieee.org). The nomination should include the complete reference of the paper, author information, a brief supporting statement (maximum one page) highlighting the contribution, the nominator information, and an electronic copy of the paper, when possible.

Review Process

Members of the IEEE MMTC Review Board will review each nominated paper. In order to avoid potential conflict of interest, guest editors external to the Board will review nominated papers co-authored by a Review Board member. The reviewers' names will be kept confidential. If two reviewers agree that the paper is of Review quality, a board editor will be assigned to complete the review (partially based on the nomination supporting document) for publication. The review result will be final (no multiple nomination of the same paper). Nominators external to the board will be acknowledged in the review.

Best Paper Award

Accepted papers in the Communications – Review are eligible for the Best Paper Award competition if they meet the election criteria (set by the MMTC Award Board). For more details, please refer to http://mmc.committees.comsoc.org/.

MMTC Communications – Review Editorial Board

Multimedia Communications Technical Committee Officers

MMTC examines systems, applications , services and techniques in which two or more media are used in the same session. These media include, but are not restricted to, voice, video, image, music, data, and executable code. The scope of the committee includes conversational, presentational, and transactional applications and the underlying networking systems to support them.