Personal Statement

Huzefa Rangwala
Department of Computer Science
George Mason University

I request that I be considered for promotion to Full Professor in the Department of Computer Science at George Mason University (GMU). The following document summarizes my accomplishments as they relate to research, teaching and service at GMU. It illustrates the significant contributions I have made to my research field, my department and GMU. I have a strong record in research, teaching, and service at the university. Testament to my all-round performance, I was awarded the 2014 Mason Emerging Researcher/Creator/Scholar Award, the 2014 Mason Teaching Excellence Award and the 2017 CS Departmental Service Award. I have provided online materials at the website `http://www.cs.gmu.edu/~hrangwal/promotion`.

1 Background

I am an Associate Professor in the Department of Computer Science at George Mason University. I was granted tenure in 2014 and promoted to Associate Professor. I joined GMU as a tenure-track Assistant Professor in Fall 2008. Prior to that, I graduated from the University of Minnesota, Minneapolis in 2008 with a Ph.D. in Computer Science. My doctoral dissertation was entitled *"Comparative Modeling using Machine Learning,"* and was supervised by Professor George Karypis. I received a M.S. in Computer Science with a minor in Bioinformatics from the University of Minnesota, Minneapolis in 2005. I received a B.E. in Computer Engineering in 2003 from Mumbai University (India).

2 Research

My research interests revolve around the areas of data mining and applications within multiple and varied domains including biology, learning sciences, social media analysis and cyber-physical sciences. I have made contributions across different application areas in (i) educational mining, (ii) microbiome and genome analysis, (iii) social forecasting and event encoding and (iv) cyber-physical systems. My research efforts have lead to successful collaborations and interactions with researchers from vastly different backgrounds; educational psychology, learning sciences, bioengineering, biochemistry, micro-biology, medicine, environmental science, digital history and information security. In these areas, I have collabrated with internal (departmental) and external colleagues, and we have been constantly working on the important problems in the fields.

A unifying theme unifying my research involves development of *s*tructured prediction methods. Structured prediction is a framework for solving classification and regression problems, in which the output/input variables are mutually dependent or constrained. Examples of dependencies and constraints include sequential, combinatorial or spatial structure in the problem domain and capturing these interactions leads to better prediction models. Several real world applications have outputs that are complex, i.e., dependencies between output labels (multi-label, hierarchical classification), or have an internal structure that is described by inter-dependent components (e.g., sequences, trees, networks, dyadic relationships)

As a long term vision, for my research goals I will continue working in the inter-disciplinary areas focusing on development of algorithms and systems for a broad and significant impact. I am leading the development of a center that brings together researchers with expertise in internet-of-things (IOT), analytics and applications in community health. I am also interested in the educational impact of my research and strive to integrate my research activities via teaching and mentoring graduate, undergraduate and high school students.

2.1 Publication Record

I have published **1** edited book, **4** book chapters, **63** peer-reviewed conference proceedings, **28** journal papers and **6** refereed workshop articles. I have **6** journal papers under review and **3** conference papers under review as well. These papers are published in premier journals and conferences of data mining and bioinformatics with highly selective acceptance rates. Since joining GMU, most of my papers are co-authored with my graduate students and some with undergraduate students. According to Google Scholar, my h-index is 19 and i-10 index is 30 with 1417 total citations on July 25, 2017.

2.2 Sponsored Research

I have actively sought funding to support my strong inter-disciplinary research with the Department of Computer Science and jointly with GMU's College of Science and College of Humanities and external institutions. I have been the PI on 3 research awards sponsored by NSF. In particular I am a PI on the highly competitive NSF sponsored BIGDATA project as well as Co-PI on NSF sponsored Cyber-Physical Sciences Project (CPS). I am the recipient of the NSF CAREER award in 2013 on my project related to metagenome analysis. I have also been a PI on a grant funded by the USDA and have been a co-PI on grants sponsored by NRL, NSF, NIH and DARPA. I have brought a total of **$6.98 million** as external funding to GMU as a PI or Co-PI. I have also received equipment funding for Mason as a Co-PI and involved in two other MRI Grants as Senior Personnel.

2.3 Collaborations

In my research efforts at GMU, I have collaborated on papers and grant proposals with several of my Computer Science Department colleagues. I have also collaborated with GMU researchers outside of my department: Jaime Lester (Higher Education), Aditya Johri (Information Tech.), Hemant Purohit (Information Tech.), Raja Kushalnagar (Gallaudet University), Naren Ramakrishnan (Virginia Tech.), Kathryn Laskey (Systems Engineering), Patrick Gillevet (Environmental Sciences), Dan Cohen & Joan Fresny T (History & New Media), Siddhartha Sikdar & Joe Pancrazio (Bioengineering), Robin Couch (Biochemistry) and external researchers: Jasmohan Bajaj (VCU Medical), Ernest Retzel (late, National Center for Genomic Research, New Mexico), Ali Keshavarzian (Rush University Medical Center), George Karypis (Computer Science, University of Minnesota), Xia Ning (NEC Labs), Gaurav Pandey (Mount Sinai School of Medicine) and Andrea Tagarelli (University of Calabria, Italy). All collaborators listed here are co-authors on papers or proposals.

2.4 Students and Impact

In my role as a faculty at George Mason University, I am most proud of all the graduate and undergraduates who have worked with me on different research projects. I have advised a total of **six** students towards completion of their doctoral degrees with **three** scheduled to graduate in the next two years. They all have an extensive and profilic publication record. In particular two of these graduated students, Anveshi Charuvaka and Samuel Blasiak have won the departmental outstanding graduate student awards. I have also served as a thesis advisor for several Masters students and had the joy of exposing our undergraduates and high school students towards their first taste of research. I am a Ph.D. committee member of 24 different doctoral students within several different departments across GMU and Virgina Tech. University.

2.5 Projects

The following summarizes my research projects with more details.

2.5.1 Educational Data Mining An enduring issue in higher education is student retention to successful graduation. National studies report that average six-year graduation rates across higher-education institutions is 59% and have remained relatively stable over the last 15 years. To alleviates these societal challenges there is a *critical need* to develop innovative approaches to enable higher-education institutions retain students, ensure their timely graduation, and are well-trained and workforce ready in their field of study. Failure to do so represents a significant problem as it deprives the country of the highly skilled workforce that it needs to successfully compete in the modern world.

The *objective of this project* is to develop new computational methods to analyze in a comprehensive manner, the large and diverse types of education and learning-related data in order to improve undergraduate education. These methods are motivated by and are designed to address three inter-related critical issues that have a significant impact on college student success. These are: (i) ***academic pathways*** towards successful and timely graduation from the student perspective, (ii) ***effective pedagogy*** by instructors, and (iii) ***retention and persistence*** of students from the institutional and advisor perspective.

We are achieving our objective by developing novel data analysis approaches for (i) modeling and identifying the dynamics of student knowledge and using them to determine the student's state of knowledge and predict the student's grades in future courses, (ii) identifying the aspects of the learning environment that each student prefers, and (iii) identifying the students whose performance is lower than what is expected given their capabilities. Figure 1 provides an overview of our project.

The line of research is *innovative*, modeling the evolution of a student's knowledge state using a dynamical state-space system model is a key innovation of this project. By clustering student knowledge states and pooling data to fit system parameters, the proposed research will capitalize on the power of statistical aggregation to circumvent per-individual data scarcity. This leads to novel *matrix factorization* and *additive latent effect models* for grade prediction. Additional technical innovations include supervised learning approaches for datasets that evolve with time, linear and non-linear multi-task learning approaches, collaborative multi-regression models with controlled grouping of the latent variables, and co-factorization approaches for incorporating diverse types of data.

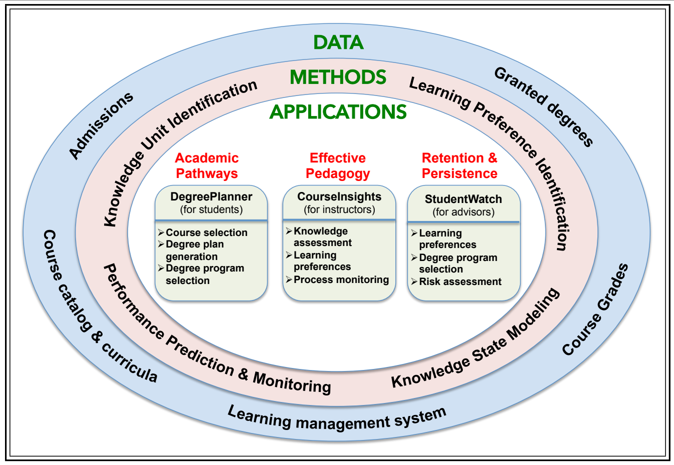

Figure 1: Project Overview.

Key Contributions

- Given a database of (student, course) dyads (i.e., pairs) with associated content features for the course, student, and course instructor, our goal is to predict grades for each student for the next enrollment term. In a comprehensive study, we explored three classical approaches for recommender systems including matrix factorization for the next-term student grade prediction task. The prediction of a student's grade on an untaken course is calculated as the product of vectors in the reduced latent subspace representing the common knowledge subspace. [IEEE Big Data 2015, JEDM 2016, IEEE Computer 2016].

- We developed an additive latent effect model by incorporating instructor information and student academic level information for the next-term grade prediction problem. The strengths of our proposed framework include the ability to enhance the standard MF methods with additional student and course-specific content information that may not be contained within the student-course grade matrix. We also developed course-specific approaches utilize the student's grades from courses taken prior to that course and incorporated additional content. [IEEE ICDM 2017, IEEE DSAA 2017, KDD Workshop 2017]

- We consider that a student's knowledge is continuously being enriched while taking a sequence of courses; and it is important to incorporate this dynamic influence of sequential courses within the academic program structures. Besides improved next-term grade prediction performance this model learns influences between pairs of courses that help in understanding pre-requisite structures within programs and tuning academic program chains. [EDM 2017]

- An automated topic modeling based inference package was developed to analyze publicly available course-related data in the form of catalog descriptions and syllabi. This activity involved investigating the best approach for extraction of ”key concepts” that were taught within a course and how these concepts fit within a program curriculum with regards to the pre-requisite structure. Further, two applications around this approach were investigated. (i) Comparison of course programs across multiple universities and (ii) Comparison of Computer Science undergraduate program with ACM published Curriculum report. [IEEE ICER 2015].

- We develop a personalized linear multiple regression (PLMR) model to predict the grade for a student, prior to attempting the assessment activity. The developed model is real-time and tracks the participation of a student within a MOOC (via click-stream server logs) and predicts the performance of a student on the next assessment within the course offering. [EDM 2016]

- In collaboration with Dr. Lester (Higher Education Department), an instrumental case study was conducted at George Mason University to understand the opportunities, barriers and incentives related to the adoption of learning analytics by higher education institutions and their faculty and staff. A qualitative case study was conducted to learn more about how students perceive and are affected by learning analytics-informed learning management systems. [Higher Ed. 2017, ASHE 2016, AERA 2016]

2.5.2 Bioinformatics

Motivation: Advances in genome-sequencing have transformed the manner of characterizing large populations of microbial communities, that are ubiquitous across several environments. The process of "metagenomics" involves sequencing of the genetic material of organisms co-existing within ecosystems ranging from ocean, soil and the human body. Orthogonally, proteomics and mass spectrometry allow the study of bio-transformations due to these microbial communities in the form of metaproteomes and metabolomes, respectively. *Annotating* microbial sequences (reads or quasi-assembled contigs) within a sample is a challenging task due to the unknown, diverse and complex nature of microbial communities within the different environments. There is a *critical need* to develop new methods that can characterize metagenome data in terms of taxonomy, function and metabolic potential.

One of the fundamental problems in bioinformatics is to build prediction models to automatically classify (annotate) a given instance. Successful annotation using computational models reduces the experimental process time and associated costs. Standard off-the-shelf prediction models involve setting up a multi-class classification problem where the input is a vectorial representation for the input domain. These models ignore the underlying dependencies between the different output class labels (hierarchy) and also ignore the complementary information present across multiple annotation databases. The *objective of this project* is to develop, and make publicly available, novel computational tools that are designed to analyze the microbial communities using the available high-throughput data and correlate the identified taxonomic, functional and metabolic profiles with the clinical phenotypic data using an integrative learning framework. This also results in key innovations within data mining techniques as they relate to hierarchical classification, multi-task learning and semi-supervised learning.

Key Contributions

- We developed a novel cost-sensitive approach (HierCost) that addressed two main issues of large scale hierarchical classification, class imbalance and training efficiency, by extending the flat classification approach using cost sensitive training examples. Although regularization methods which constrain the learned models to be close to neighboring classes according to the hierarchy have been effective, they induce large scale optimization problems which require specialized solutions. Instead, by re-defining the problem from regularization based approach to a cost sensitive classification approach (based on similar assumptions) tends to decouple the training of the models which can be trained in parallel fashion. [ECML 2015, IEEE Big Data 2016]

- Most Hierarchical Classification (HC) approaches rely on the structural relationships for learning complex models to improve the classication performance. We developed approaches that are able to better identify the set of inconsistent nodes that exists within the hierarchy, and thereby improving the HC performance. We also developed an iterative-based rewiring approach for taxonomy modication which unlike previous wrapper based approaches does not require multiple, expensive computations. Our approach is scalable and can be applied to the HC problems with high-dimensional features, large number of classes and examples. The key advantage is that the modied hierarchy can be used with any hierarchical classication approaches. The modied hierarchy in conjunction with a scalable Top-Down HC approach outperforms the flat classiers on 65% of the rare categories (i.e., classes with less than 10 training examples) [DSAA 2015, DSAA 2016, TKDD].

- In order to build effective and interpretable pipelines for prediction of patient phenotype (disease) we propose the use of Multiple Instance Learning (MIL). In addition to classifying patient phenotype, these approaches also identify individual parts of the microbiome that are indicative of that phenotype, leading to better understanding of the disease. We developed a novel, scalable and effective MIL-based computational pipeline to predict patient phenotype from microbiome-related genome data. [IEEE BIBM 2016, TCBB]

- We have also developed random-projection based methods for clustering and use the clustering results for species diversity estimation for the quick analysis of massive datasets. We developed a fast and scalable clustering algorithm for analyzing large-scale metagenome sequence data. Our approach achieves efficiency by partitioning the large number of sequence reads into groups (called canopies) using locality sensitive hashing. These canopies are then refined by using state-of-the-art sequence clustering algorithms. [BICOB 2017, SIAM SDM 2012, SDM 2013, BIBM 2012, BMC Systems Biology]

- We developed approaches to annotate sequences (proteins/DNA) across multiple hierarchies. These classification tasks predict whether an example belongs to the particular class or not. However, instead of training each of these tasks independently (single task learning), the training for all these tasks are combined using the Multi-Task Learning (MTL) approach. The MTL approach is also suited for tasks (classes) that have scarce training examples. This MTL approach leveraged the underlying relationships between the multiple hierarchies and significantly outperformed traditional prediction models for classifying sequence data within multiple hierarchical annotation databases. [IEEE ICDM 2012, TCBB 2014].
- Sequence data does not fit into the standard machine learning methods for classification and clustering which typically operate on a fixed-length representation (called vectors). We developed several approaches for sequence representation, motif discovery, sequence-level and alphabet-level classification. [IJCAI 2011,Bioinformatics 2005, ECML 2014, ECML 2013, SDM 2012]
- We developed multi-label learning approaches for protein function prediction using both the classifier integration framework (ensemble) and data integration (multiple kernel learning) framework. We also developed methods for a related problem for protein function prediction using partial knowledge of known labels. This approach was different from previously developed protein function prediction approaches that assume that the functions associated with instances within the training set (knowledge-base) are complete and fixed. In reality, we only know a subset of the protein's functions, and whether an annotated protein has additional functions is unknown. This type of multi-label prediction problem is referred to as the 'weak label' or 'incomplete class assignment' problem. [**Relevant Publications:** TCBB 2014, TCBB 2015, ACM SIGKDD 2012, ACM BCB 2012, AAAI IJCAI 2013, ECML 2013]

2.5.3 Social Forecasting and Event Encoding Forecasting societal uprisings such as civil unrest movements is an important and challenging problem. Open source data (e.g., social media and news feeds) have been proven to serve as surrogates in forecasting a broad class of events, e.g., disease outbreaks, election outcomes, stock market movements and protests. While many of these works focus on predictive performance, there is a critical need to develop methods that also yield insight by identifying precursors to events of interest.

Our key contributions are summarized as follows:

- We formulate event forecasting and precursor mining for multiple cities in a country as a novel multi-instance learning problem with a nested structure. In addition to forecasting violent events accurately, our approach leads to explainable predictions by identifying related documents in the past that can be viewed as precursors for event. Such evidence helps policy-makers and social scientists to better understand the processes governing the formation of collective identities that turn to events of interest. [KDD 2016]
- We extended the multi-instance learning framework for forecasting violent protest events. The framework is built on the hypothesis that violent crowd behavior tends to have a qualitatively different set of trigger events signaling the occurrence in the future. The framework is significantly advantageous over computer vision techniques that only detect events (not forecast them) and which require the first images of violence to be published. [TKDD 2017]
- We incorporate of spatial and temporal correlations inherent in large-scale societal event occurrences. Using a Multi-Task learning framework we alleviate the data insufficiency problem by learning multiple related tasks simultaneously and restricting all cities to share a common set of features with a consensus model. [CIKM 2017]
- We view the twin problems of event detection and extracting key sentences to enable event encoding and classification in a unified manner as a form of multiple instance learning . This enables us to identify salient event-related sentences from news articles without training labels at the sentence level. In a typical news article, there exist a small set of key sentences that provide detailed information for a specific event. Identifying

these sentences automatically is useful for succinctly summarizing the news article. Highlighting such key sentences within a visualization tool will enable a human analyst to quickly locate important information, rapidly scan the contents of the article, and make timely decisions. Additionally, the key sentences can form the basis of automated event encoding and we extract the final event based on the identified salient sentences. [CIKM 2016]

2.6 Future Research Plans

All the projects described above are continued with federal grant support and Ph.D. studies pursuing thesis work related to those topics. I am also collaborating with Prof. Sikdar in the Bioengineering Department to develop machine learning algorithms for cyber-physical systems. We are developing computational learning approaches for designing a better control system for automating a prosthetic hand as part of a NSF funded Cyber-Physical (CPS) project. There is an emphasis on real-time prediction and analysis of temporal, varying biomedical image datasets. In the future, I plan to collaborat with Prof. Parth Pathak in developing machine learning approaches for gesture recognition to control digital personal assistants (e.g., Amazon Echo) for deaf and hard of hearing individuals.

IOT Analytics

Today, consumer-centric healthcare and wellness solutions are rapidly moving towards the ability to provide healthcare as a service, where the health and wellness information of a person can be seamlessly integrated into various everyday activities. A wide array of ubiquitous sensors in the form of wearable devices (e.g., clothing and wrist devices), smartphones and infrastructure (e.g., low-cost sensors, cameras, WiFi and work station) devices are the chief enabler of this paradigm shift. These sensors collectively referred by Internet-of-Things (IOT) allow for the fine-grained sensing and inferencing of user's context, pyhsiological state, mental health and *individual needs*. Inexpensive hardware, widespread communication networks, decreased computational infrastructure costs and heavy commercial investment have lead to the following exciting projections: (i) 40% year-over-year growth in overall data captured by these sensors , (ii) IoT healthcare segment will be $117 billion and (iii) trillion IoT devices by 2020.

These sensing and detection capabalities coupled with advanced data analytics leading to intelligent intervention and persuation techniques provide an appealing end-to-end solution for healthcare technologies that benefit the public. This exponential growth in terms of devices, data, users and applications from IoT health has left a gaping hole in terms of a unifying ecosysem that systametically addresseses challenges as they relate to "data," "device," "user" and "health application". I am personally interested in data analytic challenges that emerge in this space as part of my next five-year goals. I am also leading the formation of a new Wearable Analytics and Technology for Community Health (WATCH) center at GMU bringing together a team of leading and highly productive interdisciplinary researchers and educators across four different colleges on campus.

3 Teaching

My preference for a faculty position has been driven by a desire to teach, to nurture and to inspire students. I am committed to fostering the synergy between teaching and research by providing an environment for all students to develop intellectually and professionally. I have a successful track record related to curriculum development, introduction of new teaching ideas and mentoring under-represented high school and undergraduate students. My research objectives are integrated with my overall educational goals, activities, and responsibilities as a faculty member at George Mason University (GMU). I am the winner of the *2014 Mason Excellence in Teaching Award*, *2013 Volgenau School of Engineering Teaching Award* and the *2012 Department of Computer Science Teaching Award*.

Pedagogical Activities My lectures focus on interactive and student-centered learning using a wide range of active learning techniques. To train myself with the best active learning strategies, I participated in a semester long course geared towards preparation of future faculty as a graduate student at the University of Minnesota.

Below, I describe two examples of active learning strategies that have been tailored for my graduate and undergraduate classes in computer science.

Jigsaw Activities Jigsaw is a collaborative learning activity. The topic/class module is split into several sub-parts, and a student group is responsible for reading and understanding a sub-part (called topic). Students are expected to read and study the assigned topics at sufficient depth to become experts. During class students convene in expert groups, and discuss the topics among themselves. After strengthening their ideas about the particular topic, students are asked to gather in mixed groups so that every group has at least one expert from all the sub-topics. Collectively each mixed group discusses the entire class module and learning from others. The instructor can have a few minutes of summary activity at the end of the two stages. In terms of preparation, the instructor needs to guide the discussion within different groups by asking challenging questions to individual groups or to the entire class. Jigsaws work well for classes when discussion material is subjective, and the value of it in engineering disciplines is not completely known.

As an example, I set up the jigsaw exercise to allow the class to participate in the discussion and review of several state-of-the-art clustering algorithms in a graduate data mining classes. Two weeks before the class session, students were assigned to read and thoroughly understand one of five research articles. During class time, experts from the same group were asked to convene and discuss the concepts learned. Students were then mixed into groups, so that each group had an article representative to discuss and contrast among themselves the strengths and weaknesses of each of the papers. I specifically, developed questions that mixed groups would answer regarding use of the different algorithms in different test case scenarios.

Competition-Style Assignments and Projects: I design class assignments to supplement the understanding of the subject material. Good assignments are meant to evoke questions that were not raised in class. I believe that a student's performance on assignments is a reflection of the instructor's teaching performance. I have always found that the "learning by doing" idea pays rich dividends in understanding important concepts or principles. I developed assignments where the students develop state-of-the-art algorithms explained in class and compare the performance to available public or commercial versions. The classes I teach have a predictive modeling component. I have participated in blind protein structure prediction competitions (CASP) and data mining competitions like KDD Cup. Motivated by this model, I have implemented these forms of competitions in my bioinformatics and current data mining classes.

In Fall 2016 I developed a data hackathon hosting software [1] and designed assessments where student's would engineer solutions to four challenging data mining problems in classification, clustering, feature selection and relational learning. As part of this competition, the true values were hidden and the developed models had to make a prediction and submit their results to me via a web-service. To improve the models predictive performance, students researched and engineered innovative solutions.

Assessment and Feedback While lecturing, I elicit responses from students by pausing at several points during the lecture to ask questions about the material presented. I use the write-pair-share activity and classroom assessment techniques (CAT) like surveys, muddiest point and surveys. Use of CATs allows me to adapt my class through the semester for a richer learning experience.

Curriculum development and program enhancements. I have contributed to the development of the new B.S. in bioengineering program, bioinformatics track in the B.S. in Applied Computer Science program and the newly formed M.S. in Data Analytic Engineering program (2014). Along with several faculty members and Prof. Offutt as PI we received a $900,000 award from Google to redesign our introductory programming class. The project seeks to respond to the well-documented surge in CS enrollment by creating a self-paced learning environment that blends online learning, automated assessment, collaborative practice, and peer-supported learning.

Encouraging undergraduates and K-12 students. Besides graduate student mentorship, I have involved and engaged undergraduate and high-school students within my research projects. I have mentored a total of **14** undergraduate students and **5** high-school students. These students work on well-defined projects related to my research program. These research projects are designed to expose the students to interdisciplinary research, challenge their intellect and creativity, and provide them with the satisfaction resulting from creating something novel and useful. GMU has several programs that provide undergraduate research fellowships and stipends, and I strive to support student through such initiatives (OSCAR) and independent REU supplements. I advised a high school student from Thomas Jefferson High School, Eric Tao from Fall 2011-Summer 2013. In Fall 2011, Eric won a place in the Siemens Science Regional Finals and is going to be an undergraduate major at MIT. My past experience with this approach of involving undergraduate students into my research programs has been extremely positive and rewarding. I will continue to encourage students to conduct research and pursue academic interests beyond their day-to-day course work.

[1] DMGrader is available here: http://github.com/rangwala/dmgrader

4 Service

Research Community:

I have provided extensive service to the scientific community as a reviewer for journals, conferences, and as Program Committee member of premier conferences in data mining, machine learning and bioinformatics. I take great pride in my role as a panelist and have served on a total of 9 NSF panels from 2008-2017, and a panelist for the National Institute of Health, and NSF Graduate Fellowship Program in 2011, 2012 and 2013. I have also reviewed proposals for the Swiss National Science Foundation (sNSF), North Carolina Biotechnology Center's MRG Program and Genome Alberta grants program. In 2017 I am the Tutorial Chair for the 2017 IEEE ICDM Conference. I am part of the Editorial Board Members for Machine Learning Journal (2017-2020), Pattern Recognition Journal (2017), and Eurasip Bioinformatics Journal (2015-2017). I have served as a Program Chair for the 2015 IEEE Educational Assessment Workshop, NSF CISE Career Workshop in 2014. In 2011 I was the co-Chair of the Workshop on Knowledge Discovery in Health Care and Medicine held in conjunction with ECML in Athens, Greece. In 2013 I was the program co-Chair of the workshop on Data Mining in Bioinformatics held in conjunction with KDD in Chicago, IL. In 2012, I served as the local and sponsorship co-chair of IEEE International Conference in Data Engineering (ICDE) held in Arlington, VA. I received a NSF award to support travel for students attending ICDE 2012. I also organized a career panel and a mentoring network as part of ICDE. I serve as the Senior Program Committee member for IEEE Big Data 2017 conference. I have been a program committee member for several conferences including CIKM, AAAI, KDD, ECML, ICDM and SDM and reviewer for many journals including *Transactions in Bioinformatics and Computational Biology, Transactions on Neural Networks and Learning Systems, BMC Bioinformatics, Bioinformatics, Transactions in Knowledge Discovery and Engineering, Statistical Analysis and Data Mining, Structural Biology and Journal of Bioinformatics & Computational Biology.*

George Mason University:

I have performed a significant amount of service for GMU, its student and local community. Besides my primary appointment with Computer Science, I have affiliate positions with the Bioengineering Department and Bioinformatics & Computational Biology Department. The departmental service role I cherish the most is faculty mentorship. I have served as an official mentor for **six** of my departmental colleagues, assisting them in advice related to achieving research goals, creating a new research program, writing NSF grants (CAREER), recruiting students and teaching. I have also assisted two Bioengineering faculty in writing their NSF CAREER grants (see CV). I am Chair of the Web Committee within the Department since 2016 and responsible for the development and desing of our new departmental webpage. I have been part of the recruitment committee for my department in 2014, 2016, 2017 and served on the recruitment committee for Information Technology in 2015 and Bioengineering in 2013.

I have served as part of the Computing Committee, Web Committee and the Undergraduate Studies Committee within my department. At the University level, in 2016 I served on the iPASS Leadership Committee as well as the Computing Infrastructure Committee. From Fall 2012-2014 I served on the Students as Scholars Assessment and Program Design Subcommittee. I participated and presented in two ACE Scholars Program events, two student presentations at the Celebration of Achievements events with two student presentations at the Students as Scholars Program.

5 Concluding Remarks

This statement, along with the its accompanying materials, summarizes my major accomplishments in academia at GMU. I believe it demonstrates my strong leadership in multiple research areas, a commitment to scholarly work, teaching, student mentorship, and services to my research community and to Mason. I have constantly published at top-tier conferences and journals. I have been actively supervising Ph.D. students.vI have also been actively and extensively serving to my research community in various capacities. I have also offered high levels of service to my department. I believe I have a record of strong and consistent accomplishment and commitment in all relevant areas since I joined GMU in August 2008 and tenured in 2014, and have provided evidence that I will continue such efforts in the future. I am deeply grateful for the support and mentorship that I have received from my colleagues at Mason. I respectfully request that the faculty recommend me for promotion to full professor.