From Syntax to Semantics: Exploring the Frontiers of NLP at Wilfrid Laurier

JIASHU ZHAO AND GURUNAMEH SINGH CHHATWAL
INTELLIGENT SYSTEMS FOR PREDICTIVE INFORMATION RETRIEVAL AND DATA MODELING LAB
WILFRID LAURIER UNIVERSITY, WATERLOO, CANADA
JZHAO@WLU.CA, CHHA8740@MYLAURIER.CA

I. INTRODUCTION

The Intelligent Systems for Predictive Information Retrieval and Data Modeling (INSPIRE) Lab at Wilfrid Laurier University is dedicated to advancing the fields of information retrieval (IR), machine learning (ML), and large language models (LLMs) through cutting-edge research and innovation. The lab focuses on developing new algorithms, models, and techniques to enhance how information is processed, understood, and retrieved across diverse domains. By integrating modern machine learning approaches and LLMs with traditional IR methods, the lab seeks to improve the accuracy, efficiency, and relevance of search and recommendation systems. Research at the lab includes the application of natural language processing (NLP), deep learning, and large-scale data analytics to tackle real-world challenges in information retrieval, personalization, and knowledge discovery. The lab explores the potential of LLMs in advancing tasks like semantic search, question answering, and content generation. With a collaborative environment, the lab fosters interdisciplinary research and supports the development of advanced technologies that impact fields such as search engines, e-commerce platforms, and recommender systems.

II. RESEARCH AREAS

The INSPIRE Lab at Wilfrid Laurier University focuses on a broad range of research areas that combine information retrieval (IR), machine learning (ML), recommender systems, and large language models (LLMs). Some of the key research areas include:

- Probabilistic Models for Information Retrieval: This research explores the application and enhancement of probabilistic frameworks in information retrieval (IR), focusing on the development of models that assess the likelihood of document relevance.
- Personalized Information Retrieval Techniques: This area investigates methods for adapting IR systems to individual user needs by incorporating user profiles, preferences, and contextual information.
- Graph-Based Approaches in Information Retrieval: Focused on leveraging graph theory and network structures to enhance IR systems, this research examines how to model and exploit the relationships between documents, queries, and users.
- Deep Learning for Textual Data Analysis: This research explores the application of advanced deep learning techniques for analyzing unstructured textual data. We investigate the use of neural networks, including CNNs, RNNs, and transformer-based models like BERT, to extract semantic features and improve the retrieval of relevant documents from large-scale corpora.
- Representation Learning for Textual Data: This research focuses on the development of robust embedding methods for textual data to better capture semantic relationships within language. We investigate approaches such as contextual embeddings (BERT, GPT) to improve the representation of text for downstream retrieval and classification tasks.
- Text Classification for Information Retrieval: Investigating methods to automatically categorize text into predefined labels, this research enhances text classification techniques for various applications, including categorization, spam detection, AI generated text detection, and etc.
- Analysis and Mitigation of Ranking Biases: This research examines the inherent biases that influence the ranking of search and recommendation results, including algorithmic, content, and user-related biases. We focus on developing methods to identify and mitigate these biases, ensuring more fair, diverse, and unbiased ranking outcomes across a variety of applications, from search engines to recommender systems.
- User Behavior Modeling in Recommender Systems: This area investigates techniques for analyzing and modeling user interaction and feedback within recommender systems. Our research aims to develop models that predict user preferences based on clickstream data, user behavior, and contextual cues, enhancing the personalization and relevance of recommendations over time.
- User-Centric Approaches: Focusing on improving the user experience in information retrieval and recommendation systems, this research studies how users interact with IR systems. We explore how to optimize interfaces and algorithms to enhance satisfaction, engagement, and performance, through user-centric models and feedback loops.
- Applications of Large Language Models in Information Retrieval: This research investigates and improves the use of large pre-trained language models (e.g., BERT and ERNIE) to improve various IR tasks such as semantic search and contextual document retrieval. We

explore the fine-tuning of these models to address domain-specific challenges, with an emphasis on enhancing contextual understanding and search relevance.
- Video Search and Retrieval: This research focuses on enhancing video content retrieval by integrating textual, visual, and audio data using multi-modal and multi-task learning approaches. We apply graph neural networks (GNNs) to capture relationships between video elements and user queries, improving search accuracy and efficiency.
- AI Forensics and Authorship Attribution: With the increase in the usage of LLMs in digital environments we focus on attributing the generated content to its neural author. We leverage pre-trained language models and multiple specialized features to distinguish between machine and human generated content. Furthermore, we apply this to more novel applications in public safety and secure digital communication.

III. Projects and Achievements

A. Current Projects

1) Large Language Model Enhanced Retrieval: By tapping into the generative and contextual understanding strengths of LLMs, the project aims to improve the alignment between user queries and retrieved information. The integration of advanced language models enables deeper comprehension of user intent and context, addressing ambiguities and enhancing the overall search experience. At its core, this initiative seeks to redefine how queries are understood and processed in retrieval systems, ensuring more accurate and relevant results across diverse use cases. By combining the scalability and adaptability of LLMs with robust retrieval strategies, the project sets the foundation for next-generation information systems that can dynamically adapt to complex user needs, offering groundbreaking improvements in relevance and precision.

2) Adaptive smart information systems: A key future direction for intelligent systems lies in their ability to dynamically adapt to user needs and contexts while deeply understanding the content they process. The vision is to create adaptive smart systems that go beyond simple information retrieval to interpret, organize, and present data in ways that are contextually relevant and aligned with user intent. These systems will evolve with changing trends, behaviors, and environments, ensuring continuous relevance and value.

By advancing methods for content understanding and user intent modeling, these systems will bridge the gap between raw data and actionable insights. They will enable highly personalized and intuitive experiences across search, recommendation, and decision-making platforms. Through real-time adaptation to user interactions and data nuances, these intelligent systems will redefine how users engage with information, consistently delivering precise and context-aware results across diverse applications.

3) Advanced framework to combat AI misinformation: Misinformation and fake content is a severe issue in the current society as we are succumbed to digital venues for our information consumption. Most of these areas are not regulated by any authority which gives an edge to the bad actors with an intent to spread false narrative. With the widespread use of AI and the rapid development in the quality of work that it can mimic as a human, it is serves as a great tool for proliferation of false content. It may not always be the ill intentions of a user of generative models to spread false content but due to their non-deterministic nature they are prone to hallucinate and make up facts which may not be true. The volume by which this type of content has seeped into our online ecosystems is alarming and requires frameworks that help automate the detection process.

We are currently working on a framework that uses meta and transfer learning strategy to study the common embeddings that can be used to detect fake content as well as AI generated content. The study is comprehensive of both a general strategy as well as specializing to particular use cases.

This will also lead to generation of a baseline dataset that is first of its kind for machine generated false content detection. It will be useful to further research in this and related areas. Moreover, we try to introduce a novel social based prompting strategy to create such datasets which helps in explainability of the models actions on how it learned and reduces bias.

B. Past Achievements

1) Innovative Approaches for Unbiased Learning and Ranking: Bias in information retrieval (IR) and recommender systems is a significant challenge that impacts the accuracy, fairness, and effectiveness of search results and recommendations. This project explored innovative methods to identify, understand, and mitigate various types of biases, addressing both their causes and their effects on system performance. The studies conducted within this project proposed frameworks and algorithms designed to create more equitable and reliable IR and recommender systems.

The first study focused on presentation biases, including position and context biases, which influence user interactions with ranked results. By leveraging propensity score estimation techniques, this study adjusted the probabilities of user actions to enable a more accurate and unbiased learning process. The second study advanced this work by introducing an unbiased deep neural network framework that incorporated a reweighting mechanism and a bias-correction layer to address implicit feedback biases. This approach improved the system's predictive capabilities by compensating for biases present in user-generated data.

The third study extended the scope to encompass biases from multimedia content, trust, and page layout in search result pages (SERPs). It introduced the Bias-Agnostic Whole-Page Unbiased Learning to Rank (BAL) algorithm, which used causal discovery methods to identify and correct hidden biases across multiple components of the page. Unlike traditional models limited to position bias, the BAL algorithm integrated causal inference to detect and mitigate more complex bias interactions, significantly enhancing the relevance and fairness of the rankings. These combined efforts represent a comprehensive approach to developing bias-aware systems, ensuring fairer and more

effective IR and recommendation outcomes.

2) Graph-based Search Personalization: Personalized search has become increasingly important as data volumes grow and the need to provide users with highly relevant content becomes critical. This project explored graph-based approaches to search personalization, introducing innovative techniques to better capture and utilize user preferences across diverse data types and search contexts.

The first study developed a Latent Dirichlet Allocation (LDA)-based topic-graph probabilistic model for web search personalization. By modeling both user interests and disinterests using latent topics derived from user behavior (e.g., clicks and skips), the model provided a more nuanced approach to ranking search results. This work also introduced novel methods for refining negative user profiles, ensuring that non-interests were effectively captured and incorporated into the ranking process. The result was a system capable of delivering search results that align closely with users' preferences and query intents.

Recognizing the unique challenges of video search personalization, the second study proposed a Graph Neural Network (GNN)-based framework to enhance personalized video search. To the best of our knowledge, this is the first work analyzed real user behavior specific to video search. By integrating semantic representations of queries and video content with a hierarchical aggregation strategy, the model improved the quality of learned representations. The framework also utilized a multi-task learning approach, allowing it to incorporate additional signals to optimize model parameters. These innovations resulted in more accurate personalized video search rankings, addressing the distinct demands of video-centric search scenarios while advancing the broader field of search personalization.

3) Term Association Modeling In Probabilistic Information Retrieval: Traditional probabilistic information retrieval models, such as BM25, often assume that terms are independent, limiting their ability to account for relationships between terms. This project systematically addressed this limitation by developing enhanced models that incorporate term associations, proximity, and contextual relevance to improve retrieval performance.

The first study introduced the CRoss TErm Retrieval (CRTER) model, which incorporated the concept of "Cross Terms" to model term proximity. In this approach, the occurrence of a query term was assumed to influence its surrounding text, with this influence attenuating as the distance increased. By integrating bigram and n-gram cross terms into basic probabilistic weighting models, CRTER captured the nuanced relationships between query terms and their proximity in the text. The results demonstrated significant improvements in retrieval effectiveness, particularly for queries with closely associated terms.

Building on this foundation, the project further enhanced proximity-based retrieval models by integrating contextual relevance into term proximity analysis. Traditional methods often treat all query term associations equally, neglecting the varying importance of different term associations. This study introduced measures to estimate contextual relevance between query terms using top-ranked documents. These measures were then integrated into a context-sensitive proximity model, allowing the retrieval system to distinguish and prioritize more meaningful term relationships. This refinement resulted in improved retrieval accuracy, addressing the limitations of conventional proximity-based methods and advancing the state of probabilistic information retrieval.

4) Enhance collaborative filtering by contrastive learning: This project tackled the limitations of collaborative filtering (CF) models, such as noise and sparse supervision, by incorporating contrastive learning techniques to redefine item representation learning.

The first study introduced a disentangled contrastive framework to enhance item representations by learning discriminative features. This approach addressed challenges related to noise and insufficient supervision in traditional CF models, resulting in significantly improved recommendation accuracy and robustness.

Building on this foundation, the second study proposed the Hypergraph Contrastive Collaborative Filtering (HCCF) framework, which leveraged self-supervised contrastive learning to enhance GNN-based CF models. This framework captured complex relationships among items and users, improving the discrimination power of the model. The HCCF framework demonstrated superior performance on benchmark datasets, even under conditions of sparse interaction data.

5) Machine Generated Text detection in a Multilingual Setting: In today's era of generative AI models determining data authenticity is a unique challenge. With the widespread use of openly available tools like chatGPT, gemini, metaAI etc., we now have machine generated text that is not only grammatically correct but is at par with humans. This poses a great question to accountability and attribution of the content available to us. The legitimacy of the data is a severe issue in areas such as academic integrity, misinformation, social media spamming, deepfakes etc. Thus we provide advanced methodologies to detect machine generated text in multilingual environment thus making the model more generalized and adaptable to current developments in LLMs.

We propose two unique approaches to solve this problem: a Pretrained language model (PLM) based approach and a Stylometric feature selection based method. The former is based on a multilingual PLM backbone to extract nuanced features as semantic embeddings which are inputs to a multi-layered bagged classification model. This leverages the deep semantic understanding between various languages thus providing a cohesive latent information representation layer. This strategy outperforms the state-of-the-art multilingual models for MGT detection.

The later is an approach where we curate popular and handcrafted multilingual features which are subject to rigorous selection strategy. These features are then fed to various sequential neural networks and ensemble classification methodologies to obtain a lightweight classifier that provides a great accuracy in low resource areas. Furthermore, this strategy succeeds all the non-transformer based methods by a decent

margin. Our proposed methodologies consistently outperform popular MGT detection tools such as GPTZero, DetectGPT, etc.

6) E-Commerce Data Understanding: Understanding data in e-commerce is critical for enhancing customer experiences, optimizing inventory, and improving business operations. This project explored innovative frameworks and models to address key challenges in understanding e-commerce queries, product categorization, and purchase prediction.

The first study proposed the Dynamic Product-aware Hierarchical Attention (DPHA) framework for understanding e-commerce query intents. By leveraging dynamic session information, including prior queries and clicked product categories, DPHA maps customer queries into relevant product categories. Using hierarchical attention mechanisms, it learns both query-level and session-level representations, significantly improving the classification of complex queries in dynamic shopping contexts.

To address challenges in product categorization, the project introduced the Neural Product Categorization (NPC) model. This model focuses on fine-grained categorization, tackling issues like blurred category boundaries and evolving product definitions. By employing character-level convolutional embedding and spiral residual layers, NPC extracts intricate context representations from product content. The model is trained with weak labels derived from customer behavior logs and outperforms traditional methods in classifying products into fine-grained categories.

Lastly, the project developed the Graph Multi-Scale Pyramid Networks (GMP) framework to predict users' future purchases. GMP captures multi-resolution temporal patterns and dynamic dependencies among product categories. By using a multi-scale pyramid modulation network, recalibration gating, and a context-graph neural network, GMP models hierarchical temporal factors and dynamic category dependencies. Real-world experiments demonstrated GMP's ability to outperform state-of-the-art predictive models, providing businesses with more accurate insights into customer purchase behavior.

7) Pattern Recognization: Pattern mining is essential in uncovering meaningful insights from large datasets, enabling the identification of complex relationships and trends that are otherwise not readily apparent. This project explored novel methodologies for mining and interpreting patterns in diverse domains, contributing to advancements in understanding dynamic data relationships.

The first study introduced diverging patterns, a new type of contrast pattern that highlights significant frequency changes between two datasets moving in opposite directions. To quantify these changes, the project proposed a diverging ratio and represented patterns with a four-dimensional vector. An efficient algorithm was developed to mine these diverging patterns, revealing insights that traditional frequent pattern mining approaches often overlook. This method provided a deeper understanding of contrasting trends, enabling the discovery of novel and actionable insights across various applications.

The second study focused on mining associations within multivariate time-series data, with a particular emphasis on gene expression analysis. By identifying significant changes and measuring marginal change rates, the project employed a propositional confirmation-guided rule discovery method to uncover associations among temporal changes. This approach facilitated the construction of gene interaction networks and clustering of genes with similar temporal behavior, offering valuable insights into biological processes and enabling a more comprehensive understanding of gene regulation dynamics. These methodologies demonstrate the potential of advanced pattern recognition techniques to drive discovery in complex, data-rich domains.

8) Data Understanding and Prediction for Healthcare : This research focuses on advancing healthcare by developing data-driven solutions that utilize state-of-the-art machine learning and Bayesian inference techniques. The aim is to enhance personalized medical recommendations and diagnostic accuracy by leveraging diverse data types and innovative methodologies.

The first study introduced a Bayesian-based prediction model for recommending personalized medical tests to patients. By analyzing patterns in medical data through Bayesian inference, the model integrates contextual factors such as timing and interaction types to deliver more accurate predictions. This approach is particularly effective for sparse datasets and providing a robust tool for personalized healthcare recommendations.

The second study developed a deep learning approach for skin cancer classification that combines image data with structured clinical information. Utilizing convolutional neural networks (CNNs) to extract features from skin lesion images, the model incorporates patient-specific data such as age, gender, and lesion history to enhance diagnostic precision. This multimodal approach significantly outperforms image-only methods, demonstrating the benefits of integrating multiple data sources for comprehensive and reliable medical diagnoses.

Together, these studies highlight the transformative potential of advanced data analysis in improving personalized healthcare and medical decision-making, paving the way for more effective and patient-centered solutions in the field.

9) Fraud Detection in E-commerce: Fraud activities, such as spam reviews and fake shopping behaviors, disrupt customer decision-making and harm business reputation. The project introduced a novel approach using heterogeneous graph neural networks (GNNs) to detect fraud more effectively by modeling complex relationships among diverse data types, including users, items, devices, and reviews.

The proposed framework, C-FATH, tackled critical challenges in fraud detection, such as structural inconsistencies (e.g., fraudsters often being surrounded by legitimate users) and content inconsistencies (e.g., disparate features across product categories). By employing community-based filtering, the framework grouped nodes exhibiting similar behaviors, while similarity-based sampling ensured that only relevant neighbors were considered during fraud detection. These methods enhanced the system's ability to accurately identify fraudulent entities. Combining heterogeneous graph modeling with embeddings from homogeneous graphs and

leveraging a multi-task learning strategy, the framework achieved superior results compared to traditional methods and existing GNN-based models.

IV. Partnerships and Collaborations

We actively engage in collaborations with industry partners to conduct research that addresses the practical needs and interests of users, as well as with academic institutions to advance cutting-edge, theoretical research. Our current and past codllaborating organizations are listed as follows:

- Baidu Search Science Lab
- JD Data Science Lab
- Yahoo! Labs
- York Unviersity
- Unviersity of Waterloo
- Beijing Technology and Business University

V. Funding & Acknowledgments

This lab is currently supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant, Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Early Career Supplement Grant, and Wilfrid Laurier University Start-up Grant.

VI. Vision for the future

The overarching vision for these projects is to pioneer the development of intelligent, data-driven systems that transform how we interact with and benefit from information across diverse domains. By leveraging cutting-edge machine learning, artificial intelligence, and innovative LLM-based modeling techniques, the future lies in creating adaptive, personalized, and scalable solutions that address complex real-world challenges.

We aim to advance the understanding of dynamic patterns, relationships, and behaviors within data, whether in healthcare, e-commerce, information retrieval, or recommender systems. Future systems will seamlessly integrate diverse data types, handle sparsity and noise with resilience, and adapt to evolving contexts to offer insights and recommendations that are precise, fair, and impactful.

Looking forward, the focus will be on creating interpretable, ethical, and human-centric AI systems that not only enhance decision-making but also foster trust and inclusivity. By uniting advancements in graph-based learning, contrastive techniques, and deep neural networks, we envision a future where technology empowers individuals and organizations to make smarter decisions, uncover hidden insights, and improve quality of life across sectors. This vision aims to shape the next generation of intelligent systems, driving innovation, fostering collaboration, and solving the most pressing problems in an increasingly data-rich world.

VII. Representative Publications

- He, L., Zhao, J., Gu, Y., Elbaz, M., & Ding, Z. (2024, January). A bias study and an unbiased deep neural network for recommender systems. In *Web Intelligence* (Vol. 22, No. 1, pp. 15-29). IOS Press.
- Chu, X., Hao, C., Wang, S., Yin, D., Zhao, J., Zou, L., & Li, C. (2024, May). LT 2 R: Learning to Online Learning to Rank for Web Search. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)* (pp. 4733-4746). IEEE.
- Li, Q., Su, L., Zhao, J., Xia, L., Cai, H., Cheng, S., ... & Yin, D. (2024, March). Text-Video Retrieval via Multi-Modal Hypergraph Networks. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining* (pp. 369-377).
- Mao, H., Zou, L., Zheng, Y., Tang, J., Chu, X., Zhao, J., ... & Yin, D. (2024, May). Whole Page Unbiased Learning to Rank. *In Proceedings of the ACM on Web Conference 2024* (pp. 1431-1440).
- Ren, X., Xia, L., Zhao, J., Yin, D., & Huang, C. (2023, July). Disentangled contrastive collaborative filtering. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1137-1146).
- Shao, W., Chen, X., Zhao, J., Xia, L., Zhang, J., & Yin, D. (2023, November). Sequential recommendation with user evolving preference decomposition. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region* (pp. 253-263).
- Zhao, J., Huang, J. X., Deng, H., Chang, Y., & Xia, L. (2021). Are topics interesting or not? An LDA-based topic-graph probabilistic model for web search personalization. *ACM Transactions on Information Systems (TOIS)*, 40(3), 1-24.
- Chu, X., Zhao, J., Fan, X., Yao, D., Zhu, Z., Zou, L., ... & Bi, J. (2022, April). Contrastive Disentangled Graph Convolutional Network for Weakly-Supervised Classification. In *International Conference on Database Systems for Advanced Applications* (pp. 722-730). Cham: Springer International Publishing.
- Zhao, B., Yang, B., Li, Z., Li, Z., Zhang, G., Zhao, J., ... & Bao, H. (2022, October). Factorized and controllable neural re-rendering of outdoor scene for photo extrapolation. In *Proceedings of the 30th ACM International Conference on Multimedia* (pp. 1455-1464).
- Xia, L., Huang, C., Xu, Y., Zhao, J., Yin, D., & Huang, J. (2022, July). Hypergraph contrastive collaborative filtering. In *Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval* (pp. 70-79).
- Chu, X., Zhao, J., Zou, L., & Yin, D. (2022, July). H-ERNIE: A multi-granularity pre-trained language model for web search. In *Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval* (pp. 1478-1489).
- Wei, W., Huang, C., Xia, L., Xu, Y., Zhao, J., & Yin, D. (2022, February). Contrastive meta learning with behavior multiplicity for recommendation. In *Proceedings of the fifteenth ACM international conference on web search and data mining* (pp. 1120-1128).

- Zhang, L., Shi, L., Zhao, J., Yang, J., Lyu, T., Yin, D., & Lu, H. (2022, February). A gnn-based multi-task learning framework for personalized video search. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining* (pp. 1386-1394).
- Huang, C., Chen, J., Xia, L., Xu, Y., Dai, P., Chen, Y., ... & Huang, J. X. (2021, May). Graph-enhanced multi-task learning of multi-level transition dynamics for session-based recommendation. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 5, pp. 4123-4130).
- Wang, L., Li, P., Xiong, K., Zhao, J., & Lin, R. (2021, October). Modeling heterogeneous graph network on fraud detection: A community-based framework with attention mechanism. In *Proceedings of the 30th ACM international conference on information & knowledge management* (pp. 1959-1968).
- Huang, C., Zhao, J., & Yin, D. (2021, April). Purchase intent forecasting with convolutional hierarchical transformer networks. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)* (pp. 2488-2498). IEEE.
- Wu, X., Chen, H., Zhao, J., He, L., Yin, D., & Chang, Y. (2021, March). Unbiased learning to rank in feeds recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (pp. 490-498).
- Jian, F., Huang, J. X., Zhao, J., Ying, Z., & Wang, Y. (2020). A topic-based term frequency normalization framework to enhance probabilistic information retrieval. *Computational Intelligence*, 36(2), 486-521.
- Soboroff, I. (2021). Overview of TREC 2021. In *TREC*.
- Zhao, J., Chen, H., & Yin, D. (2019, November). A dynamic product-aware learning model for e-commerce query intent understanding. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 1843-1852).
- Chen, H., Zhao, J., & Yin, D. (2019, November). Fine-grained product categorization in e-commerce. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 2349-2352).
- Huang, C., Wu, X., Zhang, X., Zhang, C., Zhao, J., Yin, D., & Chawla, N. V. (2019, July). Online purchase prediction via multi-scale modeling of behavior dynamics. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2613-2622).
- Huang, C., Zhang, C., Zhao, J., Wu, X., Yin, D., & Chawla, N. (2019, May). Mist: A multiview and multimodal spatial-temporal learning framework for citywide abnormal event forecasting. In *The world wide web conference* (pp. 717-728).
- Huang, J. X., He, B., & Zhao, J. (2018). Mining authoritative and topical evidence from the blogosphere for improving opinion retrieval. *Information Systems*, 78, 199-213.
- Jian, F., Huang, J. X., Zhao, J., & He, T. (2018, June). A new term frequency normalization model for probabilistic information retrieval. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 1237-1240).
- Miao, J., Huang, J. X., & Zhao, J. (2016). TopPRF: A probabilistic framework for integrating topic space into pseudo relevance feedback. *ACM Transactions on Information Systems (TOIS)*, 34(4), 1-36.
- Jian, F., Huang, J. X., Zhao, J., He, T., & Hu, P. (2016, July). A simple enhancement for ad-hoc information retrieval via topic modelling. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 733-736).
- Zhao, J., Huang, J. X., & Ye, Z. (2014). Modeling term associations for probabilistic information retrieval. *ACM Transactions on Information Systems (TOIS)*, 32(2), 1-47.
- Zhao, J., & Huang, J. X. (2014, July). An enhanced context-sensitive proximity model for probabilistic information retrieval. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 1131-1134).
- Zhao, J., Huang, J. X., Hu, X., Kurian, J., & Melek, W. (2012, October). A Bayesian-based prediction model for personalized medical health care. In *2012 IEEE International Conference on Bioinformatics and Biomedicine* (pp. 1-4). IEEE.
- Zhao, J., Huang, J. X., & Wu, S. (2012, August). Rewarding term location information to enhance probabilistic information retrieval. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval* (pp. 1137-1138).
- Zhao, J., Huang, J. X., & He, B. (2011, July). CRTER: using cross terms to enhance probabilistic information retrieval. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 155-164).
- Lupu, M., Zhao, J., Huang, J. X., Gurulingappa, H., Fluck, J., Zimmermann, M., ... & Tait, J. (2011, November). Overview of the TREC 2011 Chemical IR Track. In *TREC*.
- Rohian, H., An, A., Zhao, J., & Huang, X. (2009, November). Discovering temporal associations among significant changes in gene expression. In *2009 IEEE International Conference on Bioinformatics and Biomedicine* (pp. 419-423). IEEE.
- An, A., Wan, Q., Zhao, J., & Huang, X. (2009, November). Diverging patterns: discovering significant frequency change dissimilarities in large databases. In *Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 1473-1476).
- Zhao, J., Huang, X., Ye, Z., & Zhu, J. (2009). York University at TREC 2009: Chemical Track. In *TREC*.
- Gu, Y., Zhao, J., Liang, D., & Xu, Z. (2007, September). Immunity diversity based multi-agent intrusion detection. In *2007 IEEE Congress*

on Evolutionary Computation (pp. 3404-3409). IEEE.

- Zhao, J., Huang, J. X., & Hu, X. (2013). BPLT+: A Bayesian-based personalized recommendation model for health care. *BMC genomics*, 14, 1-10.
- Gu, Y., Zhou, B., & Zhao, J. (2008). PCA-ICA ensembled intrusion detection system by pareto-optimal optimization. *Information Technology Journal*, 7(3), 510-515.
- Yu, G., Jiashu, Z., & Tianjun, Z. (2006). Distributed Intrusion Detection Method Based on the Diversity of Immunity. *JOURNAL-XIAN JIAOTONG UNIVERSITY*, 40(10), 1052.
- Singh, S., & Zhao, J. J. (2022, November). A deep skin cancer classification approach using image and structured information. In *2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (pp. 930-933). IEEE.
- Zhao, J., McGrath, S., Huang, J. X., Wu, J., & Wu, S. (2018). Extracting Relevant Information from Big Data to Anticipate Forced Migration. *Highlighting the Importance of Big Data Management and Analysis for Various Applications*, 71-78.

Contact Information

Director:
Dr. Jiashu Zhao
E-mail: jzhao@wlu.ca
Address: 75 University Ave W,
Waterloo, ON N2L 3C5

Website:
https://inspire-research.github.io