**本科学历国际学生直接攻读博士学位个人陈述**

|  |  |  |  |
| --- | --- | --- | --- |
| 基本信息 | | | |
| 姓名 | 高贵焕 | **申请编号** | T202300855 |
| 院校及成绩信息（必填项） | | | |
| 请填写本科毕业高校及毕业院校在全世界的排名或在本国的排名或毕业专业在世界的排名。可参考泰晤士高等教育、QS、US News、ARWU等机构当年世界大学排名、各国高校排名或学科排名  成均馆大学在2022年QS 世界大学排名第99。 | | | |
| 请填写本科原始GPA和原始GPA分制：4.22/4.5 | | | | |
| 请填写转换后的本科4分制的GPA分值: 3.75/4 | | | | |
| 请填写本科平均成绩，并说明评分体系  本科平均成绩4.22，大学成绩评分从高到低依次为A+（=4.5）、A0（=4.0）、B+（=3.5）、B0（=3.0）、C+（=2.5）、C0（=2.0）、D+（=1.5）、D0（=1.0）、F。 | | | |
| 请填写本科成绩排名/总人数: 排名前4% | | | |
| 申请陈述 | | | |
| 在高中时能成为全校第一，在大学时又能持续保持高学分，可以取得这样的成绩是因为我一直坚持“讲效率”和“抓本质”的学习与做事原则。我一直坚持“高效”地学习知识的“本质”，在相同的时间我追求尽可能多的学习知识，并集中于学习本质知识。  高考结束后，我取得了很优异的成绩，可以走进很好的大学。但当时的我觉得大学并不是一个能够有效学习的地方，所以便放弃读大学，毅然决然地去了发展迅速且潜力巨大的中国。在陌生的环境里，边积累多种多样的经验边学习语言要比毫无目的地上大学更有益。在中国，我的汉语也以比别人更快的速度进步着，一年内便达到了当时就读的渤海大学语言学院的最高水平。学习汉语的那段时间，我发现大学好像与我想象中的大为不同，是个可以丰富知识，接触各类优秀的人才，并获得积极影响的地方。所以我决定回国重新准备高考，并顺利考上了韩国排名前列的大学——成均馆大学。  因为我一直追求优秀，所以大学期间也如愿获得了成绩奖学金。在学习上我始终尽自己的最大努力做到最好。后来，我接触到了编程。追求“效率”的编程逻辑与我的做事原则相契合，因此我被编程的魅力所吸引。我在学习编程后便马上利用所学的知识，将日常生活中需要反复处理的事情进行自动化，以提高效率。其中，包括自动举报网络恶意帖子、自动翻译、日记整理以及时间统计等程序［1］。  深陷编程和自动化魅力中的我为了实现更多的自动化，便更加深入地在人工智能专业学习。这期间，我在一家名为DSAIL的研究室实习了7个月，参与了有关微博数据感情分析的研究。在此过程中，我查找并整理了30篇与Covid-19和感情分析方向相关的主要论文以及重要概念，寻找可用的研究数据，并联系相关研究人员获得了大量研究数据[2]。在研究的过程中，我学会了论文的查找方法、阅读方法、合作及报告研究成果的方法。此外，还通过与浙江大学的合作，参与了浙江大学论文[3]的后续研究。当时参与研究的主题是通过聚集移动性数据（Mobility data）和微博数据预测 covid-19 的确诊者。我将从浙江大学获得的移动性数据，与我们分析的微博数据相整合，来探索更优质的预测covid19确诊者的方法。  在大学期间我虽然进行了多项有关训练深度学习模型的研究，但却始终觉得深度学习并不属于我。我只是在用别人做好的工具，其实并没有完全了解深度学习模型的学习过程。直到我在Sooyoung Cha教授的课上接触到了在深度神经网络(DNNs)上应用白盒测试方法考察深度学习模型性能的相关研究[4]。该研究提出了能够反映深度学习模型内部神经元活性化程度的指标——神经元覆盖率(NC)，并提议向满足该指标的方向训练深度学习模型。那一瞬间，我觉得这便是我想要研究的领域。作为一名深度学习工程师，不仅要通过结果指标和数学运算过程寻找答案，还需要观察深度学习模型的内部，了解知识的传达方式，并从中获得灵感。这就是我理解的深度学习应该追求的一个“核心”。  因此，我仔细阅读了论文《DeepXplore》[4]，并分析了该论文的研究动机和研究实施过程。为了更加了解该领域，我又进一步阅读了基于《DeepXplore》的延伸研究《DeepGauge》[5]、《test4deep》[6]和《DLfuzz》[7]等论文以及提出NC改善指标的几篇论文。我发现，研究者们并不是单纯地只关注神经元是否活跃，还利用神经元之间的关系、激活界限、以及为利用模型的训练指标提出建议。随后，我又阅读了Harel-Canada et al.[8]、Yu et al. [9]、Yan et al.[10]等学者对深度学习模型白盒测试研究的效果表示质疑的论文。此外，我还通过模仿Harel-Canada et al.[8]和yan et al.[9]公开的程序代码，掌握了在Resnet、Densenet等模型中测定NC，以及PGD(Projected Gradient Descent)对抗攻击和C&W对抗攻击的方法，还尝试了DeepGauge[6]论文中提出的NC的改善版本(K-Multisection NC,TopkNC,Top-K Neuron Pattern)，并实现了DLfuzz[7]论文中体现的模糊测试。通过这一系列的研究与实操，我发现这个领域还处于初期阶段，有很多内容极具研究价值有待进一步深入研究。  因此，博士就读期间我想从深度学习模型覆盖率之间的相关性开始研究。在现有研究中，提及深度学习模型覆盖率之间的相关关系的论文仅有2篇，也只是涉及到逻辑关系[10]和简单的相关关系[9]。但由于覆盖率是用来表示深度学习模型中的层和神经元如何学习的，因此有必要对特定层或神经元之间的相关关系进行研究。但遗憾的是，目前还没有相关研究。所以，我可以通过研究现有覆盖率之间的关系，或采用混合覆盖率、添加模糊测试等技术，以提出更好的测试方法，并帮助学界更好理解覆盖率与深度学习模型之间的关系。希望能通过这样的研究，发现DNNs内部神经元之间的相互作用，最终提出被广泛认可的与对抗攻击DNNs测试技术（PGD或C&W）效果相当的新的覆盖率。除此之外，还希望能像Guo et al.[11]一样，将这种方法应用到没有使用过覆盖率的深度学习模型上，以提高相关模型的性能。  毕业后，我打算从事有关验证深度学习模型、提高模型准确度的工作，想成为深度学习模型决策运算法则方面的工程师。这样的我可以准确判断DNN的错误部分，并解释错误产生的意义，可以在需要预防此类错误的工程技术岗位工作。我认为这将会是深度学习研究的一个重点。  *[1]在* [*https://github.com/GwiHwan-Go*](https://github.com/GwiHwan-Go) *中可检查实操结果。*  *[2] Hu, Y., Huang, H., Chen, A., & Mao, X.-L. (2020). Weibo-COV: A Large-Scale COVID-19 Social Media Dataset from Weibo. Association for Computational Linguistics (ACL).* [*https://doi.org/10.18653/v1/2020.nlpcovid19-2.34*](https://doi.org/10.18653/v1/2020.nlpcovid19-2.34)  *[3] Wu, M., Li, C., Shen, Z., He, S., Tang, L., Zheng, J., Fang, Y., Li, K., Cheng, Y., Sheng, G., Liu, Y., Zhu, J., Ye, X., Chen, J., Chen, W., Li, L., Sun, Y., & Chen, J. (2022). Use of temporal contact graphs to understand the evolution of COVID-19 through contact tracing data. Communications Physics, 5.*  *[4] Pei, K., Cao, Y., Yang, J., & Jana, S. (2019). Deepxplore: Automated whitebox testing of deep learning systems. Communications of the ACM, 62(11), 137–145. https://doi.org/10.1145/3361566*  *[5] Ma, L., Juefei-Xu, F., Zhang, F., Sun, J., Xue, M., Li, B., … Wang, Y. (2018). DeepGauge: Multi-granularity testing criteria for deep learning systems. In ASE 2018 - Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering (pp. 120–131). Association for Computing Machinery, Inc.* [*https://doi.org/10.1145/3238147.3238202*](https://doi.org/10.1145/3238147.3238202)  [6] *[3] Yu, J., Fu, Y., Zheng, Y., Wang, Z., & Ye, X. (2019). Test4Deep: An effective white-box testing for deep neural networks. In Proceedings - 22nd IEEE International Conference on Computational Science and Engineering and 17th IEEE International Conference on Embedded and Ubiquitous Computing, CSE/EUC 2019 (pp. 16–23). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/CSE/EUC.2019.00013*  [7] *[4] Guo, J., Jiang, Y., Zhao, Y., Chen, Q., & Sun, J. (2018). DLFuzz: Differential fuzzing testing of deep learning systems. In ESEC/FSE 2018 - Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (pp. 739–743). Association for Computing Machinery, Inc.* [*https://doi.org/10.1145/3236024.3264835*](https://doi.org/10.1145/3236024.3264835)  *[8] Harel-Canada, F., Wang, L., Gulzar, M. A., Gu, Q., & Kim, M. (2020). Is neuron coverage a meaningful measure for testing deep neural networks? In ESEC/FSE 2020 - Proceedings of the 28th ACM Joint Meeting European Software Engineering Conference and Symposium on the Foundations of Software Engineering (pp. 851–862). Association for Computing Machinery, Inc.* [*https://doi.org/10.1145/3368089.3409754*](https://doi.org/10.1145/3368089.3409754)  *[9] Dong, Y., Zhang, P., Wang, J., Liu, S., Sun, J., Dai, T., … Dong, J. S. (2020). There is limited correlation between Coverage and Robustness for Deep Neural Networks. In CEUR Workshop Proceedings (Vol. 2657, pp. 1–9). CEUR-WS. Retrieved from* [*https://arxiv.org/abs/1911.05904*](https://arxiv.org/abs/1911.05904)  *[10] Yan, S., Tao, G., Liu, X., Zhai, J., Ma, S., Xu, L., & Zhang, X. (2020). Correlations between deep neural network model coverage criteria and model quality. In ESEC/FSE 2020 - Proceedings of the 28th ACM Joint Meeting European Software Engineering Conference and Symposium on the Foundations of Software Engineering (pp. 775–787). Association for Computing Machinery, Inc. https://doi.org/10.1145/3368089.3409671*  [11] J. Guo, Q. Zhang, Y. Zhao, H. Shi, Y. Jiang and J. Sun, "RNN-Test: Towards Adversarial Testing for Recurrent Neural Network Systems," in IEEE Transactions on Software Engineering, vol. 48, no. 10, pp. 4167-4180, 1 Oct. 2022, doi: 10.1109/TSE.2021.3114353. | | | |
| 请提供汉语水平考试HSK等级及成绩(如果所申请项目要求提供HSK成绩)  新HSK6级:246分（总分300分） | | | |
| 母语：韩语 | | | |
| 请提供托福或者雅思成绩：托福105 | | | |
| 请列出曾获荣誉、奖项、特殊贡献或发表的论文、专著。对于论文或专著，请提供完整的引文  获奖：  National Scholarship (13 times), Korea Student Aid Foundation, Korea (Mar. 2018 ~ Jun. 2022)  Student Success Scholarship, Sungkyunkwan University, Korea (Sep. 2020)  Munhang Scholarship (2 times), Sungkyunkwan University, Korea (Mar. 2018 ~ Sep. 2019)  Scholarship for academic excellence, Sungkyunkwan University, Korea (Sep. 2018) | | | |
| 个人声明（必填项） | | | |
| 1、你是否曾被留校察看、停学、退学或接受处分？（请填写是或否）  如果是，请解释。  否 | | | |
| 2、你是否曾因任何犯罪被定罪（或现在正在针对你进行任何指控）？（请填写是或否）  如果是，请解释。  否 | | | |
| 1. 请问你是否拥有中国香港、中国澳门、中国台湾地区的身份证件或身份证明?（请填写是或否）   否  \*请注意，持有上述身份证件或证明的申请人，不可作为国际学生申请我校。 | | | |
| 我承诺个人陈述中的信息真实准确，由我本人独立完成。我知晓如有虚假不实信息，将被清华大学取消申请、录取及学习资格。    签名 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 时间 2022年11月22日 | | | |

请将此表打印并手写签名，扫描成一个PDF文件上传至申请系统。