

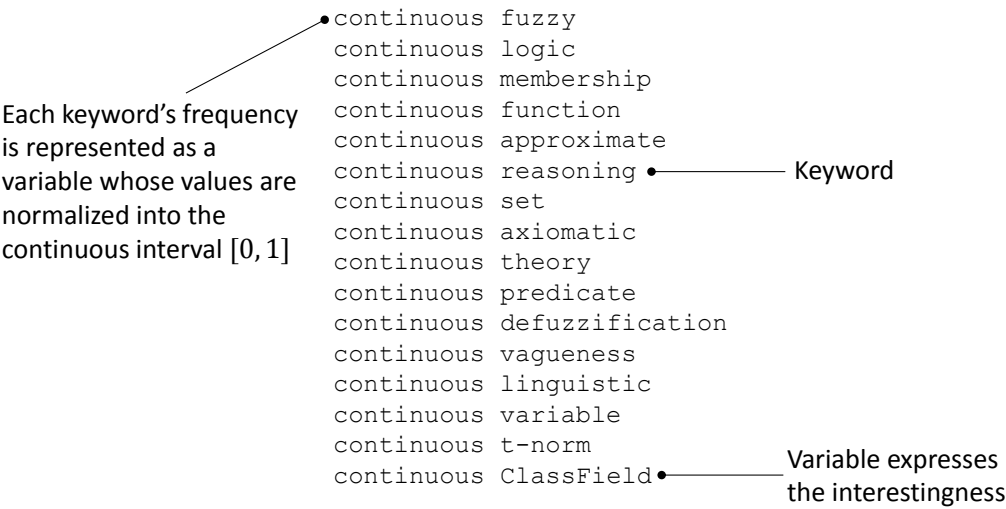
Appendix A

Training Datasets

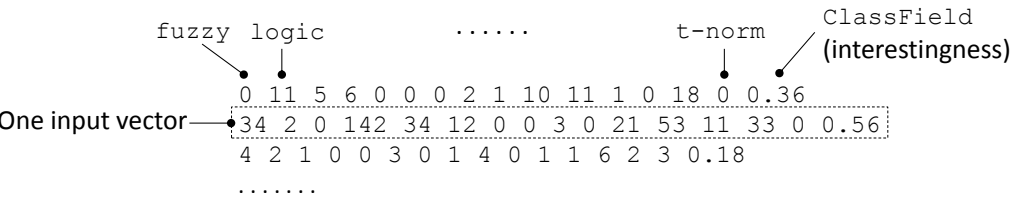
This appendix provides the details of training datasets used in all experiments of this research. All training datasets are available in directory `TrainingDataSets` at the following url

<https://github.com/TranXuanHoang/InformationFilteringSystem>

Each training dataset consists of a data definition file `infofilter.dfn` defining the list of keywords that describe the user’s research area, a data file `infofilter.dat` containing input vectors for training neural networks and a returned-after-training file `filterAgent.ser`. For example, in the dataset for training an agent majoring in *fuzzy logic*, the `infofilter.dfn` file that defines variables expressing the list of keywords (see table A.2) and the interestingness is as follows



File `infofilter.dat` saves input vectors with each in one line as the following description



File `filterAgent.ser` contains all information about the backpropagation neural network and the self-organizing map that have been trained. In order to avoid having to train the neural networks again, just place this file in the same directory where the information filtering application (see appendix B) is run in the next time.

Table A.1: List of research areas and corresponding raw keywords used to generate training datasets. **(RK) Number of Raw Keywords.** 8 training datasets are generated to train 8 agents which major in 8 different research areas. Raw keywords are given to describe these areas. The terminology *raw keywords* refers to originally given keywords (phrases) that are not split into smaller words.

#	Research Area	(RK)	List of Raw Keywords
1	Artificial Intelligence	14	artificial intelligence, machine learning, deep learning, reinforcement learning, neural network, intelligent agent, knowledge, reasoning, learning, planning, decision, natural language processing, classification, robotic
2	Neural Networks	18	neural network, multilayer perceptron, backpropagation, self-organizing map, convolutional neural network, recurrent neural network, weight update, gradient descent, adaptive learning, competitive learning, supervised learning, unsupervised learning, online learning, batch learning, hidden layer, unit, activation, feed-forward
3	Deep Learning	11	deep learning, multiple processing, multilayer neural network, convolutional neural network, backpropagation algorithm, supervised learning, reinforcement learning, recurrent neural networks, stochastic gradient descent, deep feedforward, pattern recognition
4	Computer Networks	25	computer networks, Internet, application layer, presentation layer, session layer, transport layer, network layer, data link layer, physical layer, router, gateway, bridge, port, socket, IP address, MAC address, TCP/IP, UDP, protocol, wireless LAN, DNS, link state routing algorithm, distance vector routing algorithm, multimedia, security
5	Natural Language Processing	24	natural language processing, text classification, word, lexical, dictionary, bilingual corpus, query, regular expression, n-gram tagging, syntax, context free grammar, feature structure, information extraction, information retrieval, part-of-speech tagging, hidden Markov models, precision, recall, question answering, automata, parse tree, Chomsky normal form, CKY algorithm, machine translation
6	Logic	22	logic, validity, completeness, soundness, truth table, true, false, formal deduction, propositional logic, first-order logic, negation, conjunction, disjunction, implication, all quantifier, some quantifier, syntax, semantics, premise, conclusion, tautology, theorem
7	Fuzzy Logic	10	fuzzy logic, membership function, approximate reasoning, fuzzy set, axiomatic fuzzy set theory, fuzzy predicate logic, defuzzification, vagueness, linguistic variable, t-norm
8	Modal Logic	12	modal logic, knowledge operator, belief operator, Kripke model, possibility, necessary, modal axiom, proof system, sound, complete, possible world semantic, consistent

Table A.2: Keyword lists for generating training datasets. **(K) Number of Keywords.** Each list of keywords is created from its corresponding list of raw keywords in table A.1 by splitting compound raw keywords into single words and keeping only one word if the word is duplicated.

#	Research Area	(K)	List of Keywords
1	Artificial Intelligence	19	artificial, intelligence, machine, learning, deep, reinforcement, neural, network, intelligent, agent, knowledge, reasoning, planning, decision, natural, language, processing, classification, robotic
2	Neural Networks	25	neural, network, multilayer, perceptron, backpropagation, self-organizing, map, convolutional, recurrent, weight, update, gradient, descent, adaptive, learning, competitive, supervised, unsupervised, online, batch, hidden, layer, unit, activation, feed-forward
3	Deep Learning	20	deep, learning, multiple, processing, multilayer, neural, network, convolutional, backpropagation, algorithm, supervised, reinforcement, recurrent, networks, stochastic, gradient, descent, feedforward, pattern, recognition
4	Computer Networks	33	computer, networks, Internet, application, layer, presentation, session, transport, network, data, link, physical, router, gateway, bridge, port, socket, IP, address, MAC, TCP/IP, UDP, protocol, wireless, LAN, DNS, state, routing, algorithm, distance, vector, multimedia, security
5	Natural Language Processing	42	natural, language, processing, text, classification, word, lexical, dictionary, bilingual, corpus, query, regular, expression, n-gram, tagging, syntax, context, free, grammar, feature, structure, information, extraction, retrieval, part-of-speech, hidden, Markov, models, precision, recall, question, answering, automata, parse, tree, Chomsky, normal, form, CKY, algorithm, machine, translation
6	Logic	25	logic, validity, completeness, soundness, truth, table, true, false, formal, deduction, propositional, first-order, negation, conjunction, disjunction, implication, all, quantifier, some, syntax, semantics, premise, conclusion, tautology, theorem
7	Fuzzy Logic	15	fuzzy, logic, membership, function, approximate, reasoning, set, axiomatic, theory, predicate, defuzzification, vagueness, linguistic, variable, t-norm
8	Modal Logic	18	modal, logic, knowledge, operator, belief, Kripke, model, possibility, necessary, axiom, proof, system, sound, complete, possible, world, semantic, consistent