

香川高等専門学校		開講年度	令和04年度 (2022年度)	授業科目	自然言語処理
科目基礎情報					
科目番号	4047		科目区分	専門 / 選択	
授業形態	授業		単位の種別と単位数	履修単位: 2	
開設学科	情報工学科 (2018年度以前入学者)		対象学年	5	
開設期	通年		週時間数	2	
教科書/教材	教科書: 天野 真家 著 「自然言語処理」				
担当教員	篠山 学				
到達目標					
自然言語処理の基礎的な内容を理解する。また、自然言語処理の基本的なツールを利用できるようになる。さらに、機械学習の代表的な手法や評価手法を利用できるようになる。					
ルーブリック					
	理想的な到達レベルの目安		標準的な到達レベルの目安		未到達レベルの目安
評価項目1	自然言語処理に必要な技術や理論を理解し、目的に応じてメリット・デメリットを説明できる。		自然言語処理に必要な技術や理論を理解している。		自然言語処理に必要な技術や理論をほとんど説明できない。
評価項目2					
評価項目3					
学科の到達目標項目との関係					
教育方法等					
概要	自然言語処理の基礎的な内容を理解する。また、自然言語処理の基本的なツールを利用できるようになる。さらに、機械学習の代表的な手法や評価手法を利用できるようになる。				
授業の進め方・方法	形態素解析、構文解析などから情報検索や機械翻訳までを学習する。各技術の理論を説明し、確認のための問題を解く。理論の次にGoogle Colaboratoryを使用した実践を行う。具体的には、自然言語処理ツールや機械学習とその評価手法のプログラミングを行う。使用言語はpythonとなる。pythonの基本的な文法は説明しないので、履修する場合は事前に勉強しておくこと。				
注意点	オフィスアワー木曜日7,8限目				
授業の属性・履修上の区分					
<input type="checkbox"/> アクティブラーニング		<input type="checkbox"/> ICT 利用		<input type="checkbox"/> 遠隔授業対応	
<input type="checkbox"/> 実務経験のある教員による授業					
授業計画					
		週	授業内容	週ごとの到達目標	
前期	1stQ	1週	自然言語処理の基礎	自然言語処理とは何か、自然言語処理の意義や役割について身近な例を取り上げながら理解させる。D2:1-3	
		2週	形態素解析	コンピュータに自然文を理解させるために用いられる技術である形態素解析について、その意義や仕組みを理解させる。構文解析、意味解析についても同様に理解させる。D2:1-3	
		3週	形態素解析のプログラム	自然言語処理ツールとして公開されているMeCabなどのツールを利用できる。D2:1-3	
		4週	構文解析	トップダウン法を使って構文解析ができる。D2:1-3	
		5週	構文解析の手法	CKY法を使って構文解析ができる。D2:1-3	
		6週	コーパスとn-gram	コーパスと言語モデルであるn-gramについて理解できる。D2:1-3	
		7週	構文解析のプログラム	構文解析ツールCaboChaやトークナイザSentencepiecesなどのツールを利用できる。D2:1-3	
		8週	[前期中間試験]		
	2ndQ	9週	試験問題の解答 意味解析	意味解析の目的や手法について理解できる。D2:1-3	
		10週	意味解析	選択制限(意味素や用例、連想関係)による意味解析ができる。また、シソーラスについて理解できる。D2:1-3	
		11週	意味解析のプログラム	単語をベクトル化するword2vecを利用できる。D2:1-3	
		12週	pythonのライブラリ (matplotlib、sklearn)と機械学習	pythonのライブラリ (matplotlib、sklearn)を利用できる。機械学習の概要が理解できる。D2:1-3	
		13週	サポートベクターマシン	線形サポートベクターマシンやカーネルサポートベクターマシンを実行し、評価できる。D2:1-3	
		14週	クロスバリデーション法	機械学習の評価方法のひとつであるクロスバリデーションを理解できる。D2:1-3	
		15週	決定木、LightGBM、ランダムフォレスト	機械学習の決定木、LightGBM、ランダムフォレストを実行し、評価できる。D2:1-3	
		16週	[前期期末試験]		
後期	3rdQ	1週	情報検索	情報検索や質問応答について、基本的・基礎的な知識と仕組みを習得させる。D2:1-3	
		2週	情報検索の手法	TF・IDF法を理解し、実際に重みを計算できる。D2:1-3	
		3週	情報検索の手法	転置インデックス法やベクトル空間モデルを使って情報検索できる。D2:1-3	

4thQ	4週	学習データの分割とスケール変換、次元削減	学習データの分割とスケール変換ができる。 主成分分析(PCA)を用いて学習データの次元を削減できる。D2:1-3
	5週	再現率と適合率	情報検索の評価に用いられる再現率や適合率を学ぶ。D2:1-3
	6週	再現率と適合率	情報検索のタスクによって再現率と適合率の重要度が変わること理解する。D2:1-3 自然言語処理関係の文献を読み、理解できる。E1:1,2
	7週	情報抽出	固有表現抽出の仕組みを理解できる。 質問応答システムの仕組みを理解できる。D2:1-3
	8週	教師なし学習と評価方法	様々な教師なし学習(k-means, t-SNE)について理解できる。 学習データのOne-hot-Encodingについて理解できる。 混同行列を用いた評価方法について理解できる。D2:1-3
	9週	[後期中間試験]	
	10週	試験問題の解答 機械翻訳	自然言語処理の最大の応用分野の一つである機械翻訳について学ぶ。D2:1-3
	11週	機械翻訳の手法	コーパスを利用した翻訳手法(EBMT, SMT, NMT)について学ぶ。D2:1-3
	12週	GiNZAとspacyとplotly	pythonの自然言語処理ライブラリspacyのひとつであるGiNZAを利用できる。 また、GiNZAを用いて、発話コーパスをベクトル化し、さらに機械学習により分類したものをplotlyで可視化できる。D2:1-3
	13週	機械翻訳の評価	機械翻訳システムの評価方法について人手評価と自動評価について学ぶ。D2:1-3
	14週	機械翻訳の評価	自動評価の指標(BLEU, WER)と計算方法について学ぶ。D2:1-3 自動評価の指標と計算方法の演習を行う。D2:1-3
	15週	テキストマイニング	テキストマイニングに必要なpandasや正規表現、スクレイピングについて実際にプログラムしながら学ぶ。D2:1-3
	16週	[後期期末試験]	

モデルコアカリキュラムの学習内容と到達目標

分類	分野	学習内容	学習内容の到達目標	到達レベル	授業週
評価割合					
		試験	課題提出	合計	
総合評価割合		70	30	100	
基礎的能力		35	15	50	
専門的能力		35	15	50	