

# 大連理工大学

姓名 \_\_\_\_\_

学籍番号 \_\_\_\_\_

学部 \_\_\_\_\_

クラス \_\_\_\_\_

科目番号 \_\_\_\_\_

担当教員 \_\_\_\_\_

科目名: 自然言語処理 問題種類: A 試験形式: 閉 巻

所属学部: 国際情報ソフトウェア学部 試験の実施日: 2020 年 8 月 28 日

問題用紙合計ページ数 6

	一	二	三	四							合計
配点	40	15	30	15							100
得点											

得点	
----	--

一、

問一

次の (1)-(20)のそれぞれに対して、最も適切に当てはまるものを選択肢(A)-(D)の中から一つ選んで答えよ。なお、該当する選択肢がない場合は「(D)該当なし」を選択せよ。

(1) 典型的な自然言語処理の順序として正しいものはどれか？

- (A) 形態素解析→構文解析→意味解析→応用に応じた処理  
(B) 形態素解析→意味解析→構文解析→応用に応じた処理  
(C) 応用に応じた処理→形態素解析→構文解析→意味解析  
(D) 該当なし

(2) 句構造文法  $G=(V, \Sigma, P, S)$ における  $P$  は何か？

- (A) 記号の集合  
(B) 構文規則の集合  
(C) 終端記号の集合  
(D) 該当なし

(3) 文脈依存文法 (csg) での構文規則の書き方について、一番関連が近いのはどれか？

- (A)  $A \rightarrow aA$   
(B)  $\alpha_1 A \alpha_2 \rightarrow \alpha_1 \beta \alpha_2$   
(C)  $A \rightarrow \beta$   
(D) 該当なし

(4) クラスとインスタンスの関係を示す例として最も適当なものはどれか？

- (A) 座席と乗客  
(B) アルファベットと a、b、c、...  
(C) 海と魚  
(D) 該当なし

(5) 日本語において、句「テレビを」がなりやすい格はどれか？

- (A) 動作主格  
(B) 対象格  
(C) 場所格  
(D) 該当なし

(6) 結束性で扱われる関係として正しいものはどれか？

- (A) 主語、動詞の関係 (B) 原因と結果の関係  
(C) 照応関係 (D) 該当なし  
新情報、旧情報
- (7) 修辞構造論について間違った記述はどれか？  
(A) 階層的な構造 (B) 話し手を強調する  
(C) 「中心」と「周辺」の関係 (D) 該当なし
- (8) 再現率、適合率について、間違った記述はどれか？  
(A) 適合率とは選ばれたものがどれだけ正しいかを示す (B) 適合率とは選ぶ/選ばないの正しさの割合を示す  
(C) 再現率とは正解がどのくらい選ばれるかを示す (D) 該当なし  
再現率は、正解のうち、どの程度が検索にヒットするかを示す。
- (9) 次の検索の各手順について、正しい順番はどれか？  
1. 上位 M 位までの文書を出力する。  
2. 検索質問と全ての文書ベクトルの類似度を計算する。  
3. あらかじめ各文書に対する文書ベクトルを計算しておく。  
4. 類似度の大きい順にソートする。  
5. 検索質問を、検索質問ベクトルに変換する。  
(A) 3→5→2→4→1 (B) 5→3→2→4→1  
(C) 3→2→5→4→1 (D) 該当なし
- (10) 事例を 2 クラスに分離する平面を求めることを原理とする機械学習の方式について、一番関連が深いものを選び。  
結合重み付きネットワークでの値の伝播。  
(A) 重回帰分析 線形の判定式による (B) ニューラルネットワーク  
(C) 決定木 (D) SVM
- (11) ニューロンについて、「ニューロンは他のニューロンからの信号を重み付きで総和する」という性質をなんと呼ぶか。  
(A) 非線型閾値性 (B) 急降下学習性  
(C) 線型加算性 (D) 該当なし
- (12) 発話の手順について正しい順番はどれか？  
1. 相手の行動の予測  
2. 相手の行動の制御  
3. 目的の設定  
4. 実際の発話の生成  
(A) 3→2→1→4 (B) 1→2→3→4  
(C) 3→1→2→4 (D) 該当なし
- (13) 機械学習（テキスト・マイニング）手法とその分類として正しい組はどれか？  
(A) K 平均法とクラスタリング (B) ランダムフォレストとクラスタリング 分類  
(C) 混合正規分布と回帰 (D) SVR と分類 回帰  
クラスタリング
- (14) 日本語と英語で表記の対応が間違っているものはどれか？  
(A) 最急降下法 steepest descent (B) 語用論 pragmatics

method

(C) 修辞構造論 RST

cata

(D) 後方照応 anaphora

(15)用語とその説明の組で間違っているものはどれか？

(A) 構文解析：入力に対してどのように文法規則が適用されているかを解析すること

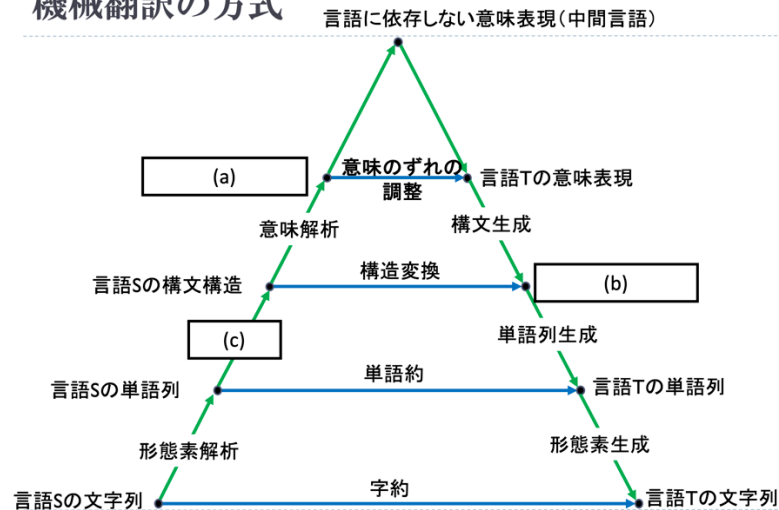
(B) 例文主導翻訳：種々の特有な表現を例文（例句）として集め、規則化する

(C) 非決定性有限オートマトン：遷移可能な状態が複数個存在する場合がある有限オートマトン

(D) 該当なし

(16)機械翻訳の方式を説明する下図の(a),(b),(c)について、最も正しい記述を選べ。

### 機械翻訳の方式



(A) (a)言語 S の意味表現、(b)言語 T の構文構造、(c)単語列解析

(B) (a)言語 S の意味表現、(b)言語 T の構文構造、(c)構文解析

(C) (a)言語 S の意味表現、(b)言語 T の構文構造、(c)構文解析

(D) 該当なし

(17)重要語句を抽出する方式  $tf \cdot idf$  について、間違った記述はどれか？

(A) 原則として、出現頻度が多い語ほど重要な語である

(B) 特定の文章に出現する頻出語ほど重要である

(C)  $idf$  値は多くの文章に表れる語ほど値が高い

(D) 該当なし

CNN(Convolutional Neural Network)とRNN(Recurrent Neural Network)

(18)略語 RNN と CNN と、それぞれの元の語について正しい対応はどれか？

ただし neural network を NN と略す。

(A) RNN: recursive NN, CNN: conceptual NN

(B) RNN: rhetorical NN, CNN: coherent NN

(C) RNN: remote NN, CNN: contents NN

(D) RNN: recurrent NN, CNN: convolution NN

(19)ある目的の対話のプランニングにおいて「信念」、「欲求」、「意図」についての記述のうち、正しいものはどれか？

(A) 複数の「欲求」は矛盾してはならない

(B) 「信念」は矛盾が生じない間は事実として扱われる

- (C) 「意図」は複数あってもいい (D) 該当なし

(20) 以下はベイズの定理である。空欄(a)、(b)に当てはまる記述を選択せよ。

$$P(C_1, C_2, \dots, C_n | w_1, w_2, \dots, w_n) = \frac{P(C_1, C_2, \dots, C_n) \boxed{(a)} P(w_1, w_2, \dots, w_n | \boxed{(b)})}{P(w_1, w_2, \dots, w_n)}$$

- (A) (a) - (b)  $w_1, w_2, \dots, w_n$  (B) (a)  $\div$  (b)  $C_1, C_2, \dots, C_n$   
 (C) (a) + (b)  $w_1, w_2, \dots, w_n$  (D) (a)  $\times$  (b)  $C_1, C_2, \dots, C_n$

得点	
----	--

二、

問二

天候	温度(° F)	湿度(%)	強風	クラス
晴れ	75	70	真	開催
晴れ	80	90	真	中止
晴れ	85	85	偽	中止
晴れ	72	95	偽	中止
雨	69	70	偽	開催
雨	71	80	真	中止
雨	65	70	真	中止
雨	75	80	偽	開催
雨	68	80	偽	開催
雨	70	96	偽	開催

上図の開催／中止に関する事例の集合  $S$  について、各問に答えよ。ただし、各計算において、必要なら  $\log_2 3 = 1.58$ 、 $\log_2 5 = 2.32$ 、 $\log_2 7 = 2.81$  を使うこと。

- (1) 開催／中止を伝えるメッセージの平均情報量  $info(S)$  を求めよ。  
 (A) 0.50 (B) 1.00  
 (C) 2.00 (D) 該当なし
- (2) テスト天気 ( $X$ =晴れ、または  $X$ =雨) の結果により  $S$  を分割した後、開催／中止を告げるのに必要な情報量の平均  $info_{\text{天気}}(S)$  を求めよ。(誤差範囲 0.01)  
 (A) 1.000 (B) 0.753  
 (C) 0.875 (D) 該当なし
- (3) 利得基準  $gain(\text{天気})$  を求めよ。(誤差範囲 0.01)  
 (A) 0.125 (B) 0.247  
 (C) 0.500 (D) 該当なし

得点	
----	--

三、

問三

次の各問で、正しい記述には○、間違った記述には×を答えよ。

- (1) 構文規則が精密になればなるほど、構文的あいまい性は少なくなる。
- (2) タイプ3文法は、文脈依存文法と呼ばれる。
- (3) プッシュダウンオートマトンの状態遷移関数は、現在の状態、入力文字、pd記憶の先頭を与え、次の状態とpd記憶の書換を返す関数である。
- (4) タイプ3文法で定義された言語を受理する有限オートマトンが必ず存在する。
- (5) 形態素とは、文法上一つの機能を持つ最小の言語単位である。
- (6) LALR(1)パーサは、pdオートマトンと似た構成を持つ。
- (7) 格文法では、語と語の意味を動詞を中心に記述する。
- (8) 文章の結束構造とは、文章が表現する意味として概念や事象のつながりを表現している。
- (9) 再現率とは、正解をどれだけ正しく選び出しているか、適合率とは、選び出されたものがどの位正しいかを示している。
- (10) 実例型機械翻訳方式では、「ある文と似た文は、元の文と同じように訳されることが多い」ということを原理としている。
- (11) 重回帰分析では、訓練事例と最も近い事例が属するクラスを採用する。
- (12) 学習方程式は、結合の強さと時間、入力の関係を示している。
- (13) 対話の解析では、根拠のある事実や真実や真であると思うことを告げると言うことを前提としている。
- (14) 文章生成では、何を言うか(what to say)と、どう言うか(how to say)の2段階がある。
- (15) 横浜と大連の時差は、2時間である。

