Institute of Systems Science
National University of Singapore

MASTER OF TECHNOLOGY IN ENTERPRISE BUSINESS ANALYTICS / INTELLIGENT SYSTEMS

Graduate Certificate Online Examination

Subject: *Practical Language Processing*

Sample Examination Questions

SECTION A

Question 1   *(Total: 23 Marks)*

You are a technical consultant to a healthcare group in Singapore, which would like to develop a customer service chatbot to be used by their clinics to automate some of the common administrative processes interacting with patients.

Patients usually call a clinic to make enquiries e.g. opening hours, describe their symptoms, find a doctor, schedule medical appointments, change/cancel existing appointments, etc.

For example, when a patient calls to make an appointment, he needs to provide his name, home address and cellphone number for the purpose of personal identification. The nurse checks if the patient is visiting the clinic for the first time. If yes, the records for the patient will be retrieved for reference during the conversation. Then the nurse asks the patient to briefly describe the symptoms, which are noted down as supporting information. Afterwards, based on the available time slots of doctors, the patient selects a preferred slot and the appointment is made.

Here is an example conversation between a patient and the nurse.

Good Morning! This is Great Health Clinic.

I'd like to make an appointment to see the doctor.

Please give me your name and telephone number.

This is Marcus Lee. My number is 91234567.

What time would you like to see the doctor?

Would tomorrow afternoon at 2 be okay?

Wait a moment, I'll check.
I'm sorry, there are no appointments available before 3pm.
We could make an appointment for 4 o'clock.

That's fine. Thank you.

Ok, 4 pm tomorrow, with Dr. Teo. You'll receive an SMS notification afterwards.
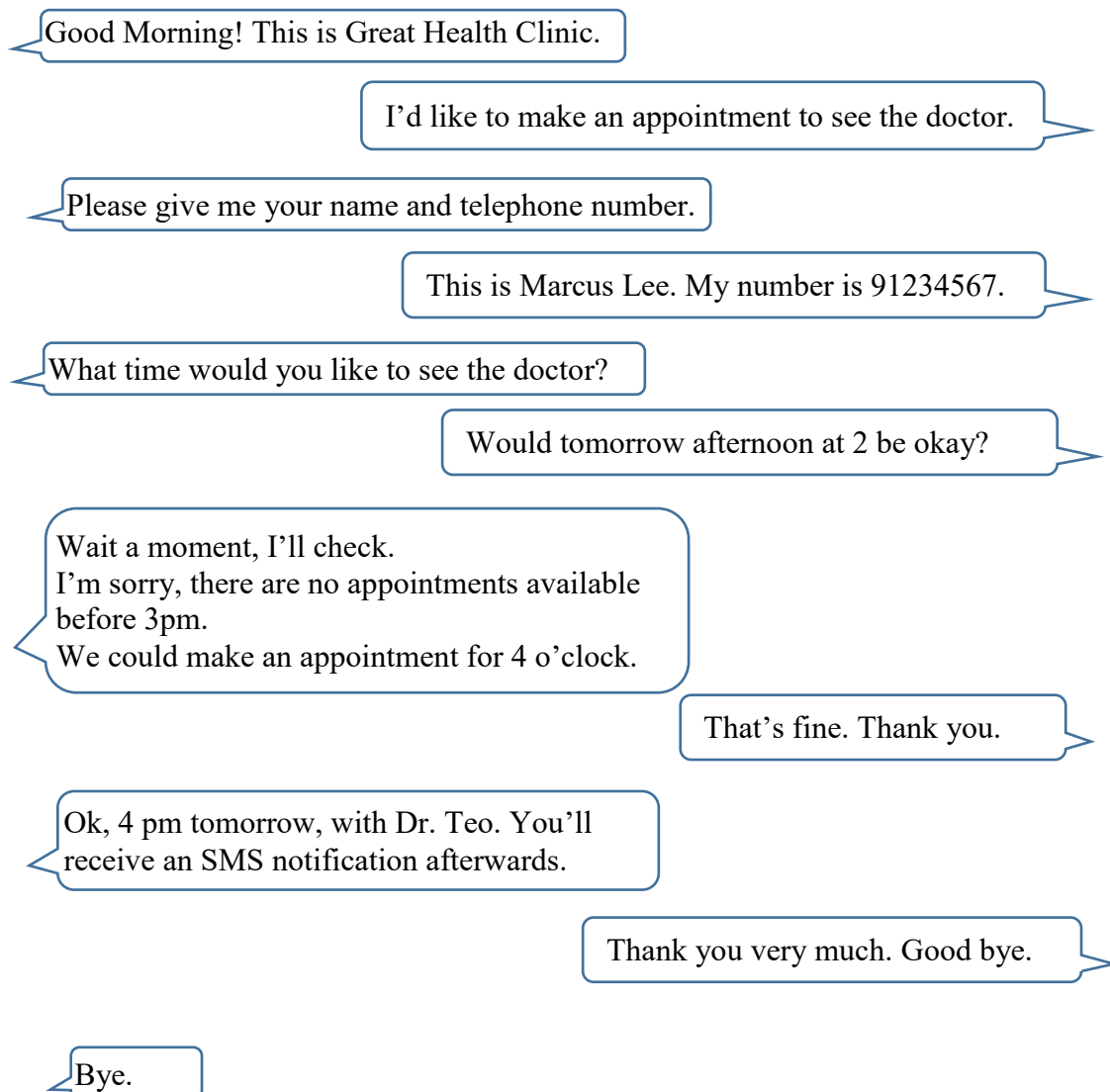
Thank you very much. Good bye.

Bye.

Figure 1. Sample conversation between caller and nurse

The flow chart shown below illustrates the major steps involved in a typical phone call for appointment scheduling. In the process of *Make an appointment,* if the patient is not a first timer, the doctor previously in charge will be prioritized for the upcoming appointment. Once the timeslot is confirmed, an SMS notification will be sent out.
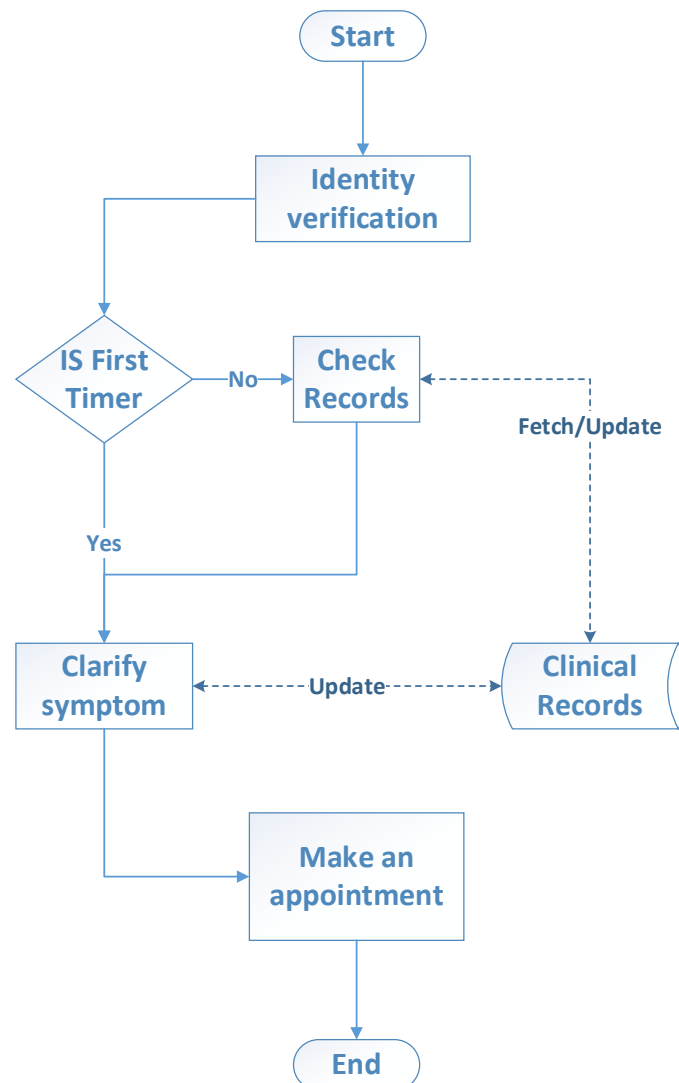

Figure 2. Process for Making Appointment

The healthcare group wants to have the chatbot voice-enabled so that the users can interact with it using speech over phones as well as using text in an app. However, they do have a concern that a caller may claim to be someone else and cause undesired change or cancellation of schedules.

a.   Based on the **context** and **flow chart** above, your teammate starts to design the chatbot by defining its intents and slots.

     1)   He defines the list of intents as the table below. Do you agree with his definitions? Justify your answer.

Intent	description
greet	Start conversation using greetings
self_identify	Identify patients by Name, Address, cellphone-number
describe_symptoms	Describe patients symptoms
Check_FristTimer	Validate patients' historical visiting record
Check_Records	Retrieve patients' historical visiting record
make_appointment	Make appointment with doctors
goodbye	End conversation using goodbyes

(2 Marks)

     2)   Suppose the user's utterance is "*My name is Marcus Lee. My address is 6 Keppel Bay Avenue, Singapore.*" What are the slots to be extracted and how to perform the extraction?

(2 Marks)

b.   Based on the **context** and **flowchart**, you first want to define the dialogue policy for the chatbot using handcrafted approach. **Draw** the *dialogue states diagram* and **index** the *states* to fully address the process starting from *Identity Verification* to *Make an Appointment*.

(4 Marks)

c.   For each of the *dialogue states* indexed in b, **list down** its associated *system actions*, and **define** the *actions*.

(3 Marks)

d.   For better scalability, instead of defining the system's *actions* for each step using handcrafted approach, you now want to redesign your chatbot such that the system's next *action* is predicted through machine learning based strategy. By indexing the Systems' actions into a list, for each turn of the dialog, it can be converted into a classification problem to pick up the most likely action from the possible action candidates list.

Your teammate proposes to use **Markov** model to predict the next action A_{t+1} based on the current user's utterance U_t, given the Markov assumption that future states depend only on the current state, not the events that occurred before it. Thus the *X_train* can be features extracted from U_t and *Y_train* will be the index of A_{t+1}.

Do you agree with this method? How would you improve this method?

(4 Marks)

e.   High quality speech recognizer is required for the voice enabled chatbot. In order to evaluate the accuracy of the different speech recognizers, you found a medicine speech dataset available, collected by a doctor in US. It contains short conversation clips recorded between doctors and patients during consultation using small hidden audio recorders, collected at several hospitals in New York. The speech data is in 16K 16bit PCM format.

Is this dataset suitable for evaluating speech recognizer in this application? Justify your answer based on the context given in the problem.

(4 Marks)

f.   To handle the concern that a caller may claim to be someone else and cause undesired change or cancellation of schedules. the healthcare group would like to add a feature to the chatbot so that it can recognize the speaker by voice.

1)   Considering that the user group of the system (patients of many clinics) is a very large group of people, what speaker recognition technology should be used in this case, *speaker identification* or *speaker verification*? Justify your choice and explain the main steps to use it.

(2 Marks)

2)   For your selected speaker recognition technology, there are options for text-dependent and text-independent solutions. If we expect the whole speaker recognition process does not take much time, which solution will you choose? Please explain the reason.

(2 Marks)

Question 2   *(Total: 27 Marks)*

Recently there has been great concerns about the COVID-19 virus. The management of the national healthcare group is especially concerned about the fake news being circulated on the pandemic on WhatsApp and other social media. Fake news tend to be sensational and have the potential to go viral and raise unnecessary fear level. You are now tasked to identify the fake news by applying NLP using the machine learning (ML) techniques that you have just learnt.

An important step in ML is feature engineering. In order to build a ML model to detect fake news, you need to understand what useful features are commonly used to build such models. You did some background reading and find the seminal paper by Liu Bing on detecting fake reviews on products. Below are summaries of the key feature sets that the authors have used to detect fake reviews:

1. Features about review content:

   - Lexical features such as word n-grams, part-of-speech n-grams, and other lexical attributes.
   - Content and style similarity of reviews from different reviewers.
   - Semantic inconsistency. For example, a reviewer wrote "My wife and I bought this car ..." in one review and then in another review he/she wrote "My husband really love ..."

2. Features about reviewer abnormal behaviors:

   - Public data available from web sites, e.g., reviewer ID, time of posting, frequency of posting, first reviewers of products, and many more.
   - Web site private/internal data, e.g., IP and MAC addresses, time taking to post a review, physical location of the reviewer, etc (a lot of them).

3. Product related features: E.g., product description, sales volume, and sales rank

4. Relationships: Complex relationships among reviewers, reviews, and entities (e.g., products and stores).

You find the information from this paper helpful. However, there is still an issue: the task in this paper is to detect fake product reviews. What works for this task may not be 'cross transferred' into detecting fake new. There needs to be similar grounding to apply fake reviews to fake news.

Table 1 below shows some examples of *actual fake news* that have circulated around during the COVID pandemic.

No	Content
1	why has mainland China significantly reduced the number of people infected in the past few days? In addition to washing their hands frequently with masks, they simply rinse their mouth with saline 3 times a day. When done, drink water for 5 minutes? This are the only way to prevent the spread of the coronavirus, which is not available on the market. Medicine, so don't buy it. Tong General Hospital, Dr. Wang Shuang:

	Neocoronary pneumonia will live in the throat for 4 days before it reaches the lungs. At this time, people will begin to cough and sore throat. Pass on this message as soon as you can save the lives of others!
2	"Yup, I was at Ya Kun Compass1 on Wed, 1 of their customers was fined $300 for sitting in the ✗space." They are very serious about it. Good to help out one another. Please let your SG friends know when u eat out w them, they don't give a chance anymore. $300 (RM900) is a lot of money as a fine. So don't sit on seats taped (/) or (X). Same as queue up.
3	This is to inform us all that the pH for corona virus varies from 5.5 to 8.5. All we need to do, to beat corona virus, we need to take more of an alkaline foods that are above the above pH level of the Virus. Some of which are: *Lemon - 9.9pH* *Lime - 8.2pH*

Table 1. Examples of fake news

Some examples of real news are also given below in Table 2. The fake and real news are not separated in the dataset collected.

No	Content
4	NEW YORK: A six-week-old infant has died of complications relating to COVID-19, the governor of the US state of Connecticut said on Wednesday (Apr 1), in one of the youngest recorded deaths from the virus. Governor Ned Lamont tweeted that the newborn was "brought unresponsive to a hospital late last week and could not be revived".
5	HONG KONG (REUTERS) - A pet cat has tested positive for the coronavirus in Hong Kong after its owner was confirmed with having the virus, the city's Agricultural and Fisheries department said, cautioning that the animal has not shown any signs of the disease. In a notice late on Tuesday (March 31), the department said there is currently no evidence that pet animals can be a source of Covid-19, the disease caused by the coronavirus, and owners should not abandon their pets.
6	NYT (CNN) Donald Trump has called himself a "wartime president" and said coronavirus is an "invisible enemy." The former CDC director has warned of a "long war ahead." New York Gov. Andrew Cuomo has called on the public to support our "troops," otherwise known as health care workers. The now-common metaphors comparing the efforts to stop coronavirus to a military war fit smoothly in many ways. Like in war, a pandemic has life-and-death decisions, an "enemy" who can strike at any time, "battles" on the "front lines" and calls for the "home front" to support the effort.

Table 2. Examples of real news

a. For each of the key feature sets summarized by Liu Bing's paper above, determine if they can be cross-transferred from detecting fake reviews to detecting fake news. Justify your answers.

(4 Marks)

b. For each of the fake news in Table 1 above, which of the key feature sets raised by Liu Bing will be useful in detecting such fake news? Explain your choices.

(3 Marks)

c. After you have studied the useful features that can be used to identify the fake news, discuss and explain what type of features engineering techniques will be useful. Limit your discussion to the 2 feature types - term-document matrix and graphical triples. Your discussion must be related to the example fake news in the above section.

(3 Marks)

d. Your management has pointed out that they are more concerned about false negatives than false positives in this case. That is, they are more concerned about *saying a fake (positive) news is genuine (negative) when it is actually fake (positive)*. This is because they are concerned about the social problems that fake news can cause if they are left to spread. As such they have instructed you to consider this when building your model. Without actually going into the laborious training details, discuss how you would try to mitigate this problem.

(3 Marks)

e. The VP of engineering looked thoroughly at your project and realized that the classification task for a fake news detection cannot be as simplistic as a binary classification like detecting fake e-commerce product reviews.

    She suggested to increase the granularity of the labels where each news article is labeled as either (i) **False** (i.e. completely false), (ii) **Mostly False** (with some truth), (iii) **Half Truth** (i.e. the same amount of truth and false statements in the article), (iv) **Mostly True** (with some falsity) and (v) **Truth** (i.e. completely true)

    Then she passed you a publication, which showed that BERT algorithms have shown to outperform other algorithms on various classification tasks. She suggests that you apply state-of-the-art (SOTA) transfer learning to build a fake news classifier.

    Specify what kind of labelled data is necessary to create the SOTA transfer learning fake news classifier. Use the given example news to illustrate the input data and labels to be used to train a BERT model.

(2 Marks)

f. You notice that the news data the healthcare group has, as shown in Table 1 & 2 on page 15, cannot be directly used to train the BERT based classifier in e). Identify the problem of the data, and propose <u>one method</u> to solve this problem.

(3 Marks)

g.   One week after trying to fine-tune a GPT-2 or BERT model, test results show that the model you've trained have 20% accuracy on the single-label multi-class classification task with 5 classes. The same results of exactly 20% accuracy is achieve after you changed several sets of hyperparameters.

As a first attempt to debug, you've checked the data and confirmed that the training, validation and test data you used were all uniformly distributed, i.e. containing the same number of *fake* and *real* data points.

There are discussion to shut down your project if your model continues to perform exactly at 20% so it is important for you to figure out what is wrong with the model, by forming a hypothesis of possible situations and debug the model.

The following questions focus on the debugging of outputs and models so that you have a reason to exist in the company.

1) You found that the last layer of the model in g) has no *activation* function and the loss function is *mean square error* (L2 loss). This isn't the typical setup for a multi-class single-label classification.

What combination of the final layer activation and the corresponding loss function would you otherwise suggest? Explain your choices.

(3 Marks)

2) Besides changing the combination of activation and corresponding loss function, what other regularization measures can be taken to train a better model? How do you order the regularization methods to be experimented? Explain the order.

(3 Marks)

3) After a change of last activation and objective function and a series of regularization measures, the final model is finally meeting the accuracy required. However, you realize that your solution, though accurate, is expensive and slow because of the usage of GPUs and the large number of layers in the default GPT-2 and BERT models.

Recalling the generalized language model where the DistilBERT and TinyBERT models were discussed, you remembered there are several very recent advancement to speed up state-of-the-art inference of models (i.e. model serving/usage). You want to try your very best to convince the management to give you more time before model deployment to resolve the speed and cost issues at inference.

Propose to the management what you will adopt to resolve the speed and cost issues at model inference. Justify your suggestion.

(3 Marks)