

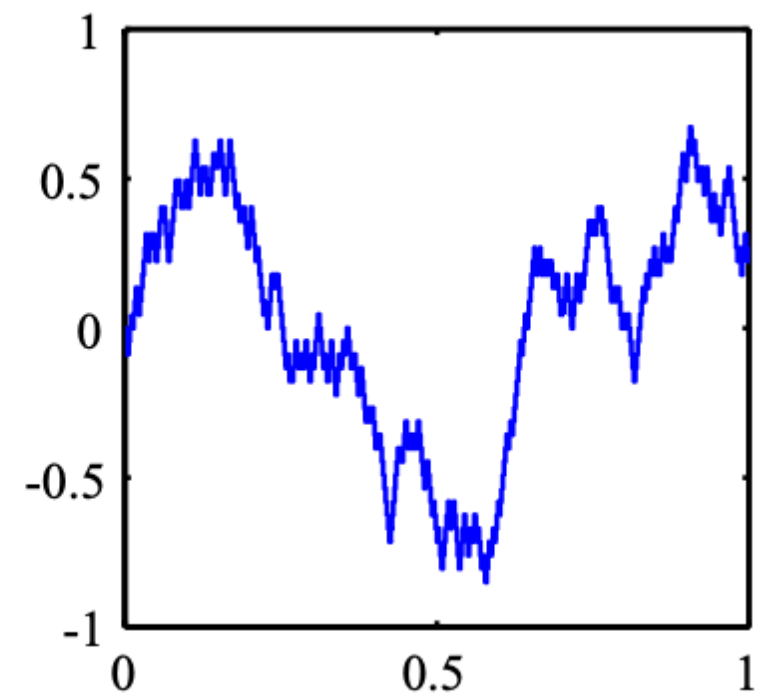
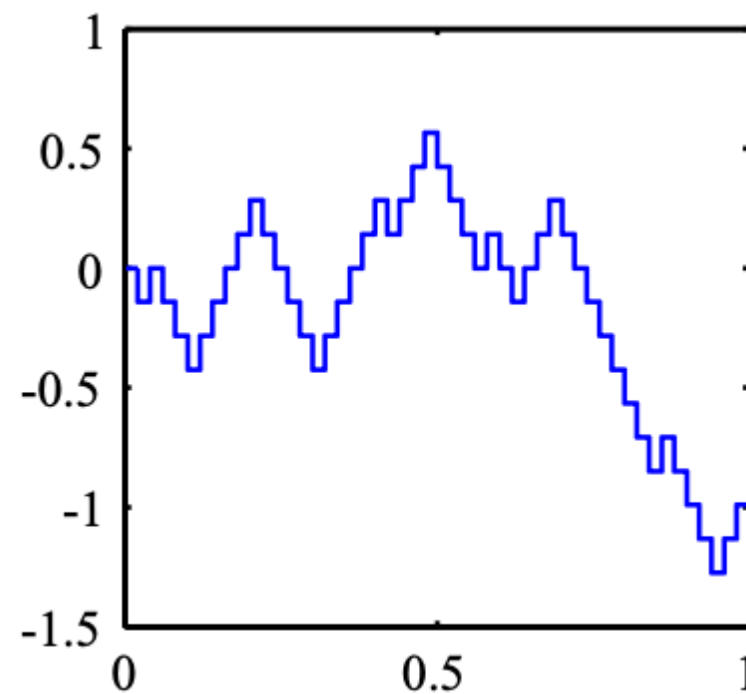
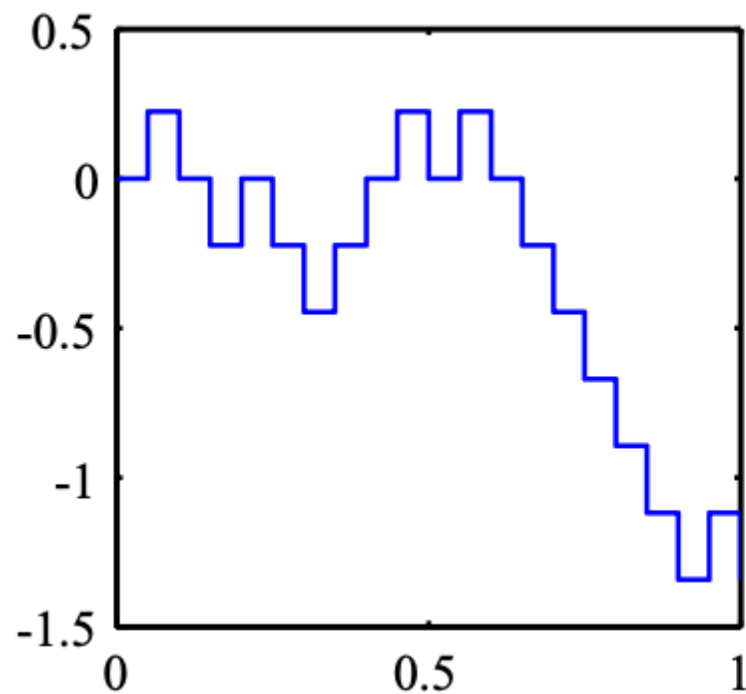
Стохастический анализ и приложения

Филиал МГУ им. М.В. Ломоносова в г. Ташкенте

D.Sc. Болонкин Максим Викторович

Про что курс?

- Курс строится на основе лекций Ramon van Handel (ссылка на конспект лекций на сайте)
- Стохастические процессы



Про что курс?

- Исчисление Ито

$$X_t = X_0 + \int_0^t \sigma_s dB_s + \int_0^t \mu_s ds$$

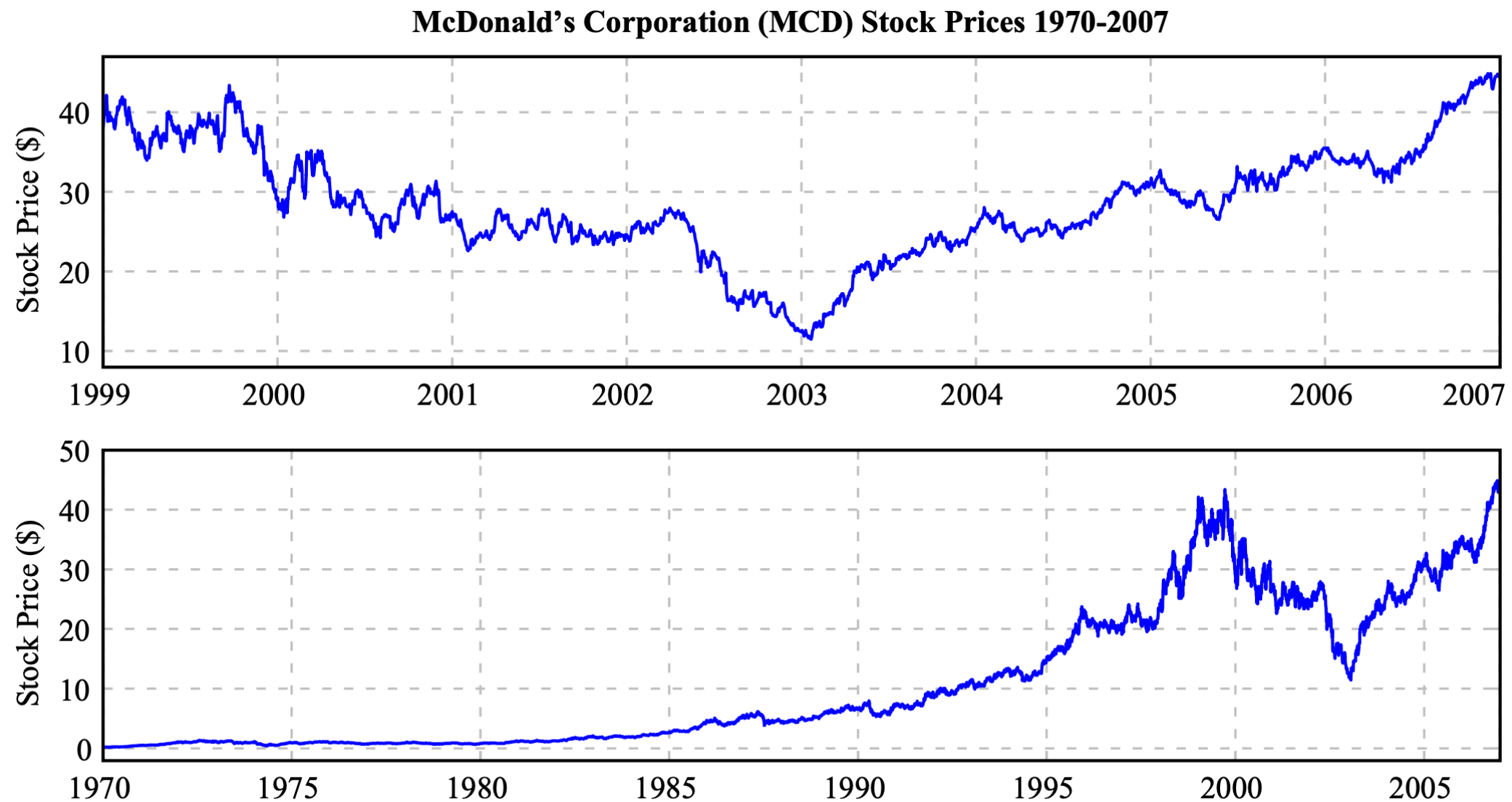
Про что курс?

- Стохастические дифференциальные уравнения

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t, \quad X_0 = x.$$

Про что курс?

- Применения в финансовой математике



- Как прогнозировать поведение ресурсов?
- Как оценивать стоимость финансовых инструментов?
- Как оптимально формировать портфолио финансовых инструментов?

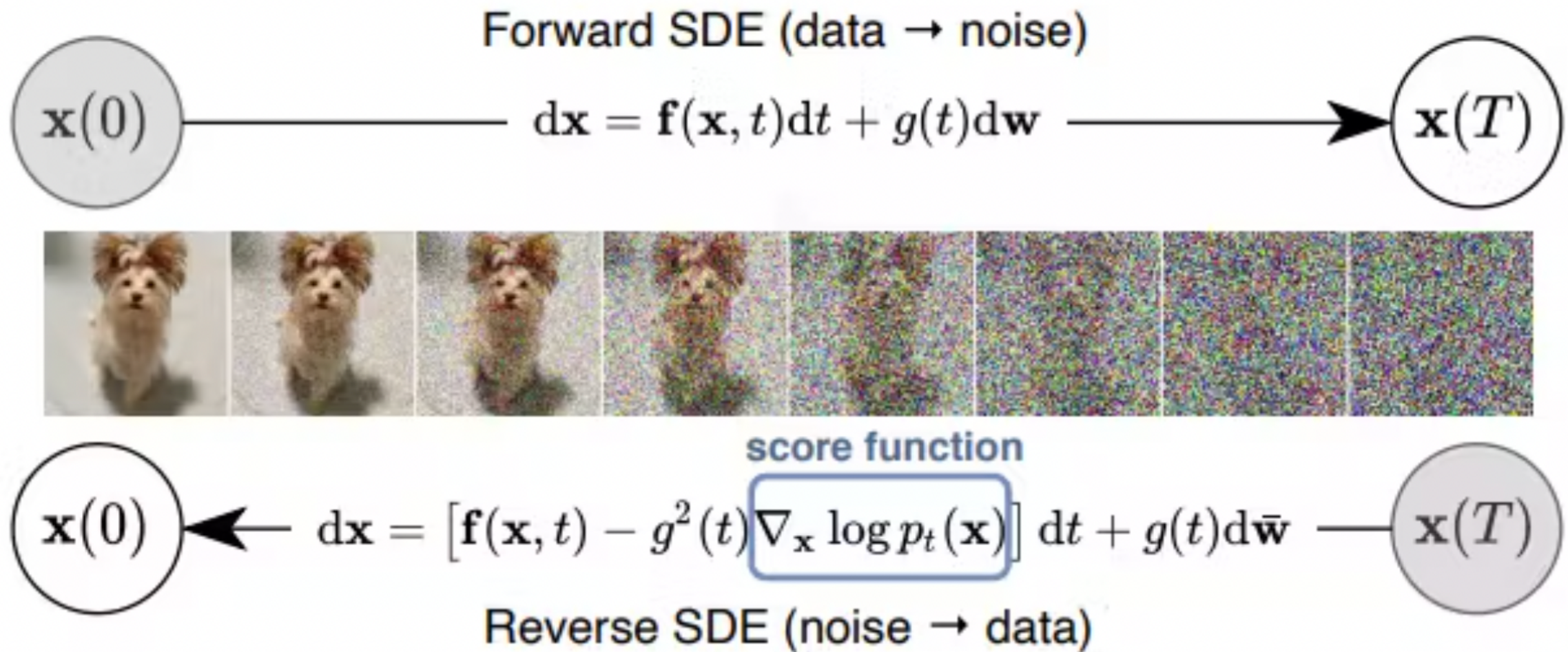
Про что курс?

- Стохастический контроль, оптимальная остановка



Про что курс?

- Генеративные модели на основе стохастических процессов



- Как генерировать данные из произвольных распределений?
- Как работают диффузионные модели?

Логистика курса

- Курс собран с нуля, поэтому, возможно, будет содержать погрешности
- Цель курса: рассказать вам что-то полезное и изучить что-то интересное самому
- Курс будет включать в себя лекции и практические занятия (формат TBD)
- Значительный объем математики: выводы и доказательства
- Контроль:
 - *ВОЗМОЖНЫЕ* промежуточные контрольные
 - *ВОЗМОЖНЫЕ* домашние задания
 - устный экзамен
- Все материалы курса будут выложены на сайте
 - <https://www.mbolonkin.info/teaching/stoch-2025>
- Вся коммуникация будет в телеграм-чате

**Как правильно выполнять
научную работу и писать
магистерскую диссертацию**

Как выбрать задачу?

- Предлагает научный руководитель:
 - Научные интересы
 - Гранты
- Предлагает сам студент:
 - Известная задача —> улучшенное решение
 - Новая задача или новый взгляд на известную задачу
 - Нужно следить за литературой!

Типы научных задач

Конкретная прикладная/теоретическая задача	Общий метод решения классов задач
• Можно ли с помощью определенного класса автоматов выполнить определенную задачу? • По определенному набору клинических признаков предсказать клинически-важный параметр (например, диагноз) • Как натренировать нейросеть для генерации узбекской музыки?	• Каковы общие свойства определенных классов автоматов? • Как улучшить предсказательные свойства моделей для случаев недостаточных данных? • Как эффективно обучать нейросети для произвольных жанров музыки?

Магистерская диссертация vs. курсовой проект

- Курсовой проект:
 - Известная задача
 - Известные решения
 - Цель: показать, что студент умеет применять изученные методы к конкретным задачам
- Магистерская диссертация:
 - Существенно новая задача и/или
 - Существенно новое решение
 - Цель: показать, что студент способен придумывать новые решения, показывать их корректность, проводить необходимые научные изыскания

Как решать задачу?

- Если задача не новая, посмотреть как ее решают другие
- Выяснить ограничения чужих решений
 - Слабый результат, есть куда улучшать
 - Не учитывает какую-то важную информацию
 - Слишком сложно, слишком долго, много ресурсов
 - Подходит только для частного случая
- Если задача новая, то посмотреть как решают аналогичные или близкие задачи
- Выяснить почему аналогичные решения не подойдут к вашей задаче

Пример “НОВИЗНЫ”

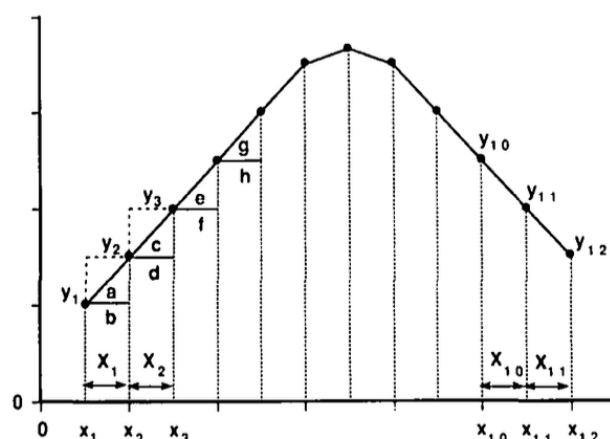


Figure 1—Total area under the curve is the sum of individual areas of triangles a, c, e, and g and rectangles b, d, f, and h.

$$\text{Area} = \frac{1}{2} \sum_{i=1}^n x_{i-1} (y_{i-1} + y_i)$$

(Tai's formula)

When the curve passes the origin: $x_0 = y_0 = 0$, $X_0 = x_1 - 0$;

When the curve intercepts Y-axis at y_0 : $X_0 = x_1 - 0$

When the curve neither passes the origin nor intercepts at y-axis: $X_0 = y_0 = 0$

Example using Tai's model:

Blood glucose determined at six time periods: (6)

time (min) 0 30 60 90 120

Glucose (mg/dl) 95 147 124 111 101

$X_0 = x_1 - x_0 = 30 - 0 = 30$;

$X_1 = 60 - 30 = 30$; $X_2 = X_3 = 30$

formulas to a standard (true value), which is obtained by plotting the curve on graph paper and counting the number of small units under the curve. The sum of these units represents the actual total area under the curve. Results are presented in Table 1. From Table 1, it is evident that total area I can not be obtained from Alder's formula. Total area II has underestimated the total area under a metabolic curve by a large margin. Total area III corresponds well (-6.1%) with the actual area estimated from the plot (total area V). However, this formula only permits a single t value, which means the time interval has to be the same.

CONCLUSIONS

Verification of Tai's mathematical model

From Table 1, it is clear that Tai's formula (total area IV) has the most accu-

rate estimation of the total area under a curve. Total area IV agrees extremely well with actual total area obtained from the graph ($+0.1\%$). Because no statistically significant differences were found between areas from these two methods, the validity of Tai's model can thus be established.

This formula also permits accurate determination of total area under the curve when the curve intercepts with Y-axis, as well as when the curve passes the origin. Furthermore, in this formula, values on X-axis do not have to be the same as the t in Wolever et al.'s formula. It allows flexibility in experimental conditions, which means, in the case of glucose-response curve, samples can be taken with differing time intervals and the total area under the curve can still be determined with precision. Thus, if different authors estimate the total area under a curve from

Table 1—Summary of results: (% area: % of total area V)

Total area	I	II	III	IV	V
Test					
Glucose	N.A.*	480 (3.3%)	13517 (94.3%)	14400 (100.4%)	14337
TEF (SM)	N.A.*	336 (3.2%)	9588 (92.6%)	10326 (99.8%)	10349
TEF (LM)	N.A.*	452 (3.2%)	13367 (94.7%)	14163 (100.3)	14115
RMR (L)	N.A.*	1157 (3.9%)	N.A.†	30040 (100.0%)	30047
RMR (O)	N.A.*	1636 (4.6%)	N.A.†	35733 (100.0%)	35725
Ave		(3.6%)	(93.9%)	(100.1%)	

Теоретическая работа

- Обычно само-очевидная задача (очевидна новизна, практическая и теоретическая значимость)
- Результаты: доказанные теоремы, леммы, утверждения
- Также могут быть полезны эмпирические подтверждения: симуляции с последующим анализом

Прикладная работа

- Новизна чаще всего не вытекает из задачи (за редкими исключениями)
- Результаты: предложенные методы решения задачи
- Необходимо убедительно показать, что
 - Предложенные методы работают правильно
 - Предложенные методы работают лучше, чем другие

Вы должны показать, зачем вы делали вашу работу

Прикладная работа

Важно показать, что ваше решение в самом деле лучше чего-то другого (в идеале — самых лучших чужих методов, но на худой конец можно сравнить с более простыми стандартными методами)

- Та же задача, чужие методы, тот же датасет
- Похожая задача, адаптированные чужие методы, тот же датасет
- Свои методы, другие датасеты (желательно с имеющимися бейслайнами)
- Свои бейслайны, тот же датасет

Прикладная работа

Пример: задача детекции лжи в сообщениях Twitter

Датасет: из статьи “Deception detection in Twitter”

Та же задача, тот же датасет, чужие методы	
Похожая задача, похожие методы, тот же датасет	
Свои методы, другие датасеты	
Тот же датасет, свои бейслайны	

Как тестировать?

- Готовые датасеты/бенчмарки
 - Решается та же задача
 - Решается другая задача, но можно адаптировать
- Собирать датасет самим
 - До 2020х позитивно влияло на шансы успешно пройти рецензию
 - Может быть дорого и долго

Инструменты

- Google Scholar
- SciHub

Google Scholar



☒ Articles ☐ Case law



sci-hub

enter your reference



open

Как писать работу?

- Заключение и введение часто пишут последними, аннотацию вообще в самом конце
- Обзор литературы должен ясно показать кто и как до вас решал эту или похожие задачи и какие недостатки этих методов побудили вас решать эту задачу
- Методы должны быть понятно описаны с необходимыми деталями, не нужно описывать общеизвестные вещи (в курсовых и дипломных можно включить какие-то описания в отдельный раздел во введении, но объем такого материала не должен быть больше или равен объему нового материала)

Как писать работу?

- Графики и таблицы должны иметь самостоятельные заголовки и обязательно обсуждаться в основном тексте работы
- Ваш код никому не интересен, оставьте его для приложения (лучше на носителе)
- Нужно показывать как хорошие результаты, так и плохие (с объяснениями)
- В списке литературы должны быть только те источники, которые вы процитировали в основном тексте

Научитесь пользоваться LaTeX

- **Overleaf** позволяет работать онлайн без установки LaTeX на свою машину
- Шрифты, математическая нотация
- Верстка похожа на программирование
- Общепринятый формат верстки научных работ

Научитесь пользоваться LaTeX

- Tikz позволяет делать рисунки

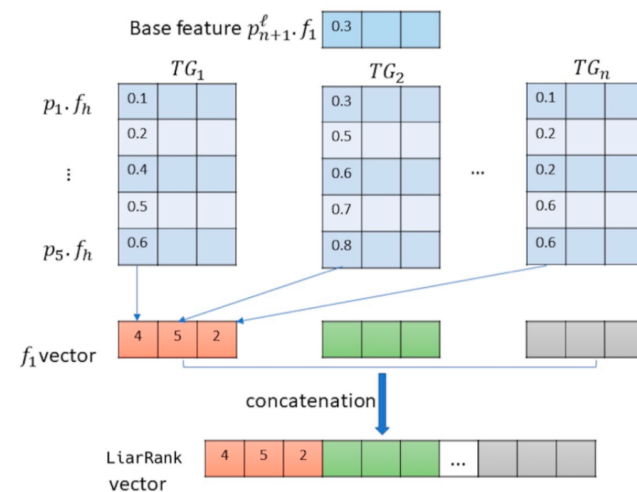


Figure 3.2: LiarRank meta-feature calculation process.

TG_3 and that value of this feature for the player p_{n+1}^l is 0.3. As there are three base features in this highly simplified example for each player, the LiarRank $sr(p_{n+1}^l, f_1)$

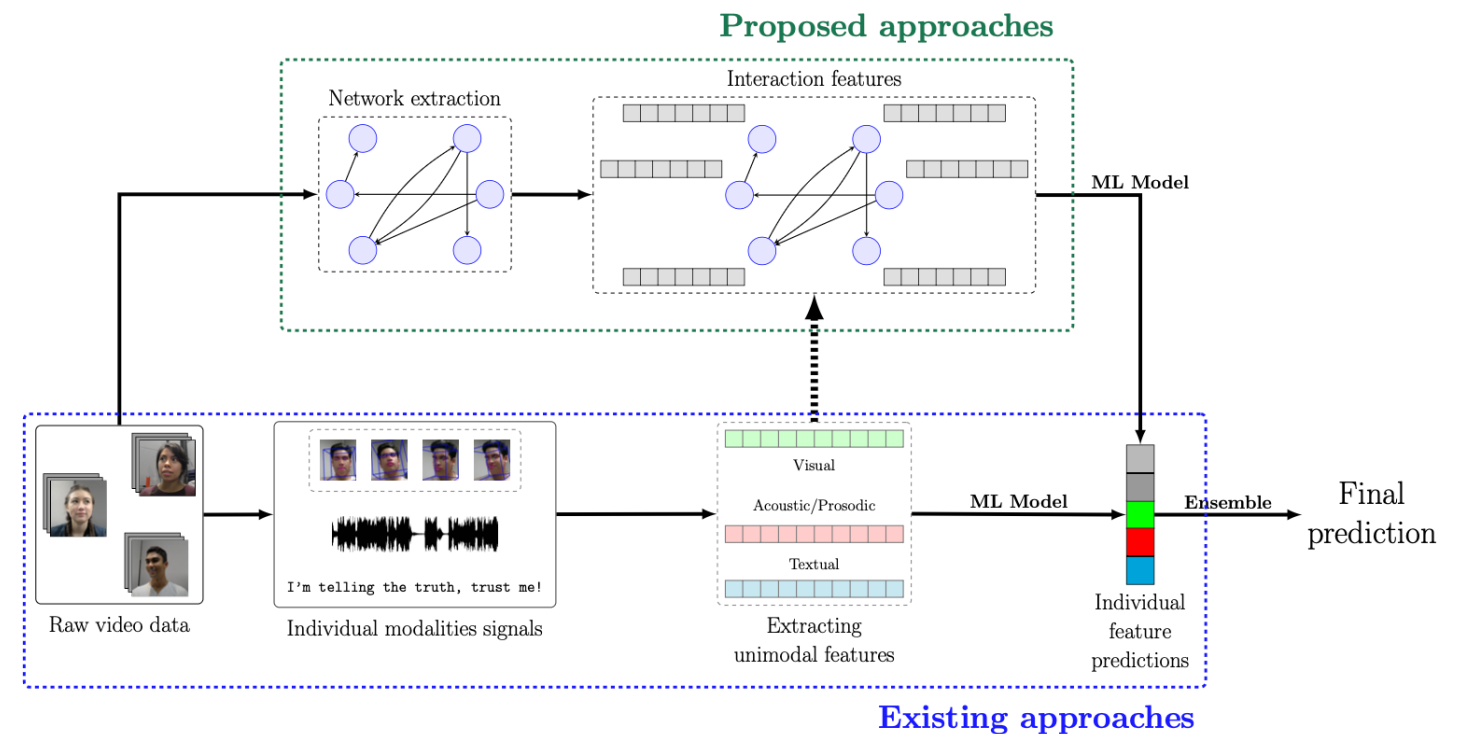


Figure 1.1: Scope of the dissertation. The majority of automated social behavior analysis works are considering each individual in the group, extracting some relevant features and making predictions about traits or actions in question. We propose to exploit the social nature of the interactions and produce features and models that

Как защищать работу?

- Понимайте свою аудиторию
- Понимайте ограничения по времени
- Нужно рассказать о чем работа и убедить аудиторию, что вы не зря делали эту работу
- Не нужно тратить время на общеизвестную теорию
 - Исключение: ваша работа непосредственно об этой теории