**Hotel recommendation system with Vietnamese reviews based on deep learning**

**Đóng góp của đồ án:**

1. **Bộ dữ liệu tiếng Việt**
2. **Phương pháp xây dựng mô hình recommend systems từ lịch sử cmt của người dùng trên các trang mạng du lịch điện thử**

**CODE**

* Topic (4 topic liên quan tới khách sạn (multi-labels), ex: cơ sở vật chất, nhân viên, vệ sinh, vị trí + sentiment)

**Pipeline:**

* Source data crawl (Hotel): traveloka, ivivu
  + Các dữ liệu cần lấy: tên user, tên khách sạn, địa điểm (Đà Nẵng, TPHCM, HN,…), rating, comment
* Dataset:
  + Crawling data
  + Tiền xử lý
  + Gán nhãn (guideline, tính độ đồng thuận)
  + Trực quan bộ dữ liệu hoàn chỉnh (Trình bày để giới thiệu bộ dữ liệu)
* Model Classification:
  + Model: PhoBERT, **Glove, ELMo**, Fastext, W2V + LSTM + DNN
  + Input: tất cả comment
  + Output: Dự đoán **3** nhãn **tích cực (tiêu cực)** và **topic**
* Model Recommend system:
  + Colaborative filtering with DL
    - User\_Id: embedding Giúp các user có sở thích tương tự lại gần với nhau hơn
    - Hotel\_Id: embedding giúp các Hotel có các tính chất giống nhau sẽ lại gần nhau hơn
    - topic: embedding giúp các topic liên quan sẽ gần nhau
    - cmt, sentiment: giúp máy hiểu hơn về ngữ cảnh + cũng như mức độ hài lòng hoặc cảm xúc của khách hàng
  + Lấy cmt làm context :
    - Context :
      * topic: (None,5) =(embedding)> (None,5,10) =(reshape)> (None,50)
      * sentiment: (None,3)
    - Concatenate: topic, sentiment
    - Dense: giảm chiều context
    - Concatenate: Context, User\_Id, Hotel\_Id
  + Output: rating và nomalize hóa

Dự đoán:

Input : profile lịch sử người dùng

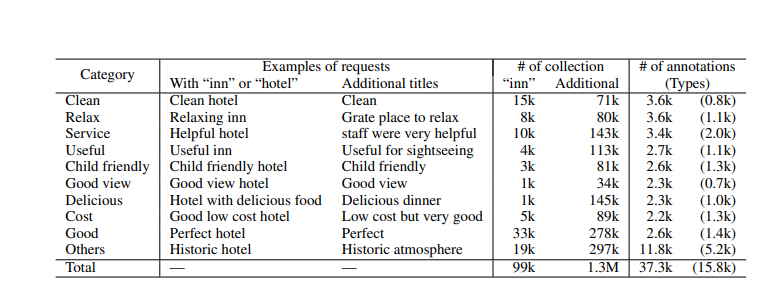
Input dự đoán: tên người dùng, tên hotel người dùng chưa đi, topic + sentiment (được dự đoán từ cmt của 1 record gần nhất)

Output dự đoán: các giá trị rating star

**BÁO CÁO (10-12 trang)**

Abstract – **0.5 trang**

1. Introduction (lấy báo cáo comprehensive qua, viết cái tổng quát về RS ở travel) – **0.75 trang**
2. Related work (3 bài báo: comprehensive lấy RS theo colaborative + 2 bài báo) – **0.75 trang**
3. Proposed aproach (vẽ 2 quy trình tiếp cận của model + giải thích) – **1 trang**
4. Dataset creation
   1. Collecting & preprocessing (có process graph) – **1 trang**
   2. Labeling (bảng topic, công thức + tính độ đồng thuận) – **1,5 trang**
   3. Dataset result (trình bày để giới thiệu bộ dữ liệu) – **0,5 trang**



*(Examples)*

1. Model neural network design
   1. Backbone (3 model + cell LSTM) – **1 trang**
   2. Topic & sentiment classification task – **1 trang**
   3. Recommendation task – **1 trang**
2. Experiments & disscussion (có phân tích lỗi) – **1.5 trang**
3. Conclusion – **0.5 trang**

**1 trang**

1. CRediT authorship contribution statement

* *(Elham Asani: Conceptualization, Methodology, Software, Writing – original draft.*
* *Hamed Vahdat-Nejad: Supervision, Conceptualization, Methodology, Writing – review & editing.*
* *Javad Sadri: Supervision, Methodology.)*

1. Declaration of competing interest

* *The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.*

1. Acknowledgments

* *We would like to thank the anonymous reviewers for their valuable feedback.*
* *Cảm ơn thầy*

1. References

**SLIDE (15 trang)**

Slide 1 : Tên đề tài + tên thành viên nhóm

Slide 2 : Introduction (show kết quả recommend từ 1 user)

Slide 3 : Proposed aproach

**Dataset**

Slide 3 + 4 : Collecting & preprocessing

Slide 4 + 5: Labeling

Slide 6 : Dataset result

**Model & Result**

Slide 7 + 8 : Topic & sentiment classification task (nói input, output, kiến trúc model)

Slide 9 : Bảng kết quả đánh giá model

Slide 13 : Phân tích lỗi

Slide 10 + 11: Recommendation task (nói input, output, kiến trúc model)

Slide 12 : Bảng kết quả đánh giá model

Slide 13 : Phân tích lỗi

Slide 14 : Conclusion (kết luận so với kết quả related work, đóng góp, hạn chế, hướng phát triển)

Slide 15 : Cảm ơn