Báo cáo - Nhập môn xử lý ngôn ngữ tự nhiên

nhập môn thống kê (Đại học Tôn Đức Thắng)

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG
KHOA CÔNG NGHỆ THÔNG TIN

BÀI TẬP LỚN/ĐỒ ÁN CUỐI KÌ MÔN …….

…tên đề tài…

Người hướng dẫn: **TS NGUYỄN VĂN A**

Người thực hiện: **NGUYỄN THỊ B – MSSV**

TRẦN VĂN C – MSSV

Lớp : **10050301**

Khoá : **17**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2014

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG
KHOA CÔNG NGHỆ THÔNG TIN

BÀI TẬP LỚN/ĐỒ ÁN CUỐI KÌ MÔN …….

…tên đề tài…

Người hướng dẫn: **TS NGUYỄN VĂN A**
Người thực hiện: **NGUYỄN THỊ B**
TRẦN VĂN C
Lớp : **10050301**
Khoá : **16**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2014

LỜI CẢM ƠN

Đây là phần tác giả **tự viết** ngắn gọn, thể hiện sự biết ơn của mình đối với những người đã giúp mình hoàn thành Luận văn/Luận án. Tuyệt đối không sao chép theo mẫu những "lời cảm ơn" đã có.

ĐỒ ÁN ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG

Tôi xin cam đoan đây là sản phẩm đồ án của riêng tôi / chúng tôi và được sự hướng dẫn của TS Nguyễn Văn A;. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình. Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

TP. Hồ Chí Minh, ngày tháng năm

Tác giả

(ký tên và ghi rõ họ tên)

Nguyễn Văn B

Trần Văn C

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

Phần xác nhận của GV hướng dẫn

_____

_____

_____

_____

_____

_____

_____

Tp. Hồ Chí Minh, ngày tháng năm
(kí và ghi họ tên)

Phần đánh giá của GV chấm bài

_____

_____

_____

_____

_____

_____

_____

Tp. Hồ Chí Minh, ngày tháng năm
(kí và ghi họ tên)

TÓM TẮT

Trình bày tóm tắt vấn đề nghiên cứu, các hướng tiếp cận, cách giải quyết vấn đề và một số kết quả đạt được, những phát hiện cơ bản trong vòng 1 -2 trang.

MỤC LỤC

DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT

CÁC KÝ HIỆU

f *Tần số của dòng điện và điện áp (Hz)*

p *Mật độ điện tích khối (C/m3)*

CÁC CHỮ VIẾT TẮT

CSTD Công suất tác dụng

MF Máy phát điện

BER Tỷ lệ bít lỗi

DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ

DANH MỤC HÌNH

DANH MỤC BẢNG

CHƯƠNG 1 – MỞ ĐẦU

1.1 Tiểu mục cấp 1

Sử dụng **kiểu chữ Times New Roman (Unicode) cỡ 13** của hệ soạn thảo Winword; **mật độ chữ bình thường**; không được nén hoặc kéo dãn khoảng cách giữa các chữ; **dãn dòng đặt ở chế độ 1.5 lines**; **lề trên 3.5 cm**; **lề dưới 3cm**; **lề trái 3.5 cm**; **lề phải 2cm**. Số trang được đánh ở giữa, phía trên đầu mỗi trang giấy. Nếu có bảng, biểu, hình vẽ trình bày theo chiều ngang khổ giấy thì đầu bảng là lề trái của trang, nhưng nên hạn chế trình bày theo cách này.

Nội dung của tiểu mục cấp 1, một mục khi chia nhỏ thì tối thiểu là 02 mục con (tức là nếu có 1.1.1 thì phải có 1.1.2); tối đa không nên quá 05 mục con.

1.1.1 Tiểu mục cấp 2

Nội dung chi tiết của tiểu mục.

1.1.1.1 Tiểu mục cấp 3

- Đây là cấp tiểu mục nhỏ nhất, không thể tiếp tục phân chia.
- Các ý trong tiểu mục được trình bày gạch đầu dòng "-".
- Các ý nhỏ hơn sử dụng bullet như sau:
    - Ý nhỏ 1.
    - Ý nhỏ 2.
- Cần lưu ý rằng đây là cấp sâu nhất, không được phép chia thành 1.1.1.1.1 .

1.1.1.2 Tiểu mục cấp 3 tiếp theo.

Nội dung của tiểu mục thứ ba, khi soạn thảo hãy dùng Styles có sẵn, để khi tạo mục lục sẽ tự động và đồng nhất mỗi khi chúng ta thay đổi format.

1.1.2 Tiểu mục cấp 2 tiếp theo

Không phải lúc nào cũng chia thành tiểu mục cấp 3, nếu như ý trình bày được gói gọn.

1.2 Nội dung của chương này

Chương này trình bày lý do chọn đề tài, mục đích, đối tượng và phạm vi nghiên cứu, ý nghĩa khoa học và thực tiễn của đề tài; cơ sở khoa học của việc chọn đề tài...;

CHƯƠNG 2 – BÀI 2

2.1 Giới thiệu bài toán

Information Extraction (IE) trong Natural Language Processing (NLP) là quá trình trích xuất thông tin có cấu trúc từ văn bản không có cấu trúc, nhằm chuyển dữ liệu từ dạng tự do thành các dạng dễ xử lý như cơ sở dữ liệu hoặc bảng biểu. Bài toán IE bao gồm nhiều nhiệm vụ con, như Named Entity Recognition (NER) để nhận diện các thực thể có tên (ví dụ: tên người, tổ chức, địa điểm), Relation Extraction (RE) để xác định các mối quan hệ giữa các thực thể, và Event Extraction (EE) để trích xuất các sự kiện và thông tin liên quan đến thời gian, địa điểm, và đối tượng tham gia. Các phương pháp giải quyết IE có thể sử dụng quy tắc xác định (rule-based), học có giám sát (supervised learning), học không giám sát (unsupervised learning), hoặc học sâu (deep learning) với các mô hình như LSTM, BiLSTM, và Transformer. Các ứng dụng của IE rất rộng, từ việc tự động xây dựng cơ sở dữ liệu từ văn bản, phân tích dữ liệu văn bản để rút ra các xu hướng, đến hỗ trợ trợ lý ảo và chatbot trong việc trích xuất thông tin quan trọng từ cuộc hội thoại. IE cũng đóng vai trò quan trọng trong các lĩnh vực như an ninh mạng, khai thác thông tin từ dữ liệu lớn, và phân tích tài liệu, giúp tự động hóa nhiều công việc nghiên cứu và kinh doanh.

Bài toán trích xuất email và số điện thoại là một dạng bài toán trong Information Extraction (IE), đặc biệt là trong Named Entity Recognition (NER), nhằm nhận diện và trích xuất các thực thể có cấu trúc từ văn bản không có cấu trúc. Một trong những phương pháp phổ biến để giải quyết bài toán này là sử dụng biểu thức chính quy (regex), vì email và số điện thoại có những mẫu định dạng rõ ràng, dễ dàng nhận diện. Tuy nhiên, phương pháp này chỉ hiệu quả khi dữ liệu có cấu trúc chuẩn và không thể xử lý các trường hợp phức tạp hoặc không đồng nhất. Bên cạnh đó, các mô hình học máy như SVM, CRF, hoặc RNN có thể học từ dữ liệu huấn luyện có nhãn để nhận diện các mẫu phức tạp hơn, mặc dù phương pháp này cần nhiều tài nguyên và thời gian huấn luyện. Các mô hình đã huấn luyện sẵn như SpaCy hoặc BERT cũng có thể được

sử dụng để trích xuất email và số điện thoại từ văn bản mà không cần huấn luyện lại từ đầu, giúp tiết kiệm thời gian và tài nguyên. Bài toán này có ứng dụng rộng rãi trong việc thu thập dữ liệu, quản lý thông tin khách hàng, hệ thống chống spam, và phân tích thị trường, giúp tự động hóa việc trích xuất thông tin liên hệ quan trọng từ các nguồn văn bản khác nhau.

2.2 Lý thuyết áp dụng

Biểu thức Chính quy (Regular Expressions)

Trong đoạn mã, biểu thức chính quy (regex) được sử dụng để trích xuất email và số điện thoại từ câu văn. Các mẫu regex xác định định dạng của email và số điện thoại, ví dụ:

- Email: r'[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}'
- Số điện thoại: r'\+?\d{1,3}[-.\s]?\(?\d{1,4}\)?[-.\s]?\d{1,4}[-.\s]?\d{1,4}'

Các mẫu này tìm kiếm và trích xuất tất cả các địa chỉ email và số điện thoại trong câu, một phương pháp phổ biến trong việc xử lý văn bản có cấu trúc.

Tiền xử lý Dữ liệu và Gắn nhãn BIO

Tiền xử lý dữ liệu là một bước quan trọng trong NLP, nơi văn bản được chuẩn hóa và chia thành các phần tử nhỏ hơn (token) để dễ dàng xử lý. Trong đoạn mã, câu văn được chia thành các từ (tokens) và các nhãn BIO được gán cho các từ này.

- BIO là một hệ thống nhãn dùng trong NER:
- B (Begin) dùng để chỉ một thực thể bắt đầu (ví dụ: B-EMAIL, B-PHONE).
- I (Inside) dùng để chỉ các phần tiếp theo của thực thể đó (ở đây không có nhãn I, vì chỉ trích xuất các thực thể đơn lẻ).
- (Outside) dùng để chỉ các từ không thuộc thực thể nào.

Mỗi câu văn sẽ được phân tích và gắn nhãn cho các từ, giúp mô hình học cách nhận diện các thực thể có tên trong văn bản.

Tokenization và Align Labels (Đồng bộ Nhãn với Token)

- Tokenization là quá trình chia một câu thành các token (từ hoặc nhóm từ) để có thể xử lý và huấn luyện mô hình. Đoạn mã sử dụng bộ token hóa từ mô hình BERT (với AutoTokenizer.from_pretrained("bert-base-multilingual-cased")) để chia câu thành các token.
- Sau khi chia câu thành token, quá trình align labels (đồng bộ nhãn) được thực hiện để gán nhãn cho từng token. Vì một từ có thể được chia thành nhiều token (như từ "email" có thể bị chia thành "em" và "ail"), nên cần đảm bảo nhãn đúng được gán cho các token liên quan.

Chuyển đổi Nhãn BIO sang Dạng Số

Chuyển nhãn BIO thành dạng số: Việc chuyển nhãn BIO thành dạng số giúp mô hình học sâu (deep learning) có thể xử lý dễ dàng hơn. Mỗi nhãn được ánh xạ thành một chỉ số, ví dụ:

- = 0,
- B-EMAIL = 1,
- B-PHONE = 2.
- Nhãn -100 được sử dụng cho các token không có nhãn (chẳng hạn, các token không thuộc bất kỳ thực thể nào).

Mô hình học sâu (Deep Learning) và BERT

- BERT (Bidirectional Encoder Representations from Transformers) là một mô hình học sâu rất mạnh mẽ trong việc xử lý ngôn ngữ tự nhiên. Mô hình BERT-base-multilingual-cased được sử dụng trong đoạn mã, có khả năng hiểu ngữ nghĩa văn bản bằng cách sử dụng cơ chế self-attention.
- AutoModelForTokenClassification được sử dụng để khởi tạo mô hình BERT cho bài toán phân loại token, nơi mô hình dự đoán nhãn cho từng token (B-EMAIL, B-PHONE, O).

Đánh giá mô hình (Metrics)

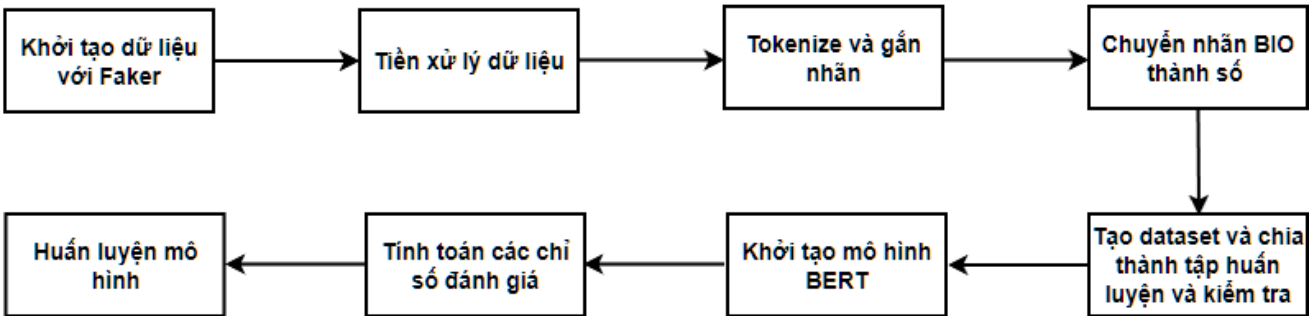
- Đoạn mã sử dụng metric seqeval để đánh giá mô hình, đặc biệt là precision, recall, f1-score, và accuracy cho bài toán NER. Đây là các chỉ số quan trọng khi đánh giá hiệu quả của mô hình NER, giúp đo lường khả năng của mô hình trong việc nhận diện đúng các thực thể.

Huấn luyện mô hình với Hugging Face Trainer

- Trainer là một lớp tiện ích trong Hugging Face giúp đơn giản hóa quá trình huấn luyện mô hình. Đoạn mã sử dụng TrainingArguments để cấu hình các tham số huấn luyện, chẳng hạn như learning rate, batch size, số epoch, và các tùy chọn liên quan đến lưu và ghi log.
- Mô hình được huấn luyện với dữ liệu đã chuẩn bị, và các thông số huấn luyện được tối ưu hóa tự động.

2.3 Quy trình triển khai, kết quả

2.3.1 Quy trình triển khai



Hình 2. 1: Quy trình triển khai

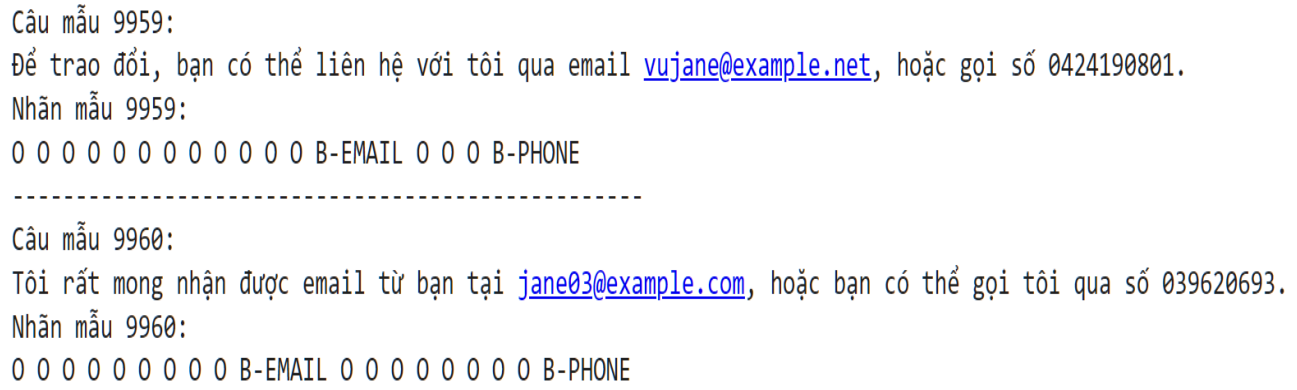
2.3.2 Kết quả

Khởi tạo dữ liệu với Faker.

```
[
    {
        "sentence": "Tôi có thể nhận email tại janemai@example.net, và bạn có thể gọi tôi qua số điện thoại +84466126381.",
        "labels": {
            "email": "janemai@example.net",
            "phone": "+84466126381"
        }
    },
    {
        "sentence": "Thông tin liên lạc của tôi: email jane22@example.org, số điện thoại +84525399885.",
        "labels": {
            "email": "jane22@example.org",
            "phone": "+84525399885"
        }
```

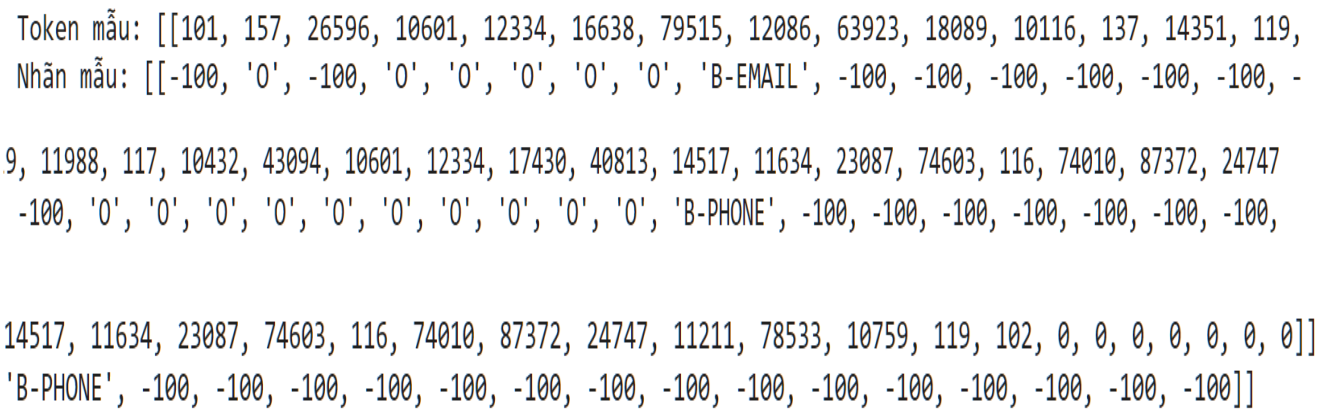

Hình 2. 2: Dữ liệu mẫu sau khi khởi tạo

Tiền xử lý dữ liệu:



Hình 2. 3: Dữ liệu sau khi xử lý

Tokenize và gắn nhãn:



Hình 2. 4: Dữ liệu sau khi tokenize và gắn nhãn

Chuyển nhãn BIO thành số:

```
[[-100, 0, 0, 0, 0, 1, -100, -100, -100, -100, -100, -100, 0, 0, 0, 0, 0, 0, 2, -100, -100, -100, -100, -100, 0, 0,

, -100, -100, -100, -100, -100, -100, -100, -100, -100, -100, -100, -100, -100, -100, -100, -100]]
```

Hình 2. 5: Chuyển nhãn BIO thành số

Kết quả huấn luyện trên tập huấn luyện:

[1500/1500 05:49, Epoch 3/3]

Epoch	Training Loss	Validation Loss	Precision	Recall	F1	Accuracy
1	0.000000	0.000009	1.000000	1.000000	1.000000	1.000000
2	0.000000	0.000006	1.000000	1.000000	1.000000	1.000000
3	0.004600	0.000005	1.000000	1.000000	1.000000	1.000000

Hình 2. 6: Kết quả trên tập huấn luyện

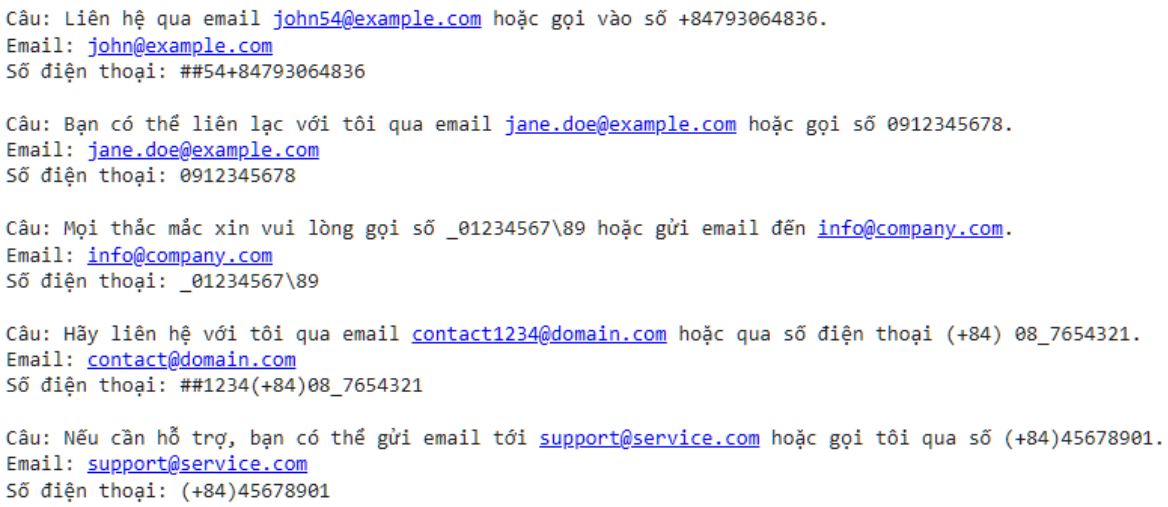
Kết quả trên tập test:

[125/125 00:05]

```
Kết quả đánh giá: {'eval_loss': 5.455024620459881e-06, 'eval_precision': 1.0, 'eval_recall': 1.0, 'eval_f1': 1.0, 'eval_accuracy': 1.0,
```

Hình 2. 7: Kết quả trên tập test

Kết quả thực tế:



Hình 2. 8: Kết quả thực nghiệm

2.4 Kết luận

Kết quả huấn luyện mô hình trong ba Epoch cho thấy một sự cải thiện vượt trội qua từng bước, với các chỉ số đánh giá đạt giá trị tối đa trong cả quá trình huấn luyện và đánh giá. Cụ thể, mô hình đạt được các chỉ số Precision, Recall, F1, và Accuracy đều là 1.0, chứng tỏ khả năng phân loại của mô hình là hoàn hảo và không có sai sót trong quá trình dự đoán trên cả tập huấn luyện và tập kiểm tra.

Trong quá trình huấn luyện, mất mát (loss) giảm mạnh từ Epoch đầu tiên đến Epoch thứ ba. Mặc dù mất mát trong quá trình huấn luyện rất thấp (cụ thể, đạt giá trị gần bằng 0), điều này cho thấy mô hình đã học rất nhanh và ổn định từ những bước đầu tiên. Tuy nhiên, kết quả này cũng có thể là dấu hiệu của overfitting, đặc biệt là khi mô hình đạt được độ chính xác hoàn hảo trong cả tập huấn luyện và tập kiểm tra. Điều này có thể xảy ra khi mô hình quá khớp với dữ liệu huấn luyện mà không thể tổng quát tốt cho các trường hợp dữ liệu mới hoặc chưa thấy.

Thời gian huấn luyện tổng cộng trong ba Epoch là hợp lý, với khoảng 350 giây cho toàn bộ quá trình huấn luyện, và mô hình đã xử lý khoảng 68 mẫu mỗi giây. Mặc dù thời gian huấn luyện ngắn, nhưng mô hình vẫn có thể đạt được hiệu suất rất cao,

điều này phản ánh sự tối ưu hóa tốt từ quá trình huấn luyện. Với tốc độ huấn luyện này, mô hình có thể dễ dàng được triển khai vào các ứng dụng thực tế với khối lượng dữ liệu lớn.

Một điểm đáng chú ý là trong quá trình đánh giá, mô hình cũng đạt kết quả rất xuất sắc. Eval_loss rất thấp (gần bằng 0), và các chỉ số đánh giá như Precision, Recall, F1, và Accuracy đều đạt giá trị 1.0, cho thấy mô hình không chỉ học tốt từ dữ liệu huấn luyện mà còn có thể dự đoán chính xác với dữ liệu kiểm tra chưa thấy trước đó.

Mặc dù kết quả rất ấn tượng, nhưng cần có những biện pháp kiểm tra thêm để xác định xem mô hình có thực sự tổng quát tốt hay không. Việc kiểm tra với dữ liệu ngoài (out-of-sample testing) hoặc sử dụng các kỹ thuật cross-validation sẽ giúp xác định xem mô hình có thực sự chống lại hiện tượng overfitting và có thể hoạt động tốt trên dữ liệu thực tế. Ngoài ra, cũng cần xem xét việc điều chỉnh các tham số như regularization, dropout hoặc tăng cường dữ liệu để làm giảm nguy cơ overfitting trong các bài toán thực tế.

CHƯƠNG 3 – CƠ SỞ LÝ THUYẾT / NGHIÊN CỨU THỰC NGHIỆM

Những nghiên cứu thực nghiệm hoặc lý thuyết: trình bày các cơ sở lý thuyết, lý luận, giả thuyết khoa học và phương pháp nghiên cứu sẽ được sử dụng trong Luận văn, Luận án;

3.1 Chèn bảng:

STT	Tiêu đề A	Tiêu đề B
1	Nội dung 1	Nội dung 4
2	Nội dung 2	Nội dung 5
3	Nội dung 3	Nội dung 6

Bảng 3.1Ví dụ cho chèn bảng

Khi cần chèn tên bảng thì chọn References \ Caption và chọn "Bảng …"

3.2 Viết tắt

Không lạm dụng việc viết tắt. Chỉ viết tắt những từ, cụm từ hoặc thuật ngữ *được sử dụng nhiều lần trong luận văn*. Không viết tắt những cụm từ dài, những mệnh đề hoặc những cụm từ ít xuất hiện. Nếu cần viết tắt những từ, thuật ngữ, tên các cơ quan, tổ chức... thì được viết tắt sau lần viết thứ nhất có kèm theo chữ viết tắt trong ngoặc đơn. Nếu có quá nhiều chữ viết tắt thì phải có bảng danh mục các chữ viết tắt (xếp theo thứ tự A, B, C) ở phần đầu luận văn.

3.3 Trích dẫn

3.3.1 Tài liệu tham khảo và cách trích dẫn

Mọi ý kiến, khái niệm, phân tích, phát biểu, diễn đạt... có ý nghĩa, mang tính chất gợi ý *không phải của riêng tác giả* và mọi tham khảo khác **phải được trích dẫn và chỉ rõ nguồn trong danh mục Tài liệu tham khảo của luận văn**. Phải nêu rõ cả

việc sử dụng những đề xuất hoặc kết quả của đồng tác giả (*đối với công trình đã công bố khác thì phải trích dẫn bình thường như một tài liệu tham khảo*). Nếu sử dụng tài liệu của người khác và của đồng tác giả (bảng biểu, hình vẽ, công thức, đồ thị, phương trình, ý tưởng...) mà không chú dẫn tác giả và nguồn tài liệu thì **luận văn không được duyệt để bảo vệ**.

Không trích dẫn những kiến thức phổ biến, mọi người đều biết tránh làm nặng nề phần tham khảo trích dẫn.

Nếu người dẫn liệu không có điều kiện tiếp cận được một tài liệu gốc mà phải trích dẫn thông qua một tài liệu khác của một tác giả khác, thì phải nêu rõ cách trích dẫn (*lưu ý phải ghi đúng nguyên văn từ chính tài liệu tham khảo và hạn chế tối đa hình thức này).* Nếu cần trích dẫn dài hơn thì phải tách phần này thành một đoạn riêng khỏi phần nội dung đang trình bày, in nghiêng, với lề trái lùi vào thêm 2 cm. Khi mở đầu và kết thúc đoạn trích này không phải sử dụng dấu ngoặc kép. Việc trích dẫn là theo thứ tự của tài liệu ở danh mục Tài liệu tham khảo và được đặt trong ngoặc vuông, khi cần có cả số trang, ví dụ [15, tr.314-315]. Đối với phần trích dẫn từ nhiều tài liệu khác nhau, số của từng tài liệu được đặt độc lập trong từng ngoặc vuông, theo thứ tự tăng dần, ví dụ [19], [25], [41], [42].

3.3.2 Qui định của Khoa Công nghệ thông tin

- Đạo văn là việc sử dụng từ ngữ hay ý tưởng của người khác như là của mình trong hoạt động học thuật nói riêng và trong hoạt động sáng tạo nói chung. Tại Đại học Hoa Sen, những hành vi sau đây được xem là đạo văn:

- Sao chép nguyên văn **02** (hai) câu liên tiếp mà không dẫn nguồn đúng quy định;
- Sao chép nguyên văn **03** (ba) câu không liên tiếp mà không dẫn nguồn đúng quy định;
- Diễn đạt lại (*rephrase*) hoặc dịch (*translate*) toàn bộ một ý nào đó của người khác mà không dẫn nguồn đúng quy định;

- Sử dụng hơn 30% nội dung của một báo cáo cuối kỳ do chính mình viết để nộp cho 2 lớp khác nhau (cùng học kỳ hoặc khác học kỳ) mà không có sự đồng ý của giảng viên;
- Sao chép một phần hoặc toàn bộ bài làm của người khác.

- Khi luận văn, đồ án, bài tập lớn, được chấm điểm, nếu bị phát hiện đạo văn thì ngay lập tức bị điểm 0. Sinh viên sẽ tiếp tục bị xử lý kỷ luật theo các qui định của Nhà trường.

TÀI LIỆU THAM KHẢO

Tiếng Việt

1. Quách Ngọc Ân (1992), "Nhìn lại hai năm phát triển lúa lai", *Di tuyền học ứng dụng*, 98(1), tr. 10-16.
2. Bộ nông nghiệp & PTNT (1996), *Báo cáo tổng kết 5 năm (1992-1996) phát triển lúa lai,* Hà Nội.
3. Nguyễn Hữu Đống, Đào Thanh Bằng, Lâm Quang Dụ, Phan Đức Trực (1997), *Đột biến – Cơ sở lý luận và ứng dụng,* Nhà xuất bản nông nghiệp, Viện khoa học kỹ thuật nông nghiệp Việt Nam, Hà Nội.
4. Nguyễn Thị Gấm (1996), *Phát hiện và đánh giá một số dòng bất dục đực cảm ứng nhiệt độ,* Luận văn thạc sĩ khoa học nông nghiệp, Viện khoa học kỹ thuật nông nghiệp Việt Nam, Hà Nội.

…….

23. Võ Thị Kim Huệ (2000), *Nghiên cứu chẩn đoán và điều trị bệnh…,* Luận án Tiến sĩ y khoa, Trường đại học y Hà Nội, Hà Nội.

Tiếng Anh

28. Anderson J.E. (1985), The Relative Inefficiency of Quota, The Cheese Case, *American Economic Review*, 75(1), pp. 178-90.
29. Borkakati R. P.,Virmani S. S. (1997), Genetics of thermosensitive genic male sterility in Rice, *Euphytica* 88, pp. 1-7.
30. Boulding K.E. (1955), *Economics Analysis*, Hamish Hamilton, London.
31. Burton G. W. (1988), "Cytoplasmic male-sterility in pearl millet (penni-setum glaucum L.)", *Agronomic Journal* 50, pp. 230-231.
32. Central Statistical Oraganisation (1995), *Statistical Year Book*, Beijing.
33. FAO (1971), *Agricultural Commodity Projections (1970-1980)*, Vol. II. Rome.

34. Institute of Economics (1988), *Analysis of Expenditure Pattern of Urban Households in Vietnam,* Departement pf Economics, Economic Research Report, Hanoi.

PHỤ LỤC

Phần này bao gồm những nội dung cần thiết nhằm minh họa hoặc hỗ trợ cho nội dung luận văn như số liệu, biểu mẫu, tranh ảnh. . . . nếu sử dụng những câu trả lời cho một *bảng câu hỏi thì bảng câu hỏi mẫu này phải được đưa vào phần Phụ lục ở dạng nguyên bản* đã dùng để điều tra, thăm dò ý kiến; **không được tóm tắt hoặc sửa đổi**. Các tính toán mẫu trình bày tóm tắt trong các biểu mẫu cũng cần nêu trong Phụ lục của luận văn. Phụ lục không được dày hơn phần chính của luận văn