

Classification of TV Programs According to Their Topics Using the XLNet Model

Quang-Long Vo¹, Ngoc-Hung-Anh Nguyen², Minh-Son Tran³, Thu-Ha Tran^{1*}

¹Ho Chi Minh City University of Technology and Education, Vietnam

²Posts and Telecommunications Institute of Technology, Vietnam

³University of Technology HUTECH, Vietnam

*Corresponding author: thuha@hcmute.edu.vn

ARTICLE INFO

Received: 4/3/2022
Revised: 21/3/2022
Accepted: 24/3/2022
Published: 28/4/2022

KEYWORDS

Classification;
TV channel;
Artificial intelligence;
XLNet;
Deep learning.

ABSTRACT

While observing the behavior of TV viewer, the HCMUTE researchers concluded that TV spectators rather watch the content of the TV channel than the channels themselves. The latters are just the containers / carriers of the precious TV information inside. Therefore to enable TV viewers to access directly the content, a new TV user graphic interface is offered to them: the TV program list is ordered by the topics of the broadcasted programs but not the channels carrying them. Therefore, viewers can now choose a program according to topics of interest, not via an intermediate step by selecting channel with dummy name / number. To classify all the broadcast programs, the researchers propose a sequence-to-sequence Model to organize these programs into one of five predefined topics / themes: feature films, news, music, sports and synthesis. The XLNet network is incorporated in the classification. TV channels currently playing the program with the selected theme will be displayed so that viewers can quickly access the wish content.

Phân Loại Tên Chương Trình Truyền Hình Theo Chủ Đề Phát sóng Sử Dụng Mô Hình XLNet

Võ Quang Long¹, Nguyễn Ngọc Hùng Anh², Trần Minh Sơn³, Trần Thu Hà^{1*}

¹Trường Đại học Sư phạm Kỹ thuật TP.HCM, Việt Nam

²Học viện Công nghệ Bưu chính Viễn thông, Việt Nam

³Trường Đại học Công nghệ, TP HCM, Việt Nam.

* Tác giả liên hệ: thuha@hcmute.edu.vn

THÔNG TIN BÀI BÁO

Ngày nhận bài: 4/3/2022
Ngày hoàn thiện: 21/3/2022
Ngày chấp nhận đăng: 24/3/2022
Ngày đăng: 28/4/2022

TỪ KHÓA

Phân loại;
Kênh truyền hình;
Trí tuệ nhân tạo;
XLNet;
Học sâu.

TÓM TẮT

Bài báo giới thiệu một dịch vụ mới trong truyền hình số dựa trên trải nghiệm thực tế khách hàng khi xem truyền hình. Mục đích chính của khán giả truyền hình là xem nội dung truyền tải về các chủ đề mà mình có nhu cầu. Để tạo điều kiện cho việc chọn lựa nội dung truyền hình muốn xem một cách nhanh chóng, nhóm nghiên cứu đã đề xuất giải pháp sắp xếp phân loại chương trình truyền hình theo chủ đề nội dung đang phát của kênh giúp người xem chọn được và xem được chủ đề truyền hình quan tâm, không phải tìm kiếm giữa nhiều kênh đang truyền (số thứ tự kênh) không có gợi ý về nội dung đang phát trực tiếp. Để phân loại tất cả tên các chương trình phát sóng, bài báo đã ứng dụng mô hình XLNet chuyển đổi giữa các câu văn (sequence-to-sequence Model) để sắp xếp các chương trình này vào một trong năm chủ đề được định nghĩa trước: phim truyện, thời sự, ca nhạc, thể thao, tổng hợp,... Các kênh truyền hình đang phát chương trình có chủ đề được chọn sẽ được hiển thị để người xem có thể nhanh chóng xem được đúng nội dung cần tìm.

Doi: <https://doi.org/10.54644/jte.69.2022.1146>

Copyright © JTE. This is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purpose, provided the original work is properly cited.

1. Giới thiệu

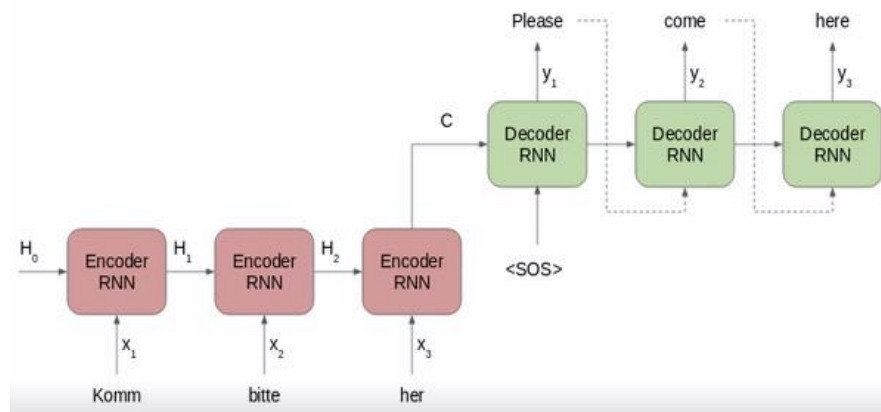
Phân loại tên của chương trình phát sóng có thể được quy đổi về bài toán lớn hơn là phân loại văn bản, phân loại câu văn hay từ vựng. Đây là các bài toán cơ bản của Xử lý Ngôn ngữ Tự nhiên (NLP Natural Language Processing).

Bài toán phân loại tên chương trình được mô hình hóa trong học sâu (deeplearning) với mô hình chuyển đổi giữa các câu văn (sequence-to-sequence Model).

Để hiểu rõ hơn về mô hình sequence-to-sequence, Hình 1 mô tả quá trình học sâu giải quyết bài toán dịch 1 câu văn tiếng Đức sang tiếng Anh dựa trên mạng RNN (Recurrent Neural Network).

Quy trình xử lý có thể được tóm tắt như sau:

- Mạng xử lý gồm nhiều mắt xích, mỗi mắt xích lại là một mạng kết nối toàn phần RNN được tối ưu hóa cho bài toán Natural Language Processing tiếng Đức.



Hình 1. Cấu trúc bên trong của một mô hình học sâu dùng cho phân tích câu văn.

- Các từ (X_i trong Hình 1) trong một câu sẽ được đưa vào mắt xích tương ứng như 1 vector của các chữ cái. Các mắt xích mạng “thần kinh” (Neural) này cũng sẽ chịu ảnh hưởng thêm từ một trạng thái trong (H_i trong Hình 1).

Trạng thái trong này sẽ được cập nhật theo từng từ của câu văn cần phân tích.

Trạng thái trong cuối cùng có thể coi là “ý nghĩa chính” (CV Context Vector) của câu văn đầu vào.

CV này được truyền tiếp đến bộ phận “giải mã” cũng là các mắt xích RNN tối ưu hóa cho Natural Language Processing tiếng Anh.

Và như thế câu văn đầu vào tiếng Đức được dịch ra tiếng Anh nhờ quy trình trên.

Mạng xử lý như trên có thể dựa trên BERT (Bidirectional Encoder Representations from Transformers), là một dạng mô hình mới của Google AI cho Natural Language Processing. BERT dùng thông tin về ngôn ngữ được học trước để xử lý các bài toán như máy tự trả lời, phân tích cảm xúc câu trả lời, tìm ý chính đoạn văn,... Tuy nhiên, trong phương pháp của mình, BERT phải giấu đi 15% số từ trong chuỗi đầu vào, điều này làm mất thông tin về quan hệ giữa các từ. Do đó có một sự khác biệt lớn giữa mô hình đào tạo trước (pre-trained) và các phiên bản điều chỉnh (fine-tuning) của nó.

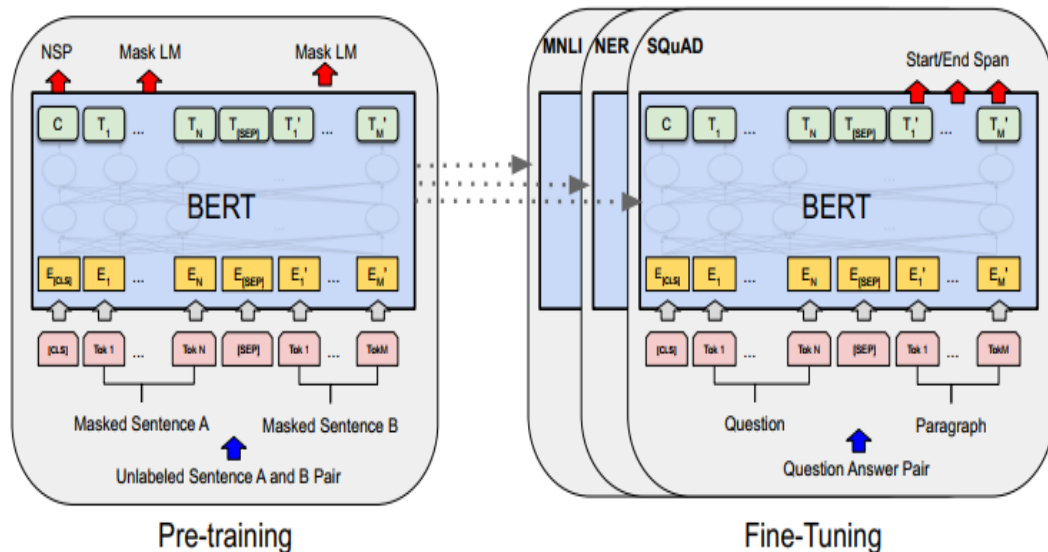
Để khắc phục vấn đề này, một mạng xử lý khác cũng đang phát triển mạnh là XLNet. Mạng này được cho rằng đã tận dụng tốt hơn BERT sự liên quan ngữ nghĩa với vị trí của từ vựng trong câu văn. XLNet cũng được cho là kháng chịu tốt hơn với điều kiện mất cân đối, không ổn định của các tập thông tin dùng cho việc luyện cho mạng “thần kinh” này. Mặc dù vậy, các mô hình vượt qua BERT bao gồm cả XLNet cũng đơn thuần là những sửa đổi trên cơ sở của người khổng lồ này; về bản chất thì kiến trúc mô hình và các nhiệm vụ không có nhiều sự thay đổi. Với cách tiếp cận khác, các tác giả của XLNet phân tích ưu và nhược của mô hình đào tạo trước dựa trên tự hồi quy (AR: auto recursive) và tự mã hóa (AE: auto encoding) và đề xuất ra mô hình vượt trội này với hi vọng có thể kết hợp được tất cả những ưu điểm tốt nhất của hai phương pháp này [1].

2. BERT

BERT [2] là mô hình biểu diễn từ vừng theo 2 chiều ứng dụng kỹ thuật Transformer. BERT được thiết kế để huấn luyện trước các biểu diễn hai chiều từ văn bản không được gắn nhãn bằng cách điều chỉnh chung cả 2 chiều trái và phải cho tất cả các lớp. BERT thể hiện các tính năng vượt trội thông qua các tác vụ thử nghiệm đánh giá ngôn ngữ tự nhiên khác nhau (GLUE, MultiNLI, SQuAD v1.1, ...).

2.1 Tinh chỉnh mô hình BERT

Một điểm đặc biệt ở BERT đó là kết quả huấn luyện có thể tinh chỉnh được chỉ với một lớp ngõ ra được thêm vào để tạo ra các mô hình vượt trội cho một loạt các nhiệm vụ, chẳng hạn như trả lời câu hỏi và suy luận ngôn ngữ.



Hình 2. Quy trình tổng thể pre-training và fine-tuning của BERT

Chúng một kiến trúc được sử dụng cho cả pretrain-model (mô hình đào tạo trước) và fine-tuning model (mô hình tinh chỉnh). Nhóm nghiên cứu sử dụng cùng một tham số huấn luyện trước để khởi tạo mô hình cho các tác vụ ứng dụng khác nhau. Trong suốt quá trình điều chỉnh thì toàn bộ các tham số của các lớp học chuyển giao sẽ được tinh chỉnh. [CLS] là một ký tự (token) đặc biệt được thêm vào trước mọi ví dụ đầu vào và [SEP] là một ký tự phân tách đặc biệt (ví dụ: phân tách các câu hỏi/câu trả lời).

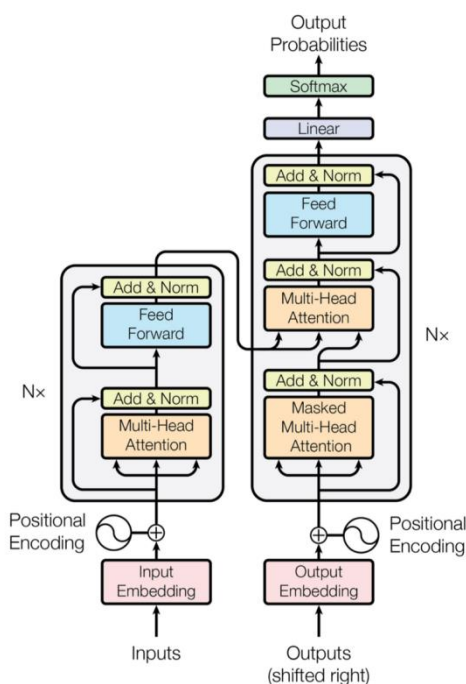
2.2 Đào tạo trước mô hình BERT

Các mô hình ngôn ngữ truyền thống từ trái sang phải hoặc từ phải sang trái không được sử dụng để Pre-training BERT mà sử dụng hai nhiệm vụ không được giám sát (unsupervised tasks). Bước này được trình bày ở phần bên trái của Hình 2.

Nhiệm vụ số 1: Masked LM (MLM) là một tác vụ cho phép tinh chỉnh lại các biểu diễn từ trên các bộ dữ liệu văn bản không được giám sát bất kỳ. MLM có thể được áp dụng cho những ngôn ngữ khác nhau để tạo ra biểu diễn nhúng cho chúng (tóm tắt cho các mệnh đề).

Nhiệm vụ số 2: Next Sentence Prediction (NSP) là một bài toán phân loại học có giám sát với 2 nhãn. Đầu vào của mô hình là một cặp câu sao cho 50% câu thứ 2 được lựa chọn là câu tiếp theo của câu thứ nhất và 50% được lựa chọn một cách ngẫu nhiên từ bộ văn bản mà không có mối liên hệ gì với câu thứ nhất. Nhãn của mô hình sẽ tương ứng với IsNext (đúng là câu tiếp theo) khi cặp câu là liên tiếp hoặc NotNext (không phải là câu tiếp theo) nếu cặp câu không liên tiếp.

Cũng tương tự như mô hình Question and Answering (Hỏi và đáp), chúng ta cần đánh dấu các vị trí đầu câu thứ nhất bằng ký tự [CLS] và vị trí cuối các câu bằng ký tự [SEP]. Các ký tự này có tác dụng nhận biết các vị trí bắt đầu và kết thúc của từng câu thứ nhất và thứ hai.



Hình 3. Sơ đồ kiến trúc mô hình BERT cho tác vụ NSP.

3. XLNet

XLNet [3] là một phương pháp học biểu diễn ngôn ngữ không giám sát mới dựa trên mục tiêu mô hình hóa ngôn ngữ hoán vị tổng quát mới, được xây dựng dựa trên phương pháp tự hồi quy (AR) tổng quát.

XLNet giống như BERT được đào tạo trước. Có thể coi XLNet như một phiên bản nâng cao của BERT, nó hoạt động tốt hơn BERT về một số nhiệm vụ xử lý ngôn ngữ tự nhiên bao gồm phân loại văn bản, trả lời câu hỏi và các nhiệm vụ khác.

XLNet sử dụng Transformer XL làm kiến trúc trích xuất tính năng, tốt hơn so với BERT, vì Transformer XL đã thêm tính năng lặp lại vào Transformer [4]. Điều này có thể làm cho XLNet hiểu sâu hơn về ngữ cảnh ngôn ngữ.

Sử dụng XLNet cho một tác vụ cụ thể rất đơn giản, chúng ta có thể tải xuống mô hình XLNet được đào tạo trước [5], sau đó sử dụng phương pháp tinh chỉnh để cập nhật mô hình được đào tạo trước cho phù hợp với tác vụ ứng dụng cụ thể.

3.1 XLNet phân loại văn bản

Quá trình thực hiện phân loại văn bản với XLNet bao gồm 4 bước:

Bước 1: Tải dữ liệu: Để phân loại văn bản, chúng ta cần tập dữ liệu mà văn bản đã được gắn nhãn.

Bước 2: Đặt dữ liệu vào nhúng đào tạo: Sau khi nhận được dữ liệu, văn bản sắp xếp thành 3 loại nhúng: Nhúng ký tự (Token embedding), Nhúng từ mặt nạ (Mask word embedding), Nhúng phân đoạn (Segmentation embedding). Quá trình thực hiện nhúng cho XLNet khác với BERT, đầu tiên các văn bản sẽ được mã hóa bằng câu văn (sentencepiece); sau đó thêm “<sep>”, “<cls>” và mặt nạ (pad mask) vào các nhúng.

Bước 3: Mô hình huấn luyện: Khi sử dụng học chuyển đổi như XLNet, quá trình đào tạo một mô hình mới với dữ liệu kết quả được gọi là “fine-tuning”, tất cả những gì cần làm là chọn một trong các mô hình được đào tạo trước XLNet và sử dụng dữ liệu hiện có để cập nhật tham số của mô hình để thực hiện nhiệm vụ đề ra. Đối với ngôn ngữ tiếng Anh, XLNet có 2 loại mô hình, mô hình cơ sở và mô hình lớn. Mô hình phiên bản lớn là mô hình lớn hơn so với mô hình cơ sở, có hiệu suất tốt hơn, nhưng cần nhiều tài nguyên và thời gian tính toán hơn.

Bước 4: Đánh giá hiệu suất mô hình: Sau khi đào tạo một mô hình mới để phân loại văn bản, muốn biết mô hình đó tốt như thế nào, chúng ta có thể đánh giá mô hình với dữ liệu mới. Dữ liệu đánh giá có thể được đặt trong quy trình khi chúng ta đặt nhóm dữ liệu đào tạo trước đây. Tốt hơn nên sử dụng 30% dữ liệu làm dữ liệu thử nghiệm để xác thực hiệu suất.

3.2 So sánh

❖ XLNet với BERT

Được đào tạo trên cùng một dữ liệu với công thức đào tạo gần như giống hệt nhau, XLNet vượt trội hơn BERT bởi một biên độ khá lớn trên tất cả các bộ dữ liệu được xem xét (Bảng 1).

Bảng 1. So sánh với BERT ([1]).

Model	SQuAD1.1	SQuAD2.0	RACE	MNLI	QNLI	QQP	RTE	SST-2	MRPC	CoLA	STS-B
BERT-Large (Best of 3)	86.7/92.8	82.8/85.5	75.1	87.3	93.0	91.4	74.0	94.0	88.7	63.7	90.2
XLNet-Large- wikibooks	88.2/94.0	85.1/87.8	77.4	88.4	93.9	91.8	81.2	94.4	90.0	65.2	91.1

❖ XLNet với RoBERTa

Kết quả được trình bày trong các Bảng 2, 3, 4 và 5 cho thấy XLNet thường vượt trội hơn BERT và RoBERTa.

Đối với các tác vụ lý luận rõ ràng như SQuAD và RACE liên quan đến ngữ cảnh dài hơn, mức tăng hiệu suất của XLNet thường lớn hơn. Sự vượt trội này có thể đến từ cấu trúc chính (backbone) Transformer-XL trong XLNet.

Đối với các tác vụ phân loại đã có nhiều ví dụ được giám sát như MultiNLI, Yelp và Amazon, XLNet vẫn có mức tăng đáng kể.

Bảng 2. So sánh với kết quả state-of-the-art trên bộ kiểm tra RACE.

RACE	Accuracy	Middle	High	Model	NDCG@20	ERR@20
GPT	59.0	62.9	57.4	DRMM	24.3	13.8
BERT	72.0	76.6	70.1	KNRM	26.9	14.9
BERT+DCMN*	74.1	79.5	71.8	Conv	28.7	18.1
RoBERTa	83.2	86.5	81.8	BERT [†]	30.53	18.67
XLNet	85.4	88.6	84.0	XLNet	31.10	20.28

Bảng 3. Kết quả trên QuAD.

SQuAD2.0	EM	F1	SQuAD1.1	EM	F1
<i>Dev set results (single model)</i>					
BERT	78.98	81.77	BERT [†]	84.1	90.9
RoBERTa	86.5	89.4	RoBERTa	88.9	94.6
XLNet	87.9	90.6	XLNet	89.7	95.1
<i>Test set results on leaderboard (single model, as of Dec 14, 2019)</i>					
BERT	80.005	83.061	BERT	85.083	91.835
RoBERTa	86.820	89.795	BERT*	87.433	93.294
XLNet	87.926	90.689	XLNet	89.898[‡]	95.080[‡]

Bảng 4. So sánh với tỷ lệ lỗi state-of-the-art trên các bộ thử nghiệm của một số bộ dữ liệu phân loại văn bản.

Model	IMDB	Yelp-2	Yelp-5	DBpedia	AG	Amazon-2	Amazon-5
CNN	-	2.90	32.39	0.84	6.57	3.79	36.24
DPCNN	-	2.64	30.58	0.88	6.87	3.32	34.81
Mixed VAT	4.32	-	-	0.70	4.95	-	-
ULMFiT	4.6	2.16	29.98	0.80	5.01	-	-
BERT	4.51	1.89	29.32	0.64	-	2.63	34.17
XLNet	3.20	1.37	27.05	0.60	4.45	2.11	31.67

Bảng 5. Kết quả trên GLUE.

Model	MNLI	QNLI	QQP	RTE	SST-2	MRPC	CoLA	STS-B	WNLI
<i>Single-task single models on dev</i>									
BERT	86.6/-	92.3	91.3	70.4	93.2	88.0	60.6	90.0	-
RoBERTa	90.2/90.2	94.7	92.2	86.6	96.4	90.9	68.0	92.4	-
XLNet	90.8/90.8	94.9	92.3	85.9	97.0	90.8	69.0	92.5	-
<i>Multi-task ensembles on test (from leaderboard as of Oct 28, 2019)</i>									
MT-DNN*	87.9/87.4	96.0	89.9	86.3	96.5	92.7	68.4	91.1	89.0
RoBERTa*	90.8/90.2	98.9	90.2	88.2	96.7	92.3	67.8	92.2	89.0
XLNet*	90.9/90.9[†]	99.0[†]	90.4[†]	88.5	97.1[†]	92.9	70.2	93.0	92.5

4. Ứng dụng XLNet vào OTT3A

4.1. Giới thiệu OTT3A

Over The Top (OTT) là giải pháp cung cấp nội dung cho người sử dụng dựa trên nền tảng Internet [1]. Lĩnh vực được ứng dụng nhiều nhất là cung cấp các nội dung truyền hình trực tuyến và các Video theo yêu cầu (VOD) tới người dùng cuối [4][6]. Ưu thế lớn nhất của công nghệ Over The Top là việc cho phép cung cấp nguồn nội dung phong phú và đa dạng theo yêu cầu của người sử dụng vào bất kỳ thời điểm nào và tại bất cứ nơi đâu chỉ với một thiết bị phù hợp có kết nối Internet cung cấp bởi bất cứ nhà mạng nào [7][8]. Ngoài ra, công nghệ này – thừa hưởng đầy đủ các ưu điểm của IPTV - còn cung cấp nhiều loại công cụ tiện ích khác mang tính ứng dụng cao như: VoIP, Mạng xã hội,... Với nhiều ứng dụng thiết thực, công nghệ Over The Top được dự báo sẽ còn phát triển mạnh trong tương lai và trở thành một trong những xu thế công nghệ thế giới [9].

Hệ thống OTT3A [10] mang 3 tính tương thích đặc trưng như sau:

Tương thích với đường truyền Internet (A1): Tự động điều chỉnh độ rộng băng tần thu (đồng nghĩa với chất lượng thu hình) một cách tối ưu theo chất lượng thực tế đường truyền. Giải quyết bài toán tối ưu cho việc chọn lựa chất lượng nội dung phù hợp nhất với sự thay đổi băng thông của đường truyền mà vẫn bảo đảm chất lượng hình ảnh.

Tương thích với số lượng kênh phát (A2): chỉ hiển thị các kênh có tín hiệu bảo đảm; Có khả năng điều chỉnh, thay đổi số lượng các kênh TV hết sức linh hoạt và tức thời nhằm phục vụ tối đa các kênh TV cho người xem. Phân loại các kênh TV theo nội dung đang phát, không nhất thiết phải theo nhóm của từng đài phát (VTV, VTC, HTV,...) nhằm giúp người xem nhanh chóng bật được chương trình yêu thích.

Tương thích chọn nguồn phát (A3): tối ưu khả năng chọn nguồn phát kênh khi được khách hàng chọn lựa, hệ thống OTT3A sẽ giám sát và truyền lệnh chọn nguồn phát từ máy thu gần nhất hoặc từ máy chủ trên mạng internet chung dựa trên cơ sở kết hợp kỹ thuật mạng đồng đẳng Peer to Peer (P2P) [11] hoặc mạng Server client.

4.2. Ứng dụng vào OTT3A

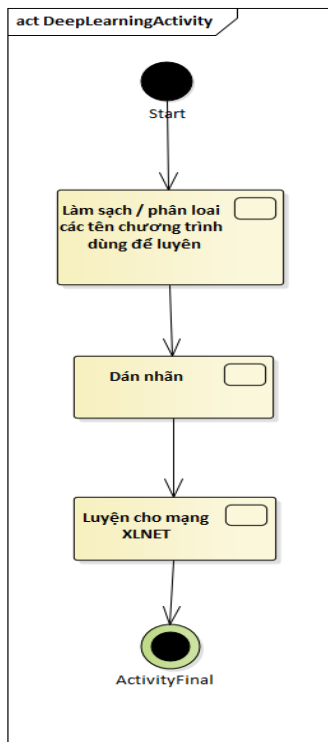
Trong khuôn khổ của bài báo OTT3A, để phân loại tên chương trình vào các chủ đề cho trước, mạng xử lý được chọn là XLNET dựa trên các lý do trên.

Bài toán phân loại tên chương trình phát sóng cũng là một bài toán rất đặt biệt. Chính vì thế, dù mạng XLNET đã được "học trước" (và học cả với ngôn ngữ Việt Nam), nhóm làm bài báo đã luyện thêm cho mạng này.

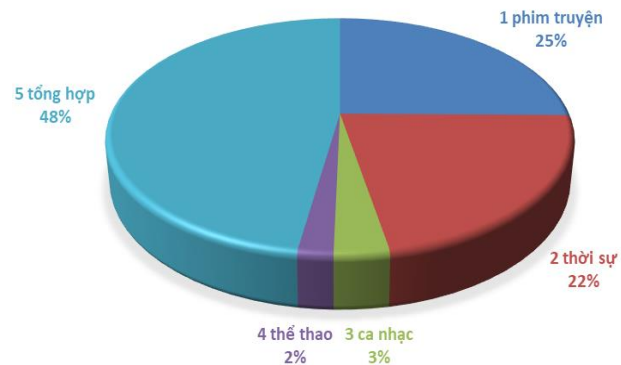
Mẫu tên chương trình được quét tự động từ trang web [12]. Sau đây quy trình như trong Hình 4 được áp dụng để định hướng cho XLNET vào bài toán đặc thù của OTT3A.

Với đầu vào là 1000 tên chương trình được phân loại thành 5 chủ đề chính như trên Hình 5, mô hình XLNET sau đây được dùng để đánh giá / phân loại các tên chương trình phát sóng hàng tuần rồi sẽ được ghi nhận vào hệ thống OTT3A CMS để phân bổ / sắp xếp chương trình phát sóng theo chủ đề.

Với mô hình XLNET đã huấn luyện thêm như trên, 961 tên chương trình được cho vào để mạng dự đoán ra chủ đề của từng chương trình. Kết quả thu được tổng kết như trên Hình 6.



Hình 4. Quy trình luyện thêm cho mạng XLNET với các tên chương trình phát sóng.



Hình 5. Sự mất cân đối trong tập tin dùng để luyện cho XLNET

Count of predict_label	Column Labels	Correction	
Row Labels	ca_nhac	phim_truyen thoi_su tong_hop Grand Total	
ca_nhac	1	11 2 17 31 3.22580645	
phim_truyen		227 1 14 242 93.8016529	
the_thao			6 14 20 0
thoi_su		3 74 134 211 35.07109	
tong_hop		1 21 435 457 95.1859956	
Grand Total	1	242 104 614 961	

Hình 6. Kết quả dự đoán phân loại tên chương trình phát sóng

Cột Row Labels là tên đúng của một tựa đề chương trình phát sóng. Bốn cột tiếp theo ghi lại số lần tựa đề của một dòng nào đấy được đánh giá (sai hay đúng) thuộc về giá trị của cột này. Ví dụ ở dòng có giá trị "ca_nhac" (chủ đề Ca nhạc) trong cột Row Labels, sẽ chỉ có 1 tựa đề được phân loại đúng là Ca nhạc: có số 1 ở cột ca_nhac. Còn lại thì 11 tựa đề coi là phim truyện (phim_truyen), 2 là thời sự (thoi_su) và 17 là tổng hợp (tong_hop). Như vậy xác suất đánh giá chính xác cho các tựa đề thuộc nhóm Ca nhạc là 3,23% (cột Correction).

Có thể thấy trừ chủ đề phim truyện và tổng hợp có xác suất chính xác hơn 90%, các nhóm chủ đề khác đều khá kém, nhất là nhóm Thể thao (0%). Cũng có thể nhận thấy nhóm này chiếm tỉ lệ nhỏ nhất trong tập thông tin dành huấn luyện cho XLNET.

Tên chương trình truyền hình sau khi được đánh giá phân loại tự động, sẽ được kiểm duyệt thủ công 1 lần nữa để chỉnh sửa cho chính xác chủ đề của các chương trình phát sóng. Sau đó, danh sách các tên chương trình này cùng các thông tin phụ trợ như thời gian, thông tin mô tả chính,... sẽ được ghi nhận vào hệ thống quản lý nội dung CMS của hệ thống OTT3A. Người xem có thể tìm nội dung cần xem qua chủ đề nhờ các bước xử lý như trên.

5. Kết luận

Một tính năng quan trọng trong đặc tính tương thích nội dung phát sóng là khả năng giúp người xem tìm ra nội dung muốn xem dựa trên chủ đề của chương trình phát sóng. Như vậy người dùng sẽ không phải chọn trước hết 1 kênh truyền hình nào đấy để lúc đó mới biết nội dung đang chiếu có hợp với mong muốn của mình không. Bằng tính năng tương thích theo nội dung, người dùng được trực tiếp chọn chủ đề mình muốn xem để tiếp cận thẳng vào nội dung. Kênh truyền hình chỉ là lựa chọn thứ cấp khi có nhiều nội dung chung chủ đề và phát trên các kênh khác nhau.

Để thực hiện được tính năng này, 1 vấn đề đặt ra là sự phân loại chính xác và nhanh chóng toàn bộ tên của chương trình phát sóng. Vì chương trình phát sóng thường được lên kế hoạch trước 2 tuần, khoảng 3000 tựa đề cần được phân loại một cách chuẩn xác và nhanh chóng. Bài báo đã vận dụng bài toán Natural Language Processing để xử lý việc phân loại này.

Cũng cần phải thấy rõ đây là 1 bài toán rất đặc thù và chưa được giải quyết như 1 vấn đề cổ điển trong xử lý ngôn ngữ tự nhiên. Tên tựa đề thường là 1 mệnh đề không trọn vẹn (không đủ thành 1 câu trọn nghĩa), việc phân tích bằng học sâu sẽ không thể cho kết quả rất tốt. Tựa đề là tiếng Việt Nam, các kiểu chơi chữ trong tựa đề tựa đề thêm súc tích, ... cũng gây ảnh hưởng không nhỏ đến kết quả dự đoán của hệ thống. Việc mất cân bằng trong chủ đề phát sóng của các đài (thể thao và ca nhạc chỉ chiếm phần nhỏ trong toàn chương trình truyền hình) cũng tác động đến sự sai lệch trong phỏng đoán chủ đề.

Việc sử dụng mô hình XLNet được huấn luyện trước với nhiều từ tiếng Việt hơn ([13], [14]), việc tự động tìm kiếm qua các nguồn tìm kiếm trên Internet để hiểu thêm môi trường cho những tựa đề của chương trình, làm phong phú thêm về số lượng và chất lượng của tập tin học thêm cho mô hình AI,... là những hướng đi khả thi và hứa hẹn cũng tốt hơn việc nhận biết/phân loại tên chương trình một cách nhanh chóng và chính xác.

TÀI LIỆU THAM KHẢO

- [1] XLNet–Người không lờ thay thế vị trí của BERT, 06/07/2019, URL: <https://trituenhantao.io/kien-thuc/xlnet-nguoi-khong-lo-thay-the-vi-tri-cua-bert/>
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
- [3] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, Quoc V. Le, XLNet: Generalized Autoregressive Pretraining for Language Understanding, Đại học Carnegie Mellon, Nhóm trí tuệ AI của Google.C. Waldenor, Is OTT Disrupting Television? Master Thesis, Stockholm, June 7th 2013.
- [4] Giới thiệu về chuyên đổi câu trong xử lý ngôn ngữ tự nhiên Transformers In NLP | State-Of-The-Art-Models (analyticsvidhya.com)
- [5] Mã nguồn của mạng XLNET <https://github.com/zihangdai/xlnet>. J. C. Whitaker, Interactive TV Demystified, 2001 McFraw-Hill, ISBN 0-07-136325-4
- [6] A. PUNCHIHEWA, Tutorial on IPTV and its latest developments, ICIAFS January 2011.
- [7] L. BRINGUIER, White Paper OTT Streaming – 2nd edition, September 2011, Anevia
- [8] T. OHANIAN, Over-the-Top Considerations: Functionalities and Technologies, Cisco Systems, NAB 2014.
- [9] D. RIDDLESSEN, A. D. SINGLETON, Broadband speed equity: A new digital divide, 2014, Applied Geography, Elsevier.
- [10] Trang web chính thức của dự án OTT: <https://ott.atlassian.net>
- [11] Y. CUI, B. LI, K. NAHRSTEDT, "Layered Peer-to-Peer Streaming", Proc. NOSSDAV'03, Monterey, USA, pp. 162-171, Jun 2003.
- [12] Lịch phát sóng các đài truyền hình VN: Lịch phát sóng truyền hình hàng ngày 24/7 nhanh nhất | LịchPhatSongTiVi
- [13] Mô hình phân tích đoạn văn tiếng Việt: PhoBERT: The first public large-scale language models for Vietnamese – VinAI Research
- [14] Xử lý ngôn ngữ tự nhiên dùng mô hình đa ngôn ngữ: XLM-RoBERTa: The alternative for non-english NLP | by Branden Chan | deepset-ai | Medium



Vo Quang Long was born in 1996. Majoring in Electronics and Communication, graduated in 2019 from Ho Chi Minh City University of Technology and Education. Email longvo2396@gmail.com , Contact phone 0777116785



Nguyen Ngoc Hung Anh was born in 1989 in Long An, He graduated in major in Information Systems at Posts and Telecommunications Institute of Technology, in Ho Chi Minh City, 2020. Graduated with a Master's Degree in 2022; Email hunganhbc@gmail.com. Phone 0902717074



Son Minh Tran was born in 1973. He received the PhD degree in Telecommunication in 2004 and became an assistant professor at the Budapest University of Technology, Hungary till 2006. He then worked as a research fellow on video compression algorithms and enhanced features for digital video broadcasting at the ARTEMIS Department at TELECOM & Management SudParis, France. He spent 10 years doing research on video security in Nagra France, Group Kudelski, leader in DTV security. Rejoining Ho Chi Minh University of Technology and Education as senior research fellow since 2018, he continues his activity on video processing. He has a portfolio of 14 granted and pending patents on the Secured Video Transmission/ Processing and watermarking.



Tran Thu Ha was born in 1966, holds in PhD Industrial electronics, automation - telecommunications in 1996 at Kiev Polytechnic University, Ukraine. After that, she worked at the Ho Chi Minh city University of Technology and Education since 1997 as lecturer at Electrical and electronics department. She got a title Assoc. Prof. PhD in 2011. She spent the time for researching and teaching in the fields: in electronics industrial for using neural network, fuzzy logic in auto-control, sliding control, PID, robot, voice control, IOT noise cancellation. In the area of telecommunications: conducting research some problems on teletex, image processing, noise cancellation in signal transmission systems; research on Signal processing, watermarking video security, video processing.