SmartSales: An AI-Powered Telemarketing Coaching System in FinTech

Yuanfeng Song[1,2], Xuefang Zhao[2], Di Jiang[2], Xiaoling Huang[2], Weiwei Zhao[2]
Qian Xu[2], Raymond Chi-Wing Wong[1], Qiang Yang[1,2]
[1]The Hong Kong University of Science and Technology [2]AI Group, WeBank Co., Ltd, China
{songyf,raywong,qyang}@cse.ust.hk

ABSTRACT

Telemarketing is a primary and mature method for enterprises to solicit prospective customers to buy products or services. However, training telesales representatives is always a pain point for enterprises since it is usually conducted manually and costs great effort and time. In this demonstration, we propose a telemarketing coaching system named *SmartSales* to help enterprises develop better salespeople. Powered by artificial intelligence (AI), SmartSales aims to accumulate the experienced sales pitch from customer-sales dialogues and use it to coach junior salespersons. To the best of our knowledge, this is the first practice of an AI telemarketing coaching system in the domain of Chinese FinTech in the literature. SmartSales has been successfully deployed in the WeBank's telemarketing team [1]. We expect that SmartSales will inspire more research on AI assistant systems.

CCS CONCEPTS

• **Information systems** → **Expert systems**;

KEYWORDS

AI-Powered coaching system, Telemarketing, FinTech

ACM Reference Format:
Yuanfeng Song, Xuefang Zhao, Di Jiang, Xiaoling Huang, Weiwei Zhao, Qian Xu, Raymond Chi-Wing Wong, Qiang Yang. 2021. SmartSales: An AI-Powered Telemarketing Coaching System in FinTech. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21), October 20–24, 2021, Virtual Event, China.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3474085.3478555

1 INTRODUCTION

Recent years have witnessed a huge rise in artificial intelligence (AI) coaching system in the sports [1, 10] and education [8] area. However, when it comes to telemarketing, the primary method for FinTech enterprises to solicit prospective customers to buy products, the telesales representatives training procedure is still a pain point since it is usually conducted manually and costs great effort. The reason is that the experience of senior salespersons is hard to accumulate.

[1]https://www.webank.com/#/home

ACM ISBN 978-1-4503-8651-7/21/10.
https://doi.org/10.1145/3474085.3478555

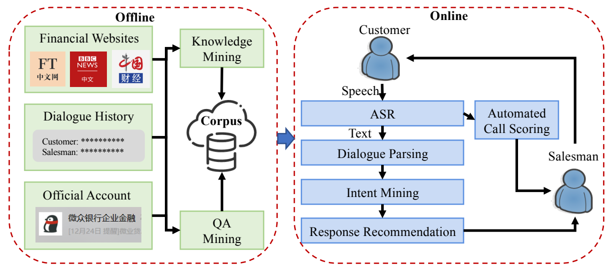


Figure 1: The workflow of the SmartSales

In this demonstration, we propose the first AI-powered telemarketing coaching system named *SmartSales* in the domain of Chinese FinTech. SmartSales aims to accumulate the experienced sales patter from customer-sales dialogues and then use this patter to coach the junior salespersons. As shown in Fig. 1, SmartSales comprises two components: offline knowledge accumulation and online coaching. The knowledge accumulation part aims to mine valuable knowledge as well as question-answer pairs, and then amalgamates them into the knowledge base. Meanwhile, the coaching part focuses on the real-time mining of the clients' intent and generating responses with the knowledge base as guidance for the salesmen. The success of SmartSales relies on the seamless integration of various natural language processing (NLP) techniques such as question-answer (QA) mining, automatic speech recognition (ASR), natural language understanding, response generation/recommendation, sentiment analysis, profanity-detection, and automated call scoring and the like. Through this demonstration, users will experience the working mechanisms of the AI coaching system and how the advanced NLP techniques can benefit telemarketing in the FinTech industry.

2 SYSTEM OVERVIEW

SmartSales includes various critical components, and in this section, we will briefly introduce their details.

QA Mining & KnowlegeBase Construction As shown in Fig. 2, automated QA mining includes two steps: question identification and answer detection. For the question identification step, a classifier is trained to distinguish meaningful questions with the non-question from the dialogues, and then the questions are further filtered to ensure smoothness and quality. Candidate answers are prepared for the selected answers, and we consider two types of similarity features to detect the matched one, including lexical (e.g., the ratio of overlapping words and TF-IDF) and semantic (e.g., sentence embedding and topic distribution [2]). Besides dialogue history, SmartSales also crawls knowledge from other sources such as financial websites and WeChat official accounts into its knowledge base.

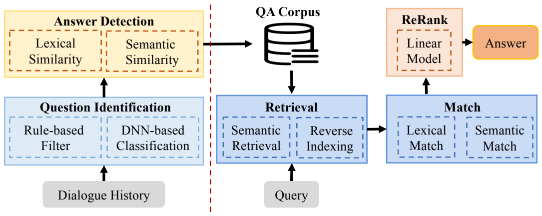


Figure 2: The QA Mining and Recommendation

ASR Given the real-time acoustic input, the ASR module decodes its corresponding transcripts as follows:

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} (\log P_{LM}(\mathbf{w}) + \lambda \log P_{AM}(\mathbf{a}|\mathbf{w})), \quad (1)$$

where P_{LM} is a language model (LM), P_{AM} is an acoustic model (AM), and λ is a trade-off parameter. In SmartSales, we build the AM using the Kaldi "Chain" model [6, 7]. To further improve the performance, we also use an advanced L2RS mechanism [9] integrated with BERT [3] to do the N-best rescoring step.

User Intent Mining We manually identify 35 intents which belong to 4 categories, and then the intent mining problem is transformed into a classification one. We extract word level and character level features using Word2Vec [5], then feed them into a convolution neural network (CNN) to get the high-level representation H^w and H^c, and finally train a classifier represented as:

$$y = Softmax([H^w; H^c]W^T), \quad (2)$$

where y is the intent and W^T is trainable weight matrices.

Response Recommendation As shown in Fig. 2, the response recommendation models employ a two-step approach, that is, retrieval and ranking. The essential words are extracted from the client's dialogue, as query q, to retrieve the top-n candidates, and then various features for candidates are obtained from pre-trained models. Eq. 3 is used to measure the similarity S_i of the i-th feature between query and candidates, and a pairwise learning-to-rank model [4] is used to rerank these candidates. The overall correlation is calculated by a feedforward neural network using Eq. 4, and the model is trained with the loss function defined as Eq. 5. Finally, the top 3 candidates are displayed to the end-users.

$$S_i(q,a) = Cos(q,a) = \frac{\|q \cdot a\|}{\|q\| \cdot \|a\|}, \quad (3)$$

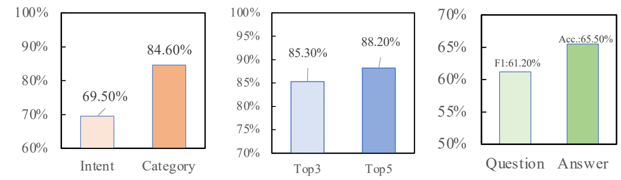
$$Sim(q,a) = \sigma(\sum_i W_i S_i(q,a)), \quad (4)$$

$$loss = \max(0, 1 - Sim(q,a^+) + Sim(q,a^-)), \quad (5)$$

where a^+ and a^- are two response candidates, and a^+ is more relevant than a^- with respect to query q.

Automated Call Evaluation Automated call evaluation uses algorithms to analyze phone conversations. Insights gained can help identify the best sales leaders and enhance their effectiveness by timely feedback. SmartSales supports profanity detection and positive/negative comment detection with a coarse-to-fine retrieval and matching approach. It also supports call quality scoring.

Performance Analysis We build a corpus constructed from the real dialogue between telesales and customers in the Chinese Fintech domain. The original data came from 3-months WeChat official account logs and speech data with around 15K recordings. We first obtained the transcripts by the ASR module and then recruited two workers to annotate the intent and question-answer pair label. After



(a) Intent Detection (b) Response Rec. (c) QA Mining

Figure 3: The System Performance of SmartSales

the human-annotated QA corpus had been built, we processed the data into clusters and tagged negative/positive labels for question pairs in the same cluster for the reranking task. Finally, we roughly obtained 10K of intent data, 5K QA pairs, and 15K ranking pairs. We split the dataset into 80% for the training set and reserved the rest as the testing set. The experimental evaluations in terms of accuracy and F1 are presented in Fig. 3.

3 SYSTEM DEMONSTRATION

The functionalities are demonstrated with a screenshot in Fig. 4.

Real-time ASR Decoding After the call is made, the real-time decoding result by an ASR system will be displayed on the homepage.

User Intent Mining & Response Recommendation The ASR decoding results will be forwarded to the user intent mining and response recommendation modules, and the analysis results will be displayed on the webpage.

Call Evaluation & Scoring Typical functionalities include the sentiment analysis results of both the clients and the telesales.

KnowledgeBase Maintenance The salesman can maintain a local knowledge base, which contains personalized question-answer pairs. This information will also be used to update the machine learning algorithms regularly.

User Management Different users have different management interfaces, depending on their role. The system administrators can create accounts for an institution and assign local administrators. The local administrators can access the user's page in the administration console to view, create, and modify users. Sales users can modify their user profiles.

Data Analysis Report This functionality aims to summarize the statistics of the salespersons and the quality of calls by these salespersons. The summary report will be shown to the administrator.

Figure 4: A Screenshot of SmartSales

Acknowledgement We thank anonymous reviewers for their helpful comments. The research of Raymond Chi-Wing Wong is supported by GZSTI16EG24.

REFERENCES

[1] Chris Baraniuk. 2015. Rise of the AI sports coach.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 4171–4186.

[4] Tie-Yan Liu. 2011. *Learning to rank for information retrieval.* Springer Science & Business Media.

[5] Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). http://arxiv.org/abs/1301.3781

[6] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, et al. 2011. The Kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*. IEEE Signal Processing Society.

[7] Daniel Povey, Vijayaditya Peddinti, Daniel Galvez, Pegah Ghahremani, Vimal Manohar, Xingyu Na, Yiming Wang, and Sanjeev Khudanpur. 2016. Purely sequence-trained neural networks for ASR based on lattice-free MMI.. In *Interspeech*. 2751–2755.

[8] Johannes Schneider. 2020. Human-to-AI Coach: Improving Human Inputs to AI Systems. In *Advances in Intelligent Data Analysis XVIII - 18th International Symposium on Intelligent Data Analysis, IDA 2020, Konstanz, Germany, April 27-29, 2020, Proceedings (Lecture Notes in Computer Science)*, Michael R. Berthold, Ad Feelders, and Georg Krempl (Eds.), Vol. 12080. Springer, 431–443. https://doi.org/10.1007/978-3-030-44584-3_34

[9] Yuanfeng Song, Di Jiang, Xiaoling Huang, Yawen Li, Qian Xu, Raymond Chi-Wing Wong, and Qiang Yang. 2020. GoldenRetriever: A Speech Recognition System Powered by Modern Information Retrieval. In *Proceedings of the 28th ACM International Conference on Multimedia*. 4500–4502.

[10] Jianbo Wang, Kai Qiu, Houwen Peng, Jianlong Fu, and Jianke Zhu. 2019. AI coach: Deep human pose estimation and analysis for personalized athletic training assistance. In *Proceedings of the 27th ACM International Conference on Multimedia*. 374–382.