SoMin.ai: Social Multimedia Influencer Discovery Marketplace

Aleksandr Farseev
ITMO Univerity
St. Petersburg, Russia
farseev@gmail.com

Kirill Lepikhin
Social Miners Research
Singapore
kirill@somin.ai

Hendrik Schwartz
Social Miners Research
Singapore
hendrik@somin.ai

Eu Khoon Ang
Social Miners
Singapore
eukhoon@somin.ai

Kenny Powar
Rebel Owl
Singapore
kpowar@rebel-owl.com

ABSTRACT

In this technical demonstration, we showcase the first ai-driven social multimedia influencer discovery marketplace, called SoMin [4]. The platform combines advanced data analytics and behavioral science to help marketers find, understand their audience and engage the most relevant social media micro-influencers at a large scale. SoMin harvests brand-specific life social multimedia streams in a specified market domain, followed by rich analytics and semantic-based influencer search. The Individual User Profiling models extrapolate the key personal characteristics of the brand audience, while the influencer retrieval engine reveals the semantically-matching social media influencers to the platform users. The influencers are matched in terms of both their-posted content and social media audiences, while the evaluation results demonstrate an excellent performance of the proposed recommender framework. By leveraging influencers at a large scale, marketers will be able to execute more effective marketing campaigns of higher trust and at a lower cost.

ACM Reference Format:
Aleksandr Farseev, Kirill Lepikhin, Hendrik Schwartz, Eu Khoon Ang, and Kenny Powar. 2018. SoMin.ai: Social Multimedia Influencer Discovery Marketplace. In *2018 ACM Multimedia Conference (MM '18), October 22–26, 2018, Seoul, Republic of Korea.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3240508.3241387

1 INTRODUCTION

The past decade has testified a rapid growth of the Internet. One can observe the drastic expansion of social networking services, where millions of users publish and consume information regularly. Built upon such growth, social media marketing industry has correspondingly developed its capabilities of helping marketers in content personalization and deliverance [11–13]. However, the growing amount of irrelevant content, such as unrelated advertisement and spam, made social media users more and more reluctant towards perceiving sponsored search results and online advertisement, such as “Google AdWords” [10] and “Facebook Sponsored Ads” [1].

MM '18, October 22–26, 2018, Seoul, Republic of Korea

ACM ISBN 978-1-4503-5665-7/18/10.
https://doi.org/10.1145/3240508.3241387

To mitigate such customer skepticism, marketers often leverage on human-centric content delivery channels, where Influencer Marketing clearly dominates over other marketing strategies. Indeed, it has been shown that 92% of consumers are more likely to trust brands that advertise via influencer channels [17] rather than those who has adopted conventional marketing strategies. Unfortunately, the limited availability of influencer search platforms and the absence of audience-based and content-based influencer matching technology, result in tremendous amounts of manual work performed, the corresponding high marketing agency service costs and low efficiency of the conducted marketing campaigns.

Aiming at bridging the aforementioned research and industrial gaps, in this technical demonstration we propose an online influencer discovery platform that would perform influencer matching at both content and audience levels simultaneously. To accomplish such semantic search, we have utilized our previously-proposed multi-source re-ranking approach [6], which is able to perform well-balanced and source-consistent recommendation over multiple data representations. The content representation was gained via extracting multi-modal hot topics, named entities, and image concepts [15] from recent influencer-posted and brand-intended content, while the audience representation was obtained via computing distributions of the brand and influencer social media followers over automatically-profiled behavioral user attributes [3, 7, 8]. The above technology was integrated into the cloud-based social multimedia influencer discovery marketplace SoMin, which delivers semantic-based influencer search to the corporate users.

The overall data processing pipeline in SoMin is as follows: (a) the distributed cloud crawlers harvest recently posted user-generated content (UGC) from multiple social networks with respect to a particular brand and its corresponding social media influencers [2]; (b) user profiles are predicted via SoMin's Social Multimedia Analytics API endpoints [15] for all brand's and influencers' followers, which allows for drawing brand-specific and influencer-specific bahavioural audience distributions; (c) multi-source multi-modal topic, named entity recognition, as well as image concept detection models [15] extract textual and visual content representations, which are used for drawing influencers' semantic content distributions as well as the semantic description of the brand-intended marketing message; and (d) the multi-modal re-ranking engine [6] recommends semantically-matching social media influencers with respect to the brand-intended marketing message and its audience. To the best of our knowledge, SoMin is the first ai-driven social

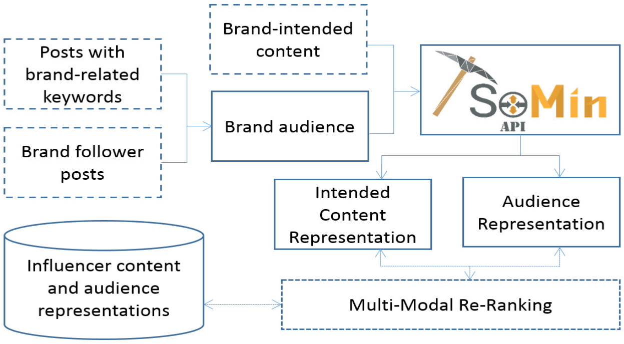

Figure 1: Block diagram of the SoMin platform

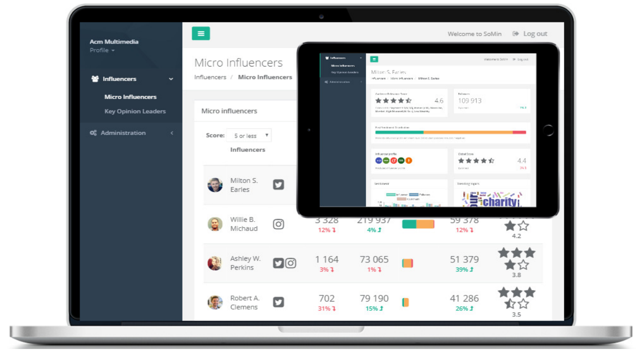

Figure 2: The SoMin online GUI snapshot

multimedia influencer discovery platform that performs social media influencer matching based on audience and marketing content simultaneously.

2 SOMIN

2.1 Platform Description

We now briefly describe the SoMin platform. The block diagram of the platform is shown in Figure 1. In the SoMin influencer discovery marketplace, the brand and influencer audiences are re-crawled and re-computed every 12 hours, while the influencer matching is executed in real time. At any given time frame, the following steps are taken (the continuously executing operations are marked with the "*"):

I. Brand and Influencer-Related Content Crawling*: The crawling process consists of two steps: (a) a **seed set** of brand and influencer-related followers as well as users who has recently mentioned Brand-related keywords is collected[1]; and (b) Twitter [16] and Instagram [14] APIs are used to perform content monitoring with respect to specified geographical regions (i.e. in our case, Singapore), so that cross-social network account mapping can be performed when users publish messages containing re-posts from other social networks (i.e. publish Instagram images on Twitter). By following the described user identification approach, we continuously harvest the data from over one million of Twitter, Instagram, YouTube, Pinterest, and Foursquare accounts across the region.

II. Audience and Content Representation*: Brand-intended and influencer-posted **content**, as well as the content posted by users from the seed set, are represented as follows: visual data (Instagram and Twitter images) is represented in a form of image concept distribution [15]; textual data (Twitter tweets, Instagram image captions, and Foursquare check-in comments) is represented in a form of Latent Topic and Named Entity distributions [15]. All the above distributions are further concatenated in one real-valued vector and normalized over all seed set users for the brand and each of its influencers independently. Brand and influencer **audiences** are represented by applying psycho-graphic user profiling (i.e. automatic detection of Psychological Demographics, MBTI Personality, Emotion Categories, Interests) [15] to each user in the system followed by averaging of real-valued attribute probabilities among seed set users for the brand and each of its influencers independently.

III. Semantic Brand-Influencer Matching: The audience and content representations are fed into Multi-Modal Re-Ranking engine [6] that produces the ordered lists of Influencers with respect to the given brand and the intended marketing content. For example, a query for influencer search in fashion domain may look like:

```
BRAND_FOLLOWERS: '@OlaySkin, @olay'
BRAND_KEYWORDS: 'olay, cleansing, hydration, moisture'
SEARCH_GEO_REGION: 'Singapore, Malaysia'
INTENDED_CONTENT: 'My morning routine includes my
multi-tasking Olay skin moisturizer with SPF available.'
INTENDED_VISUAL_CONTENT: 'https://bit.ly/2MBXayi'
```

The above query will result in the ranked list of fashion-related influencers in Singapore and Malaysia who's audience matches the audience of Olay Twitter[2] and Instagram[3] accounts while the recently-posted content is similar to the content that Olay intends to use for advertisement (e.g. @AIMEESONG, @OliviaPalermo, @xeniatchoumi, etc.).

3 THE DEMONSTRATION

Using our web GUI [5], SoMin users can specify multi-modal queries as described in the example from Section 2.1. The influencer matching response will be given in a form of a ranked list of top-matching influencers, where each influencer will be listed together with the brief description of his/her audience and the corresponding audience and content relevance score. By selecting influencer from the list, users will be routed to the "Influencer Details" dashboard that visualizes influencer's audience and recently-posted content in terms of Hot Topics, Named Entities, Behavioural Attributes, Image Concepts, and Sentiments of their audience. Finally, SoMin users will be able to specify various filters for tailoring Social Influencer matching engine for their specific needs. For example, the platform could be set up to output only those influencers who are active in Singapore, do not represent any brand at the moment, and did not advertise for competitive brands in the past.

[1] Assuming that some of the harvested users might be citizens of the European Union (EU), in order to preserve the compliance of our technical demonstration with General Data Protection Regulation (GDPR) [9], we've limited the data collection process in this demo to monitor Twitter and Instagram data via their public APIs [14, 16], while the posts from other social networks were obtained via harvesting cross-social network posting activities [7].

[2] http://twitter.com/olayskin

[3] http://www.instagram.com/olay

4 CASE STUDY IN ASIA

4.1 Experimental Setup

To further evaluate the performance of our influencer recommendation engine, we have conducted a case study on real-life data collected for one of the current platform users. The user is a PR Agency operating in Southeast Asia. The main purpose of using the platform was to understand the Customer Audience Segmentation and to find the key Social Media Influencers in Asia Pacific region for the brand operating in the restaurant industry (i.e. "Morton's The Steakhouse") and its three potential competitors (i.e. "Wolfgang's Steakhouse", "Lawry's The Prime Rib", and "Wolfgang Puck").

To extract targetable brand customer segments as well as to retrieve key relevant social media influencers, the platform was set up to monitor key popular keywords and hashtags (e.g. #mortons, #mortonssteakhouse, #mortonssingapore, #steak) related to the brand and its competitors in Twitter and Instagram. Where aplicable, the brand's and brand competitors' social media account followers were also included in the analytics pipeline. In total, SoMin.ai has analysed $315,251$ Social Media users who has collectively posted $13,627,016$ textual and $6,024,310$ image/video posts. The total number of discovered social media influencers is 870, which is about 0.3% of the overall population snapshot.

We have chosen top 20 messages posted by "Morton's The Steakhouse" on their Twitter timeline as a test query for the platform. Three independent annotators (professional marketers) were asked to annotate the recommendation results obtained for the 20 queries to the platform as "relevant" or "non-relevant"[4]. "Relevance" was defined as the similarity of the content in the query and the recent influencer-posted content as well as the potential match of the chosen Brand's customer segment (i.e. in our case, "Masculine, Mature, Average Income, Non-Logical, Principal, Idealists") to the Influencer's follower audience.

To gain an insight into the quality of SoMin Influencer Recommendation, we used "Precision at K" (P@K) metric, which is defined as the portion of relevant documents among the top K recommended documents. The reason behind the evaluation metric choice is its ability to give a good insight into the recommendation quality without prior knowledge on the total number of the relevant items in the whole dataset.

4.2 Evaluation

Evaluation was performed for different values of K. The average values among the 20 queries values of **P@K** are **0.87**, **0.9**, and **0.83** for **K** = **3**, **5**, and **10**, respectively. The results suggest that the Recommender System effectively solves the problem of Influencer Recommendation based on customer audience and intended marketing content simultaneously. We would also like to highlight that the highest performance was achieved for $K = 5$, while for the case of $K = 10$, the recommendation quality is comparably lower. Such an observation can be explained by the relevance of the content in the performed queries. Specifically, some query messages (e.g. bottles with liqueurs posted on the "Morton's The Steakhouse" Twitter wall) were not directly related to the brand's market domain (i.e. Steaks and Dining), which, in turn, might bring less relevant recommendation results at the tail of the recommendation list (e.g. Alcohol Drinks-related Social Media Influencers). Overall, we would like to highlight the high recommendation quality achieved for all the examined values of K, which allows for the successful use of SoMin platform for preparation and execution of the real-world influencer marketing campaigns.

[4]The inter-agreement between annotators was computed as 0.857 (Cohen's κ metric), which verifies a substantial agreement between annotators.

5 CONCLUSION

In this technical demonstration, we have presented the first ai-driven social multimedia influencer discovery marketplace, called SoMin [4]. The main aim of the platform is to help marketers find, understand their audience and engage the most relevant social media micro-influencers at a large scale. The experimental results demonstrate that SoMin is able to achieve the excellent recommendation performance when matching influencers in terms of both their posted content and follower audiences. The high recommendation performance, testified in the real-world settings, allow marketers to boost their digital efforts by conducting influencer marketing campaigns of higher trust at a lower cost.

6 ACKNOWLEDGEMENTS

This work was financially supported by Government of Russian Federation (Grant 08-08)

We gratefully acknowledge the assistance of Maestro Dr. Daron Benjamin Loo for assisting with preparation of the manuscript.

REFERENCES

[1] Facebook. 2018. Facebook ads. https://www.facebook.com/business/products/ads. Online; accessed 17 May 2018.
[2] Aleksandr Farseev, Mohammad Akbari, Ivan Samborskii, and Tat-Seng Chua. 2016. 360° user profiling: past, future, and applications. *ACM SIGWEB Newsletter* Summer (2016).
[3] Aleksandr Farseev and Tat-Seng Chua. 2017. Tweet can be Fit: Integrating Data from Wearable Sensors and Multiple Social Networks for Wellness Profile Learning. *ACM Transactions on Information Systems (TOIS)* (2017).
[4] Aleksandr Farseev and Lepikhin Kirill. [n. d.]. SoMin.ai: An AI-Driven Influencer Discovery Marketplace. https://somin.ai. Online; accessed 17 May 2018.
[5] Aleksandr Farseev and Lepikhin Kirill. 2018. SoMin: Social Multimedia Influencer Discovery Marketplace. https://demo.somin.ai. Online; accessed 1 May 2018.
[6] Aleksandr Farseev, Denis Kotkov, Alexander Semenov, Jari Veijalainen, and Tat-Seng Chua. 2015. Cross-Social Network Collaborative Recommendation. *Proceedings of the 7th ACM International Conference on Web Science* (2015).
[7] Aleksandr Farseev, Liqiang Nie, Mohammad Akbari, and Tat-Seng Chua. 2015. Harvesting multiple sources for user profile learning: a big data study. In *Proceedings of the 5th ACM International Conference on Multimedia Retrieval.*
[8] Aleksandr Farseev, Ivan Samborskii, and Tat-Seng Chua. 2016. bbridge: A big data platform for social multimedia analytics. In *Proceedings of the 2016 ACM on Multimedia Conference.* ACM, 759–761.
[9] EU GDPR. 2018. General Data Protection Regulation (GDPR) Portal. https://www.eugdpr.org/. Online; accessed 17 May 2018.
[10] Brad Geddes. 2014. *Advanced Google AdWords.* John Wiley & Sons.
[11] Francesco Gelli, Xiangnan He, Tao Chen, and Tat-Seng Chua. 2017. How Personality Affects Our Likes: Towards a Better Understanding of Actionable Images. In *Proceedings of the 2017 ACM on Multimedia Conference (MM '17).* ACM, 10.
[12] Francesco Gelli, Tiberio Uricchio, Marco Bertini, Alberto Del Bimbo, and Shih-Fu Chang. 2015. Image Popularity Prediction in Social Media Using Sentiment and Context Features. In *Proceedings of the 23rd ACM International Conference on Multimedia (MM '15).* ACM, 4. https://doi.org/10.1145/2733373.2806361
[13] Francesco Gelli, Tiberio Uricchio, Xiangnan He, Alberto Del Bimbo, and Tat-Seng Chua. 2018. Beyond the Product: Discovering Image Posts for Brands in Social Media. In *Proceedings of the 2018 ACM on Multimedia Conference (MM '18).* ACM.
[14] Instagram. 2018. Instagram Media API Endpoints. https://www.instagram.com/developer/. Online; accessed 13 Aug 2018.
[15] SoMin.ai. 2018. SoMin Social Multimedia Analytics API. http://dev.somin.ai. Online; accessed 17 May 2018.
[16] Twitter. 2018. Twitter Search API. https://developer.twitter.com/content/developer-twitter/en.html. Online; accessed 13 Aug 2018.
[17] Steven Woods. 2016. # Sponsored: The Emergence of Influencer Marketing. (2016).