Study on application of open source intelligence from social media in the military

Zh Y Ju1,2, Q Li2, H Y Liu1, X M Cui2 and Zh H Wang1

1School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
2Beijing Institute of Special Electromechanical Technology, Beijing 100012, China

Email: haiyangliu@bjtu.edu.cn

Abstract. With the gradual popularization of information technology and network technology, various types of open source information, especially social media information, have shown an explosive growth trend. Social media intelligence not only possesses ideological and value-level values, but also directly serves military defense. It will help stimulate open source intelligence to play a huge role in combat operations that combining communications and psychological knowledge to explore and exploit frontier technologies in artificial intelligence, such as machine learning and data mining, and conducting research on social media big data collection, analysis, mining, and prediction technologies. This paper focuses on the preliminary analysis of the application of open source information such as social media in the field of military intelligence.

1. Introduction

Sherman Kent, the “father of intelligence analysis” in the United States, pointed out in 1947 that about 80% of the intelligence used by the intelligence community every day comes from public sources. In his 1949 monograph “Strategic Intelligence for American World Policy”, he pointed out that although some intelligence may be obtained through secret channels, a large amount of intelligence must come from bland and fair observation and research [1]. In December 1982, the United States Secretary of Defense stated that No. 5240.1-R “Procedures governing the activities of DOD intelligence components that affect United States persons” considers the use of publicly available intelligence as one of the types of intelligence used by intelligence agencies [2]. Leaders in the IC (U. S. Intelligence Community) recognized that the challenges and dynamics of the 21st century would bring more demand for OSINT (Open Source INTelligence), not less. FBIS (Federal Broadcasting Information Service) Deputy Director J. Niles Riddel, at the First International Symposium on Open Source in 1992, acknowledged changes in OSINT resulting from the rise of personal computing, large-capacity digital storage, capable search engines, and broadband communication networks. He believed all these factors would lead to exponential growth in the commercialization of information [3]. At the same event, then CIA (Central Intelligence Agency) Deputy Director Admiral William Studeman called for “a revolutionary change in the Intelligence Community’s approach to opensource management, collection, processing and dissemination.” [4] [5] In 2012, the U.S. Army has made OSINT operations a necessary component of Army intelligence operations, “OSTNT is integral to Army intelligence operations” [6]. The importance of OSINT is self-evident.
The use of open source information will not only supplement high-value intelligence information, but will also redefine the concept of forewarning and become the basis for future intelligence analysis. The risks and costs associated with intelligence obtained through a comprehensive analysis of publicly available information are relatively small, and this is where the value of OSINT staff lies. In fact, it is not open source information analysis that supports secret information. On the contrary, secret information is often used to support open source information analysis. Doing so may be beneficial in obtaining the necessary warnings in the current and future ambiguous operational environment [7].

1.1. Development of U.S. Open Source Intelligence

The earliest open source intelligence research began with a defense-oriented agency. On February 26, 1941, the United States established the Foreign Broadcast Monitoring Service (FBMS) to monitor and analyze the propaganda programs of the Axis powers. On July 26, 1942, the Federal Broadcasting Supervision Office was renamed the Federal Broadcast Information Service (FBIS), which was almost closed at the end of World War II, and was taken over by the War Department on January 1, 1946. One year later, under the National Security Act of 1947, it was transferred to the CIA. Twenty years later, the CIA’s official history description of FBIS described its fundamental organization and responsibilities … basic operation and methods” as largely unchanged.

From the creation of FBIS until the 1990s, the purview of open-source analysis within the IC was primarily monitoring and translating foreign-press sources. There are some important differences between this historic character of OSINT—the first generation of OSINT—and today’s second generation. Collection of material was a significant emphasis of the first-generation OSINT effort. FBIS operated 20 worldwide bureaus to allow it to physically collect material for exploitation. Embassies also provided a platform for collection of material. In addition to diplomatic officers, defense attachés served as overt collectors. The Defense Attaché System was consolidated in 1964-1965 under the authorities of the Defense Intelligence Agency (DIA). Some mission functions were reduced in order to focus on high-priority material, but the remaining requirement to process this material was primarily translation; although, it should be noted that FBIS has served some analytic functions—primarily trend analysis—since its founding.

FBIS’s work provided critical insights and decision points for the military during the Cold War, including the first indications of the Soviet removal of missiles from Cuba, early warning of the Soviet withdrawal from Afghanistan, and context on crises in Hungary and Czechoslovakia. Eighty percent of the information used to monitor the collapse of the Soviet Union has been attributed to open sources. The end of the Cold War resulted in budget cuts for most IC institutions, but it created a particular crisis for FBIS and the open-source mission. At the same time the volume of OSIF (Open Source InFormation) was increasing dramatically, FBIS was rapidly losing resources. In 1997, FBIS was at risk of dissolution as part of CIA budget cuts but was saved by a public campaign led by the Federation of American Scientists.

A working group at the Office of the Under Secretary of Defense for Intelligence (OUSD-I) in 2004 found deficiencies in open-source policy and doctrine, training, and management. Furthermore, U.S. Department of Defense (DoD) open-source requirements were underrepresented and underfunded within the IC.18 These findings prompted the establishment of the Defense Open Source Council (DOSC), which was appointed as the primary government mechanism for DoD OSINT through DoD Instruction 3115.12.

On November 1, 2005, the Director of National Intelligence (DNI) created the Open Source Center (OSC) and designated the CIA as its executive agent, later redesignating it as the “functional manager.” Based at the CIA and replacing FBIS, the OSC brand gave OSINT practitioners greater license to expand beyond news monitoring and translation. The creation of the OSC fulfilled a requirement of the Intelligence Reform and Terrorism Prevention Act of 2004, which specifically called for the creation of an intelligence center dedicated to “the collection, analysis, production, and dissemination of open-source intelligence.” The DNI is charged with ensuring that OSIF and analysis is effectively and efficiently used by the IC.
In October 2015, the OSC was renamed the Open Source Enterprise (OSE), and it was brought under a newly created Directorate for Digital Innovation (DDI) within the CIA. The inclusion of the OSE in a directorate focused on cyber threats and digital technology could enhance the institution’s embrace of technology and analytic tools [5].

1.2. The status of open source intelligence in joint operational doctrines

The joint operational doctrine is a code of conduct, operating standard, collaborative process, and control system for the five branches of the U.S. Army, Navy, Air Force, Marine, and Coast Guard operations. It is a regulatory document promulgated by the US Army Federation. The U.S. military has compiled a total of 109 joint operations doctrines, divided into three levels, namely general outlines, professional classification guidance, and professional doctrines. The regulations are divided into 6 majors, namely J1 personnel, J2 intelligence, J3 operations, J4 logistics, J5 operations plans, J6 command, control, communication and computer systems [8].

The first level (general outline), including JP1 “Joint Operations of the United States Armed Forces” and JP0-2 “Joint Operations of the Armed Forces”, mainly determine the principles, guidelines, command system, and division of responsibilities of joint operations.

The second level (6 professional guides) including: JP1-0 “Personnel and administrative support doctrine for joint operations”: Determine joint personnel and administrative support duties and work processes for joint operations; JP2-0 “Joint Combat Intelligence Support Joint Doctrine”: Determine the principles and joint responsibilities of joint operational intelligence sharing, describe the organizational structure of joint intelligence, and guidelines for intelligence work; JP3-0 “Joint Combat Doctrine”: Determine the basic principles of joint operations and guidelines for the formulation of action plans, elaborates matters to be considered in the implementation of joint operations in wartime, establishes principles for non-war military operations, and elaborates considerations for avoiding operations. JP4-0 “Joint Combat Logistics Support Regulations”: Determine the authority and responsibilities of joint operations logistics units, list the logistic work list, clarify the process of formulating the logistics plan, and the process and system of theater-level logistics establishment. JP5-0 “Ordinances for the Formulation of Joint Operations Plans”: Determine the procedures and principles for the formulation of joint operations plans, including how to formulate plans, how to prepare execution plans, planned backup plans and withdrawal plans, and determine the review standards for joint operations plans. JP6-0 “Doctrine for C4 (Command, Control, Communication, and Computer Systems) System Support in Joint Operations”: Determine the tasks, goals, and work lists of the C4I (Command, Control, Communication, Computer Systems and Intelligence) system, determine the evaluation standards for the quality of the C4I system, determine the structure and infrastructure of the C4I system, the operating procedures and responsibilities of each specialty, as well as control and assessment methods.

The third level (92 Joint operations doctrines), correspond to the 6 professional guidance doctrines on the second level.

J1 has 3 doctrines, which mainly describe joint personnel and administrative issues, including reporting structure, religious support, and financial management.

J2 has 6 doctrines, which mainly describe joint intelligence support, combat intelligence support, target-specific intelligence support, combat space intelligence preparation, and joint operations national intelligence support. JP2-0 “Joint Intelligence”; JP2-01 “Joint and National Intelligence Support to Military Operations”; JP2-01.3 “Joint Intelligence Preparation of the Operational Environment”; JP2-03, “Geospatial Intelligence Support to Joint Operations”. Two other intelligence doctrines were not made public.

J3 has 60 doctrines, which mainly include four aspects of joint operations, joint operations support, joint operations defense, and joint non-war operations. Including:

There are 25 doctrines on joint operations, including amphibious operations; ground operations; joint interception operations; joint maritime operations; carrier helicopter operations; joint special operations; joint fire support; Joint tactics, technology and procedures for close air support; joint
nuclear operations; joint tactics, technology, and procedures for space operations; joint airborne and air assault operations; joint command and control; etc.

There are 19 doctrines on joint operations support, including joint information warfare; joint command and control warfare; joint engineering operations; joint electronic warfare; joint airspace control of combat zones; joint psychological warfare; military deceit operations and so on. Among them, JP3-13.2 “Military Information Support Operation” belongs to the joint information warfare part.

There are 9 doctrines on joint operations defense, including joint anti-aviation and anti-missile operations; joint tactics, techniques and procedures to suppress enemy air defense; joint theater missile defense; joint air defense operations; joint rear area operations; nuclear and biological defense and so on.

There are 7 doctrines on joint non-war operations, including counter-terrorism; peacekeeping operations; anti-drug operations; non-combat evacuation operations; humanitarian assistance; joint tactics, technology and procedures for domestic support operations and so on.

J4 has 19 doctrines, which mainly include joint transportation and logistics support issues. J5 has 3 doctrines, which mainly include joint tactics, techniques, and procedures for developing a battle plan; and the implementation system of joint operations plan. J6 has 1 doctrine, which mainly describes the use of C4I systems in campaign and tactical.

Through the above structural analysis, it can be seen that in the six professional fields, intelligence work ranks second only to the personnel major in the US military joint doctrine. The volume is second only to the combat and logistics majors, and in the combat major, intelligence support has a place in joint support. Open source intelligence, in JP2-0, constitutes an intelligence discipline together with geospatial intelligence, human intelligence, signal intelligence, surveying and signature intelligence, technical intelligence, and counterintelligence.

2. Research status at home and abroad

In recent years, social media has become one of the most representative data sources for big data. Overseas, according to YouTube blog and media homepage data, as of May 2019, the number of monthly active viewers has exceeded 2 billion, mobile users have more than 200 million daily visits, and users upload more than 500 hours of new videos to the platform per minute on average. According to official statistics from Instagram, as of June 2018, Instagram has 1 billion active users per month, and the daily active users of the Stories function have exceeded 500 million. In less than 8 years, more than 50 billion photos were uploaded. Domestically, according to Tencent’s third quarter 2019 financial report, the combined monthly active accounts of WeChat and WeChat were 1.151 billion. According to the 2018 WeChat data report, WeChat has an average of 45 billion daily messages and 410 million audio and video calls. According to Sina Weibo’s third quarter 2019 financial report, there were 497 million monthly active users and 216 million daily active users. The average daily posting volume of Weibo video/live broadcast was 1.5 million, the average daily posting volume of photos was 120 million, and the text posting volume was 130 million.

Multiple information warfare capabilities will enhance the user’s influence. Among them, using social media as a link between information operations and network operations is a very important technical means. With the popularity of intelligent terminals, data information is growing at an exponential rate, the characteristics of the information contained in it also show new changes.

2.1. Research status abroad

The Defense Advanced Research Project Agency (DARPA, Defense Advanced, Research Project Agency) launched the TIA (Total / Terrorism Information Awareness) [9] program in 2002, which aims to use computer technology to analyze and process social public opinion; In 2012, DARPA launched the “Social Media in Strategic Communication” (SMISC) [10] project to help the U.S. military better understand the hot events that occur in real time on social media deployed by the U.S. military and implement large-scale social networks Media campaign. The US military needs SMISC to be able to quickly identify new rumors or hot events on social networks, and then determine who is
posting this information. The SMISC system can distinguish whether information on the social network that is not conducive to the US military is accidental or purposeful behavior of hostile countries and hostile groups.

In 2006, the US Department of Homeland Security organized the development of a super data analysis and processing system—ADVISE [11]. The system can analyze a large amount of information from online blogs and emails and government archives and intelligence records to find suspicious signs or regular conclusions. Coincidentally, the Japan Institute of Information Analysis developed the WISDOM system, indexed about 100 million Japanese web pages, and was able to give detailed trend statistics and analysis results.

In 2010, the U.S. intelligence service cooperated with Google to conduct real-time monitoring of the Internet, using publicly available information on the Internet to predict the trend of sensitive events. The Army has awarded BAE Systems a $4.37 million five-year task order for OSINT support, the company announced Oct.15 during the 2019 Association of the United States Army conference. The contract ranges from providing OSINT solutions to training Army personnel to how to treat open source data, whether it’s social media, news or other publicly available information. BAE systems will take the tools it’s developed from pathfinder program, prior contracts with agencies such as DARPA as well as internal innovations to provide the Army with the tools they need. BAE Systems will also help the Army determine the appropriate policies and governance structures to apply to OSINT as a separate data type. As part of its support, the company will also establish and manage a secure cloud hosting environment [12].

2.2. Research status at home
In order to meet the major needs of the national and social public security fields, the State Key Laboratory of Complex System Management and Control of the Chinese Academy of Sciences has developed a real-time intelligence collection and analysis system for terrorist activities, the Dark Web [13]. Through systematic research and development, the system can conduct terrorist organization behavior monitoring and early warning based on open source information [14]. Bu XY from Wuhan University conducts doctoral research on the subject of “counter-terrorism intelligence acquisition based on open source information integration” [15]. Fu JL and others used text mining and network analysis methods to analyze the main characteristics of “East Turkistan” activities from OSINT [16].

The National Development and Reform Commission funded a large-scale information search research project for massive network information; The National 863 Program has funded research on key technologies for cyber information security and cyber public opinion analysis and early warning, and the construction of a demonstration verification system; The National Natural Science Foundation of China has funded research on new theories and methods for web analysis and mining of web text public opinion; The National Social Science Foundation of China has funded research on early warning systems for emergencies. In the summer of 2017, the Chinese government released the “New Generation Artificial Intelligence Development Plan”, seeking to develop artificial intelligence applications in the field of defense, such as remote sensing image processing, making command decisions, solving military redundancy, and studying defense equipment, etc.

The massive amount of open source information on the Internet makes information with real intelligence value likely to be overwhelmed by invalid information. Therefore, effective, timely, and intelligent intelligence analysis methods are required to meet the processing requirements of OSINT in a big data environment. The intelligence profession is facing the challenge of big data. The key issue is how to exploit the analytical potential of open source data, rather than confidential information that is closely guarded. Open source information will provide early warning of enemy threat intentions in future threat environments [7].

As for the basic technology in the field of artificial intelligence that generates OSINT, many researches have been carried out at home and abroad. Many countries have raised the development of artificial intelligence to national strategies, and strongly supported them from the aspects of policy orientation, strategic planning, and capital budgeting. In recent years, technologies such as
computational vision, natural language processing, and robotics have developed by leaps and bounds. Governments have begun to deploy artificial intelligence into the military and defense fields. DARPA has launched a large number of basic technology research projects, exploring and developing related technologies for independently acquiring, processing information, extracting key features, and mining association relationships from different types of multi-source data such as text, images, sound, and video. In early 2017, the US Department of Defense established the Algorithmic Warfare Cross-Functional Team (AWCFT) to “accelerate the ability of the Department of Defense to integrate the use of big data and deep learning”. For example, Project Maven has used deep learning-based computer vision algorithms to achieve automatic processing of images and videos. US researchers have successfully identified an East Asian country surface-to-air missile base based on deep learning. The Russian military has been actively promoting the development of intelligent robotics, Russian President Putin recently announced that “who can become the leader in the field of AI, he will become the ruler of the world.”

3. From open source data to OSINT

3.1. Definition of OSINT

The research work of OSINT started in the early 1940s, and its connotation has been continuously expanded in research.

According to [17], open source intelligence (OSINT) is the intelligence discipline that pertains to intelligence produced from publicly available information that is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence and information requirement [17]. OSINT also applies to the intelligence produced by that discipline.

According to [18], OSINT namely relevant information derived from the systematic collection, processing, and analysis of publicly available information in response to known or anticipated intelligence requirements.

According to [19], OSINT is intelligence that is produced from publicly available information and is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific IR. Publicly available information is information that anyone can lawfully obtain by request, purchase, or observation.

OSINT is developed using media and Web-based sources. OSINT processing transforms (converts, translates, and formats) text, graphics, sound, and motion video in response to user requirements. OSINT is also developed from information collected by commercial companies that use their own assets or purchase information from independent contractors who monitor media [19].

Luo Qingchang, a former secretary of Li Kenong, pointed out in “Introduction to Information Science” that OSINT refers to “information collected and used from publicly available materials or news in order to meet the needs of intelligence work” [20].

3.2. Categories of OSINT

U.S. military divides OSINT used in military operations into three categories: strategic intelligence, campaign intelligence, and tactical intelligence. Strategic intelligence includes 9 types: political intelligence, economic intelligence, social intelligence, transportation intelligence, telecommunications intelligence, military geography, armed forces intelligence, personal (biographical) intelligence, and scientific and technological intelligence; campaign intelligence refers to geographic resources, regional situation, social factors, religious composition, logistical support, etc. in the theater's jurisdiction; tactical intelligence refers to information needed to plan tactical operations, including geospatial information, meteorological conditions, public psychology, and civil foundation facilities, transportation networks, imaging systems, etc. [21].

OSINT products are categorized by intended use and purpose. Categories can overlap and some publicly available categories of intelligence products. As [17], intelligence products can be categorized
as following: Indication and warning, current intelligence, general military intelligence, target intelligence, scientific and technical intelligence, counterintelligence, and estimative [22].

OSINT can provide the following functions for military operations: (1) it can provide situational awareness of the combat environment; (2) it can obtain information about threat characteristics, terrain, weather, and civilians; (3) it can build the required intelligence knowledge system; (4) it can provide basic knowledge and understanding of potential threat actions or intentions in a specific combat environment; (5) it can generate intelligence knowledge as the basis for the integrated functions of the military, such as battlefield intelligence preparation.

3.3. Production of OSINT
The big data environment formed by social media has brought opportunities and challenges to the development of OSINT work. The IC leaders saw that open source was becoming an unwieldy intelligence discipline. Admiral Studeman described other collection disciplines as “highly structured” but declared that “open source is not a tightly integrated discipline” and that “open source information collectors, processors, and users have been diverse and decentralized groups spread across the breadth and depth of the Community.” The IC did not have knowledge of its own unclassified holdings and capabilities, and it had no means for sharing OSIF [5].

Big data research states: “Using big data is better than using complex models” [23]. Big data reflects the relevant information of people, events or activities from different perspectives. Fusion of these data together, the use of advanced technology in the field of artificial intelligence for related analysis can more fully reveal the connection of things, dig new models and relationships, and generate OSINT for the military.

Hua [24] and others believe that in addition to the original process, the intelligence analysis process in the big data environment places more emphasis on information collection and analysis processing, which contains a series of processes including definition of intelligence requirements, formulation of intelligence plans, information retrieval and data collection, multi-source information fusion and cleaning, information analysis and content mining, information display and intelligence extraction, report writing and information transmission. Specifically, the basic process is as follows:

First, identify the task type. Determine the mission theme of the intelligence, analyze the situation of the intelligence mission, capture the characteristics of the intelligence users, and then translate the intelligence needs into intelligence requirements and clearly define them.

Second, develop an intelligence plan. After identifying the requirements, according to the requirements, determine the process, construct the index system, plan the intelligence time, form the intelligence team, choose the appropriate research method, and choose the corresponding technology and tools. According to the intelligence mission, determine the source channel, scope, scale, and type of information retrieval and data collection, then formulate collection strategies and implement collection, and evaluate the scale, effectiveness, and authenticity of the collected data.

Third, data preprocessing. Integrate data from multiple sources and different structures, filter duplicate data, identify duplicate names and alias data, split and extract data, and perform a series of operations such as leak detection and filling, dimension reduction and so on.

Fourth, data mining analysis. Analyze and mine the preprocessed data to form an intelligence report with decision support or reference value, and pass accurate information in an appropriate manner at the right time.

Big data reflects the relevant information of people, events or activities from different perspectives. Fusion of these data together for related analysis can more fully reveal the connection of things, dig new models and relationships, and further generate open source intelligence to serve the battlefield.

4. Application of OSINT from Social Media in the Military
Information technology can be used in conflicts that may occur in the future, and a variety of information warfare capabilities will enhance user influence, including computer network warfare, electronic warfare, psychological warfare, and social media weaponization. Among them, a very
important technical means is to use social media as a link between information and cyber warfare. Social media, as an important open source information channel for people to disseminate information and express their views, contains a wealth of useful information. In recent years, it has become one of the most representative data sources for big data. Social media can help army commanders gain intelligence, accurately understand the situation on the battlefield, and change the way they fight. The army can understand the real situation by analyzing the information on social media. Commanders examine specific words to determine if the message is trying to use persuasive text or deliberately create tension to have a direct impact on the battlefield. There are many reasons for the sudden change in the position or behavior of the enemy. If the army can determine the cause of the enemy's change based on the specific entity or specific time that people are talking about, they can take accurate action [25].

4.1. **Guide military operations**

In the summer of 2014, “Pro-Russian separatists” began activities in eastern Ukraine. Russia has repeatedly denied that its regular troops entered Ukrainian territory, but the facts reflected on social media indicate that the Russian government has been lying. Young Russian soldiers posted selfies on Instagram, and the photos quickly flowed to Instagram and Twitter, as well as many YouTube videos. They did not know that the metadata containing the photo attributes positioned their photo location within the Ukrainian border.

In May 2015, a member of the Islamic State (IS) posted a selfie on a social network. At the time of the release, he was careful not to reveal details such as the IS headquarters building. Although the photo did not directly reveal the building's latitude and longitude information, according to the analysis of the photo’s “metadata” by the US Intelligence Group 361, combined with the previous big data, the building was quickly located, and the air force fired three missiles to destroy it. From posting a post on the social network to 3 missiles hitting the target, the entire process took less than 24 hours.

In March 2018, a key threat project of the Institute for the Study of War and the American Enterprise Institute released a research report, “Russia's Military Posture Ground Forces Order of Battle” [26], which researched and combed open source information about the Russian ground forces' planned structure and deployment. Russia is in the process of reorganizing some of its military forces. The project confirmed the deployment and location of the Russian conventional army through OSINT, and wrote this battle sequence report.

United States Air Force intelligence officials said their intelligence collection efforts must analyze large amounts of Internet data, including social media, to closely track enemy movements. It will be the most likely situation that the US military will encounter in the future that the combination of conventional and non-conventional operations such as the Russian-Ukrainian conflict, mixed with a strong public relations conflict style, and this situation is most difficult to cope with [7]. The prediction clues generated by open source information such as social media cannot constitute the final intelligence product by itself, but it can be used to guide traditional intelligence surveillance and reconnaissance forces to investigate and verify the beginning of open source information.

The U.S. military believes that major changes should be made in the way military intelligence professionals think about issues. Public information is as valuable as restricted data. The new generation of intelligence agents who grew up in the ubiquitous information technology and social media environment, they are more likely to break through the constraints of traditional thinking. Meanwhile, when training in the use of intelligence gathering tools, innovation should be encouraged in order to better collect and analyze open source information. Specific intelligence analysis training should include: restatement of the problem, causal flow chart, weighted permutation, “devil” propaganda and many other techniques, such as “14 powerful techniques for solving problems” written by Morgan Jones in The Thinker Toolkit [7].

4.2. **Launch a cyber psychological warfare**

Social media has become a battleground for psychological warfare.
In March 2018, the British company Cambridge Analytica was exposed to leaked personal information of Facebook’s 50 million users to manipulate multi-country elections including the 2016 US elections. According to the “National Review” report, during the 2016 US presidential election, the company started from Facebook’s 270,000 users, using nearly 185 friends each as a diffusion path, predicting individual behavior based on public data points and models, and pushed personalized political ads to 50 million users. This is not only a new cyber security incident, but also a new challenge to the ethics and law of intelligence agencies [27].

“Islamic State” and its predecessors borrowed from the jihadist organization to shape “online attacks” into a new situation of “jihad”, using social media to carry out a large number of activities, including extremist propaganda, global recruitment, launching intelligence wars and psychological warfare. They played social media at the limit to serve its conventional wars, lone wolf terrorism, and information warfare. While “Islamic State” uses social media to transmit and carry out terrorist information, counter-terrorism forces are also collecting intelligence based on social media and carrying out precise attacks on terrorism accordingly. Through data mining technology and analysis of suspicious people's social media, potential terrorist members can be identified, targeted, and struck [28].

Personalized recommendations keep everyone from knowing what others see, and under the manipulation of the algorithm, everyone can only see what the operator wants them to see. In this closed-loop information feeding process, the minds of target users can be deeply affected [29]. Mainstream social media has a huge user base, comprehensive data accumulated over the years, powerful data mining technology, and active third-party applications on social media platforms, which making it easy to become a fertile ground for precise placement and manipulation. Social media can form portraits of interests, showing the interests of individuals and small groups. For small-scale groups, the interest graph can be used to predict user behavior; for large-scale groups, it can be used to analyze socialization trends.

It will have a profound impact on mastering the initiative of psychological warfare that using basic technology in the field of artificial intelligence to study open source information generated by social media and producing OSINT. In terms of intelligence guidance, research “protects data and systems” through existing capabilities and emerging methods”, instills uncertainty or misleading information into local, and confuses and deceives opponents before or during cyber attacks.

4.3. Military involved public opinion monitoring
On November 22, 2017, the RYB kindergarten in Chaoyang District of Beijing was exposed to child abuse, and there were rumors about military-involved content, saying that “the family members of the head of the ‘Tiger Group’”and “there are soldiers involved in child care indecent incident”and so on. On the 23rd, the military immediately involved in the investigation, confirming that the lover of the kindergarten principal was a “Tiger Team” cadre who had changed jobs in early 2016. The kindergarten land is not for the military. No officer or relative of the military is involved in the kindergarten or other management work, nor are the children of the military officer or soldier attending the kindergarten. On the evening of the 24th, the “Tiger Regiment” political commissar responded positively to this hot issues through the China Military, clarifying that “it is circulated online that the kindergarten principal is a member of the active military family, which is inconsistent with the facts” and “the officers and soldiers were not found to be involved in so-called indecent acts” [30]. In this case, military involved public opinion was controlled in a timely and effective manner, without continuous fermentation and deterioration.

The maliciously created military involved public opinion will damage the image of the military, affect military-civilian relations, and even affect major military operations. In the era of big data, in order to implement military involved public opinion monitoring, two basic tasks need to be focused on: (a) Network-based open intelligence collection and management. Utilize web crawlers and deep web mining technology, based on APIs published by social media such as Internet forums, Weibo, and WeChat, and combine with ontology knowledge base, text analysis, semantic web, and multimedia
analysis to collect and monitor sensitive military involved information on the Internet. (b) Intelligent analysis and processing of intelligence. Filter and analyze military involved information collected and monitored, mine hot topics with seditious nature from massive data, conduct risk assessment of possible social public opinion, and provide effective decision-making and network public opinion situation control suggestions based on the spread characteristics of hot news.

In addition, in the analysis and processing of social media open source intelligence, attention should be paid to strengthening human-computer interaction and virtual-real interaction. Machines can help during the analysis, but they cannot replace everything. Machines can better simulate human intents and thus predict human behavior more accurately, but there are still significant differences between imitated intentions and true interpretations of original intentions. Therefore, while vigorously promoting the automation of intelligence processing, we must introduce human knowledge for effective supervision, strengthen human-machine interaction and virtual-real interaction, and reduce possible machine bias and excessive dependence on data [7].

5. Conclusion
With the gradual popularization of information technology and network technology, various types of open source information, especially social media information, have shown an explosive growth trend. It is necessary and important that attaches great importance to the role of OSINT in the military intelligence system, and adopts a military-civilian integration approach to exploit OSINT. It will have a disruptive impact on intelligence architecture and help stimulate OSINT to play a huge role in combat operations that combining communications and psychological knowledge to explore and exploit frontier technologies in artificial intelligence, such as machine learning and data mining, and conducting research on social media big data collection, analysis, mining, and prediction technologies.

This paper introduces the development of open source intelligence in the United States, then analyzes the status of open source intelligence in the joint operational doctrines of the US military, and summarizes the basic technologies for generating open source intelligence through social media in recent years at home and abroad. The process of producing open source intelligence from social media data in the era of big data is explained, and a preliminary analysis is focused on the application of open source information such as social media in the field of military intelligence.

Acknowledgements
This work is supported by National Natural Science Foundation of China (No. 61672086, 61771058).

References
[1] Kent S. Strategic Intelligence for American World Policy Princeton: Princeton University Press 1949
[2] 1982 Procedures governing the activities of DOD intelligence components that affect United States persons
[3] Riddel, JN. 1992 Remarks at the First International Symposium, ‘National Security and National Competitiveness: Open Source Solutions,’ Open Source Solutions Inc. December 2
[4] Studeman and Admiral William 1992 Teaching the Giant to Dance: Contradictions and Opportunities in Open Source Within the Intelligence Community Open Source Solutions Inc. December
[5] Williams H. J. and Blum I. 2018 Defining second generation open source intelligence (OSINT) for the defense enterprise Rand Corporation
[6] ADP2-0. 2019 Intelligence. Headquarters, Department of the Army.
[7] Zhang C. 2018 Military Digest. 17 pp 59-63
[8] Wen XM. Introduction to the US Joint Operations System Doctrine. 2015-9-17.
[9] Total / Terrorism Information Awareness (TIA)
[10] Social Media in Strategic Communication
[11] Mark Clayton. 2006 US plans massive data sweep The Christian Science Monitor
11

[12] Strout N. BAE wins open source intelligence contract from the army. C4ISRNET.

[13] “Dark Web” Prototype System. Center for Social Computing and Parallel Management Chinese Academy of Sciences 2005-2009.

[14] Wang FY, Li XC, and Mao WJ et al. 2013 Social Computing Methods and Applications. Zhejiang University Press

[15] Bu XY. 2013 Research on the Counterterrorism Intelligence Acquisition Based on Open Source Information Integration. Wuhan University

[16] Fu JL, Liu WL, and Zheng XL et al. 2014 Acta automatica sinica 40(11) pp 2456-68

[17] Headquarters, 2004 Department of the Army. Intelligence, The Army Field Manuals FM 2-0

[18] Joint Publication 1-02. 2016 Department of Defense Dictionary of Military and Associated Terms

[19] Joint Publication 2-01 2017 Joint and National Intelligence Support to Military Operations

[20] Luo QC. 2007 Introduction to Information Science. Beijing: Current Affairs Press

[21] Department of Defense Intelligence Bureau Open Source Intelligence Center. 1996 Department of Defense Open Source Intelligence Program

[22] Headquarters, 2012 Department of the Army. Open-source Intelligence. ATP2-22.9

[23] Halevy, A. Norvig, P. and Pereira F. 2009 The unreasonable effectiveness of data. Intelligent Systems 24(2) pp 8-12

[24] Hua BL. 2014 Intelligence Analysis in the Big Data Era in the Context of PRISM. Library and Information 5 pp 2-6

[25] Zhang HY. 2018 US media analysis: The significance and challenges of applying social media to the battlefield. Security Insider

[26] Harris C, and Kagan FW. 2018 Russia's Military Posture Ground Forces Order of Battle. Institute for the Study of War

[27] Dougherty MB. 2018 Cambridge Analytica and Social Media Panic. National Review

[28] Li HW. 2018 Sociology of Religions. 5 pp 224-42

[29] Jiangnan Times. Facebook is far from China, but social media that controls your mind is close to you

[30] Xiao JB and Li L. “Tiger Team” political commissar Feng Junfeng responds to rumors of military involvement in the RYB Kindergarten incident.