Who Did What: Editor Role Identification in Wikipedia

Diyi Yang[♣], Aaron Halfaker[♢], Robert Kraut[♠], Eduard Hovy[♣]
♣ Language Technologies Institute, Carnegie Mellon University {diyi,hovy}@cmu.edu
♢ Wikimedia Foundation ahalfaker@wikimedia.org
♠ Human-Computer Interaction Institute, Carnegie Mellon University robert.kraut@cmu.edu

Abstract

Understanding the social roles played by contributors to online communities can facilitate the process of task routing. In this work, we develop new techniques to find roles in Wikipedia based on editors' low-level edit types and investigate how work contributed by people from different roles affect the article quality. To do this, we first built machine-learning models to automatically identify the edit categories associated with edits. We then applied a graphical model analogous to Latent Dirichlet Allocation to uncover the latent roles in editors' edit histories. Applying this technique revealed eight different roles editors play. Finally, we validated how our identified roles collaborate to improve the quality of articles. The results demonstrate that editors carrying on different roles contribute differently in terms of edit categories and articles in different quality stages need different types of editors. Implications for editor role identification and the validation of role contribution are discussed.

Introduction

Distributed work teams in online communities have become increasingly important in creating innovative products, such as GNU, Linux and Wikipedia. Millions of volunteers participate in the online production communities, exchange their expertise and ideas, and collaborate to produce complex artifacts. Better understanding of the participants and how they behave can make these communities more successful. For example, in Wikipedia, editors take up different responsibilities, when editing articles, based on their interest and expertise. Some, for example, might add substantive new content to articles while others may focus on copy-editing. Systems designed to route work to appropriate Wikipedia editors have focused on matching editors to articles that are topically similar to ones they have already worked on (Cosley et al. 2007). These task recommenders, however, have for the most part ignored the type of work that the editors can do.

This paper develops new methods to identify roles that editors exhibit when contributing to Wikipedia and then tests whether work done by editors occupying different roles affects article quality. This knowledge can then be used to create more sophisticated task recommender systems that take both article content and editing skill into account.

The problem of identifying editors' roles in Wikipedia has attracted significant attention. Numerous studies have discussed how to identify roles based on users' behavioral regularities and social network signatures (Welser et al. 2007). Most research classifies editors based either on their edits in different namespaces (Welser et al. 2011) or via the user attributes such as access privileges (Arazy et al. 2015), personalized barnstars (Kriplean, Beschastnikh, and McDonald 2008), etc. Classification based on users' attributes is relatively accurate, but this information is not available for many active editors and is insufficient in explaining the nature of an editor's work. While classification based on edit histories can be constructed for most active editors, current approaches focus on simple edit counts and access privileges fail to provide a finer grained description of the work actually performed in an edit. For example, it cannot tell the difference between an editor who copy-edits or rephrases a paragraph and an editor who inserts markup, template or information to an article.

In this work, we extend Daxenberger's (Daxenberger and Gurevych 2012) fine grained taxonomy of edit types to differentiate editors who occupy different editing roles. In our taxonomy, edits are distinguished contextually in terms of the object being edited (e.g. information, template, reference, etc.) and functionally, in terms of the edit operation (e.g. insert, delete, modify, etc.). Specifically, we developed 24 edit categories to understand how different users perform the editing task collaboratively. We then described the development and validation of methods for the automated measurement of these edits categories revealed in users' edits.

Building on this automated measurement of edit types, we use a graphical model analogous to LDA topic modeling analysis to identify the latent roles editors occupy, much as documents comprise topics. Just as documents are mixtures of topics, editors are mixtures of roles. The roles that editors occupy generate the edits they perform, just as the topics that comprise a document determine the works in it. In contrast to studies that employed either clustering analysis or principle component analysis to extract user roles (Liu and Ram 2009; 2011), our role modeling treats an editor as comprising multiple roles at the same time. This approach makes the role more interpretable in capturing the versatil-

Dataset	# Revisions	# Editors	# Article	Anonymous User Included	Time Period
Annotated Edit Category Corpus	953	728	-	YES	2014.06.10 - 2015.06.10
Editor Modeling Revision Corpus	626,761	38,520	172,740	NO	2014.12.01 - 2014.12.31
Article Quality Prediction Dataset	-	22,633	151,452	NO	2015.01.01 - 2015.06.30

Table 1: Dataset Description

ity and dynamics of editors.

The collaborative contribution and interaction behaviors of such roles matters a lot in shaping the health and sustainability of Wikipedia. As a further step, we investigated how the collaboration of editors carrying on different roles predicted the quality changes of articles and some differences in the number of requisite roles for improving the quality of articles. The results demonstrated that different sets of roles are needed in the different quality stages of article. In detail, articles in Start or Stub[1] stages require more Substantive Expert to help with the content; articles in A or Good stages show a lack of Wikipedia Gnomes[2] to repair the broken links and make things run more smoothly.

To sum up, this work lays a foundation for future research to automatically identify a fine granularity edit types for Wikipedia editors, to extract a mixture of editor roles and to encourage specific role setting to improve the quality of articles. It also helps in how to develop intelligent task routing systems to recommend users to tasks that match their expertise.

Related Work

A role is a bundle of tasks, norms and the behaviors that are expected of those who occupy a position in a social structure (Biddle 2013). Roles are major mechanisms through which project members, including volunteers in large online communities, coordinate complex activities. Theory on coordination in groups and organizations emphasized role differentiation, division of labor and formal and informal management (Kittur and Kraut 2010).

Previous social roles studies in online communities can be understood through the content of interaction and through the use of behavioral and structure cues (Kittur and Kraut 2008; Welser et al. 2011). For example, a variety of roles have been identified in online discussion forums (Fisher, Smith, and Welser 2006; Yang, Wen, and Rose 2015; Welser et al. 2007), including *answer people*, *questioners*, *leaders*, etc. Another similar line of work studies the identification of roles in the context of a social network (Bamman, O'Connor, and Smith 2013), e.g. *celebrity, newbie, lurker, troll*, etc.

In the context of Wikipedia, Welser et al. (2011) used both qualitative and quantitative methods to identify four roles in this online community: substantive experts, technical editors, vandal fighters, and social networkers. In contrast, Arazy et al. (2015) utilized the access privileges in Wikipedia and developed a set of twelve roles based on Wikipedia's organizational structure. Kriplean et al. (2008) showed that informal awards can be used to encourage and reward different types of valued work, and suggest that these Barnstars might be a good way to identify emerging types of works and different roles in Wikipedia. However, such role discoveries based on superficial edit types, structural signatures or access privileges suffer from either weak ability in differentiating editors or not readily accessible profile information. They are also inadequate in capturing what is actually edited and how editors collaborate in the construction process (Qin, Greene, and Cunningham 2014; Liu and Ram 2009).

Existing studies on capturing the intentions behind a textual change (Faigley and Witte 1981) suggest that edit types that each editor contributes to an article can also be considered to uncover the expected and enacted behaviors of an editor (Liu and Ram 2009). For example, Daxenberger et al. (2012) automatically assigned edit categories such as grammar, paraphrase or vandalism to edits in a document. Their taxonomy of edit categories (Daxenberger and Gurevych 2013; Pfeil, Zaphiris, and Ang 2006) is acquired through the differentiation and operationalization of surface edits and text based edits. However, relatively little research except (Liu and Ram 2009) has gone into how such edit categories define and interpret specific roles in their coordinative contribution to editing articles.

Researchers have developed a number of techniques for identifying social roles online, generally employing either clustering analysis or principle component analysis. For example, Welser et al. (2011) grouped editors based on the types and content of their edits, as well as their user pages. Liu and Ram (2009) utilized a K-Means approach to classify contributors based on their actions in editing article pages. However, relatively little research has discussed the multifaceted property of a user, namely, one can perform multiple social roles simultaneously. Graphical models used in uncovering the hidden topics in a document (Blei, Ng, and Jordan 2003) can be leveraged here to acquire a mixture of user role representation, which are quite reasonable in profiling the versatility and dynamics of editors.

Our research also extends earlier research on role modeling by introducing evaluation criteria. Although earlier attempt to deduce the roles structure in Wikipedia have generated roles with face validity that are loosely consistent with expert's classifications, they provide no metrics to evaluate the quality of the roles. In the current paper we validate the methods we used by (a) estimating the percentage of the variance in low-level editing behavior the roles account for and (b) examining whether roles are useful for predicting changes in the quality of articles.

[1] https://en.wikipedia.org/wiki/Template:Grading_scheme

[2] https://en.wikipedia.org/wiki/Wikipedia:WikiGnome

Research Question and Data

Our major research goal is to find a set of social roles associated with editors in Wikipedia based on our developed taxonomies of edit categories. Then we plan to investigate how these roles and their collaborative participation affect the quality and coordination of users' contribution. Our analysis is conducted on three datasets from English edition of Wikipedian, as shown in Table 1. Specifically, we will train a multi-class classifier to assign edit types for edits inside a revision on the Annotated Edit Category Corpus. Then apply the learnt model to the Editor Modeling Revision Corpus and identify editors' repeating patterns of activity. The Article Quality Prediction dataset is used to investigate how the collaboration of editor roles affects the changes of article quality.

Predicting Edit Categories

Previous research to identify editors' roles in Wikipedia based these assessments primarily used edit counts in different namespaces, structure signatures (Welser et al. 2011) and access privileges (Kriplean, Beschastnikh, and McDonald 2008), without making assumptions about the type of work that a particular edit entailed. To address the inadequacy, we first introduce a fine-grained taxonomy of the types of edits editors make to Wikipedia articles (i.e., pages in Wikipedia namespace 0). We then describe new a machine-learning model to automatically identify the semantic edit categories (e.g., adding new information versus vandalizing and article) associated with each edit. These classifiers map low-level features of the edits, including the number of added or removed tokens, misspelling words, and comment length to a multi-label classification, representing the edit categories which an edit belongs. We then use this classification of edit types as well as other information about the type of work editors do in Wikipedia as input into our role classifier. The development and validation of this machine-learning model are described in more detail as below.

Edit Categories Construction

Basing our research on Daxenberger et al. (2012), we distinguished between revisions and edits. A Revision is created whenever an editor makes changes to a Wikipedia page. An Edit is a coherent local change and regarded as one single editing action. Each edit is associated with a set of labeling of edit categories, representing in which aspects it has been changed. A revision can contain multiple edits. For each pair of adjacent revisions, we collected a set of edits that has been made to transform from its parent revision into this revision.

Figure 1 provides an overview of our edit taxonomy, on the basis of these studies (Daxenberger and Gurevych 2012; 2013). In this work, we annotated a set of edits rather than revisions. In general, this taxonomy considers actions (insert, delete, modify) applied to different objects in Wikipedia (e.g., information, templates or references), leading to 24 distinct edit types. The two top-level layers summarize whether these edit categories are meaning-preserving or meaning-changing.

Of the meaning-preserving edits, *Grammar* (G) means the edit is correcting spelling or grammatical errors, as well as fixing punctuation. When an edit attempts to paraphrase words or sentences, it is categorized as *Rephrase* (P); if such edit only moves entire lines without changes, it is defined as *Relocation* (R). For edits that try to operate with the markup segments, such as "===History===", depending how it affects the markup, we divide them into three subcategories, *Markup Insertion* (M-I), *Markup Deletion* (M-D) and *Markup Modification* (M-M).

Meaning-Changing edits depends upon how an edit affects the textual information content, we generated three categories: *Information Insertion* (I-I), *Information Deletion* (I-D), and *Information Modification* (I-M). Similarly, we acquired the remaining categories *Template Insertion* (T-I), *Template Deletion* (T-D), and *Template Modification* (T-M), *File Insertion* (F-I), *File Deletion* (F-D), *File Modification* (F-M), *External Link Insertion* (E-I), *External Link Deletion* (E-D), *External Link Modification* (E-M), *Reference Insertion* (R-I), *Reference Deletion* (R-D), *Reference Modification* (R-M), *Wikilink Insertion* (W-I), *Wikilink Deletion* (W-D), and *Wikilink Modification* (W-M).

Our taxonomy breaks Daxenberger's 'Reference' category (Daxenberger and Gurevych 2012) into three finer-grained categories: External Link refers to links from articles to web pages outside Wikipedia, Wikilink refer to links to another page within the English Wikipedia and Reference describes the source of the information, to help the reader who wishes to verify it, or to pursue it in greater depth[3]. Note that we utilized the Revision Scoring package [4] to identify Relocation, and did not include the category of relocation into our prediction stage.

Feature Space Design

The Annotated Edit Category Corpus contains 1997 edits. We annotated it based on a written annotation guideline. The annotation task is framed as a multi-label classification. That is, each edit will be assigned to one or more edit categories. For example, if an edit added a sentence to an article, this edit might involve insertion of information only or the insertion of information, a Wikilink insertion and a reference simultaneously. An edit containing the three components would be multi-labeled as I-I, W-I and R-I. To assess the validity of the annotation, we compared the annotations of 63 randomly sampled revision edits made by the first author and by an expert Wikipedian. Despite the difference in Wikipedia editing experience between the hand coders, the agreement between the annotations was substantial (Cohen's Kappa = 0.723; see (Landis and Koch 1977) for rules of thumb for evaluating strength of agreement using Kappa).

The machine learning goal was to classify an edit into one or more of the edit categories based on characteristics of the text changed, the comments editors used to describe their edits, and characteristics of the edit. To capture these char-

[3] https://en.wikipedia.org/wiki/Help:Referencing_for_beginners

[4] http://pythonhosted.org/revscoring/index.html

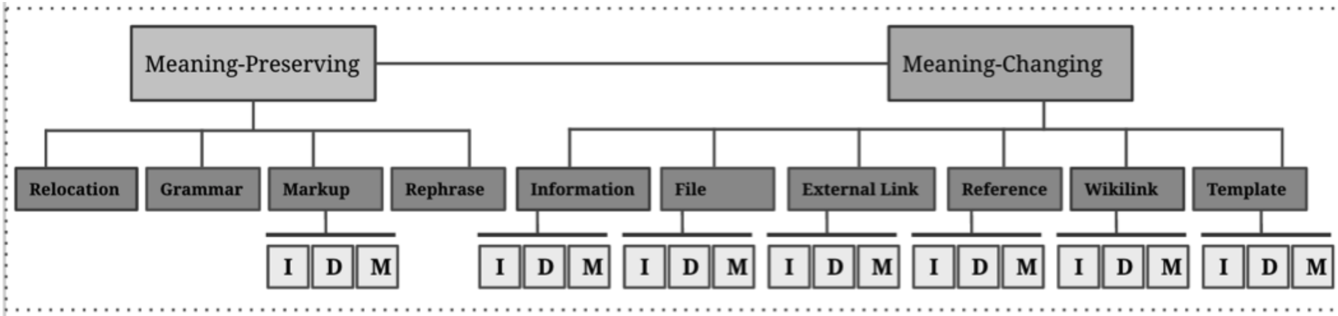

Figure 1: **The Taxonomy of Edit Categories**. Note: Insertion is abbreviated as I, Deletion as D and Modification as M

acteristics, we developed the following features[5]:

Is minor whether the revision is marked as minor change.

Comment length the number of characters in the revision comment.

Typo mention : whether the comment contains 'typo' or 'grammar'.

Is user registered : author is registered or is IP user.

Number of edits : the number of edits in this revision.

Number of tokens/capitals/digits/whitespace : the number of tokens/capitals/digits/whitespace in a segment.

Types of POS tag : the number of distinct POS tags.

Semantic similarities : the maximum, minimum and average semantic similarities between segments within an edit.

Misspelling words : the number of misspelling words in the segment.

Operation type : the number of INSERT/DELETE operations.

Segment length : the length of INSERT/DELETE segments.

Operation in template : whether the edit happens in the segment context of template such as '{{}}'.

Operation in file : an edit happens in the segment context of file such as '[File /Image/ Media:]'.

Operation in markup : an edit happens in a markup[6] segment context, such as '===', '==', '<div>', '</div>', '<span>', etc.

Operation in reference : an edit happens in a reference[7] segment context '<ref>', '</ref>'.

Operation in external link : an edit is performed in the segment context of external link[8] 'http://' or 'https://', etc.

Operation in wikilink (internal) link : an edit happens in an internal link[9] context such as '[[', ']]'.

Template/markup/reference/file/external/wikilink in segments: the number of designed markers related to template, markup, reference, file, external, wikilink that are contained in the segment.

Given the input feature representation of an edit, we then built a machine-learning model for this multi-label classification. Specifically, we used two of the multi-label classifier implemented in Mulan (Tsoumakas, Katakis, and Vlahavas 2010) using ten fold cross validation. We used the RAkEL ensemble method classifier, described in (Tsoumakas and Vlahavas 2007). It randomly chooses a small subset with k categories from the overall set of categories. We compared this with the MLkNN classifier, which is based on K Nearest Neighbor method. Table 2 shows the evaluation metrics including Recall, Precision, micro-averaged F1 score and AUC (Area under Curve). Both methods gave classifications that agreed with the human judgments, indicated by the AUC score of 0.865 and 0.906 respectively. We chose to use RAkEL method in order to acquire a relatively better performance in terms of F1 Score.

	Recall	**Precision**	**F1**	**AUC**
RAkEL	0.575	0.730	**0.643**	0.865
MLkNN	0.363	0.724	0.482	0.906

Table 2: Edit Categories Prediction Results

Modeling Editor Roles

Our edit taxonomy and its automated measurement only describe the types of work that an editor does when writing or revising the article pages that the general public associates with the encyclopedia. However, in addition to what Kittur and colleagues call this "direct production work" (i.e., edits to articles) (Kittur et al. 2007; Kittur, Pendleton, and Kraut 2009), Wikipedia requires a lot of behind-the-scene

[5]Here, Operation represents the action (Insert or delete) of an edit. Segment means the textual content that has been operated by a user. Segment Context is a piece of article content where the Segment is situated in (we collect the Segment content together with around100 characters before and after its content).

[6]https://en.wikipedia.org/wiki/Help:Wiki_markup

[7]https://en.wikipedia.org/wiki/Template:Ref

[8]https://en.wikipedia.org/wiki/Help:Wiki_markup#Externallinks

[9]https://en.wikipedia.org/wiki/Help:Interwiki_linking

Derived Roles	Representative Behavior
Social Networker	Main talk namespace, user namespace, reference modification
Fact Checker	Information deletion, wikilink deletion, reference deletion, file deletion, markup deletion external link deletion
Substantive Expert	Information insertion, wikilink insertion, markup insertion, reference insertion, external link insertion, file insertion, template insertion
Copy Editor	Grammar, paraphrase, relocation
Wiki Gnomes	Wikilink modification, Template insertion, markup modification, Wikipedia talk namespace, category namesapce
Vandal Fighter	Reverting, user talk namespace, reference insertion, external link deletion, paraphrase
Fact Updater	Template modification, reference modification, file namespace
Wikipedian	Wikilink insertion, Wikipedia namespace, template namespace, file insertion

Table 3: Derived Editor Roles and Their Representative Edit Types

administrative and coordination work to be successful, and what might be termed the indirect work has been increasing as a percentage of all work done in Wikipedia (Kittur et al. 2007). To a first approximation, one can identify indirect work by the namespace in which it is done. For example, discussion of changes to articles is typically done in namespace 1 (article talk pages), discussion and changes to Wikipedia policies are done in the Wikipedia talk and Wikipedia name spaces (5 and 4 respectively), and much editor-to-editor communication occurs in the user talk namespace (namespace 3). To allow our role models to represent indirect work, such as social interaction, community support, and maintaining standards in our role models, we included the number of edits editors made in each Wikipedia namespace[10] into the role models.

We also include the number of reverts (i.e., returning a Wikipedia to a prior state) and vandalistic edits editors made in the role model. Unlike (Daxenberger and Gurevych 2012), we did not create new classifiers to infer these edit types from editing activity. Rather we take advantage of two utilizes written by the Wikimedia Foundation that accurately measure this activity. Mediawiki-utilities Revert Check API [11] measures revert. The Vandalism API [12] returns the probability that a given revision is vandalism; we considered revisions with a vandalism probability scores larger than 0.85 to be vandalism. Reverts and vandalism was assigned to each of the edits comprising a single revision (i.e., all the edits done between consecutive saves to a Wikipedia page).

Role Identification Method

Our objective is to identity the roles that editors play, clustering editors who share patterns of work, using the types of edit they make in articles, their revert and vandalism, and edit counts in other namespaces. For this purpose, we used the graphic model underlying the Latent Dirichlet Allocation (LDA) method. LDA is an unsupervised, statistical generative model that can be used to discover hidden topics in documents as well as the words associated with each topic (Blei, Ng, and Jordan 2003). It assumes that each document is generated as a mixture of latent topics and each topic is characterized by a multinomial distribution over words. In the context of Wikipedia, an editor, represented by his or her edit history of edit types, reverts, vandalistic edit and edits in non-article namespaces, is analogous to a 'document'. The number of edits of different edit types, reverts, vandalistic edits and edits in non-article namespaces, is analogous to word frequency within the editor "document". The latent roles derived are analogous to an LDA topic.

Here, roles are based on repeating patterns of activities or 'structural signatures' and are analyzed in action, based on the work itself. Roles that editors occupy generate the edits they perform; editors occupying the same roles have similar patterns of work. Unlike the use of the term role in sociology, our definition did not include expectations from role partners (Orlikowski 2000) because in Wikipedia informal roles do not include strong expectations. Just as in an LDA topic model, where each document comprises multiple topics and each word can appear in multiple topics, an editor in Wikipedia comprises a mixture of roles, which may vary from one article to another, from one namespace to another or even within a single article. This approach is more realistic than previous ones that assumed that each editor occupies only a single role at a time and renders our extracted social roles more interpretable when describing editors' versatility and dynamics.

Derived Roles Exploration and Validation

We trained a LDA model on the Editor Modeling Revision Corpus. We experimented with driving from 5 to 15 roles (i.e., topics in the LDA software) and evaluated the interpretability of the produced latent roles based on human judgment. Qualitatively, we first visualized the top ranked edit types for each role, and then authors interpreted the results based on whether such work types are coherent in explaining the given roles. We ended up with 8 roles and selected the edit-types and namespaces that are most likely to correspond to a role. We summarized the results in Table 3. Two experts familiar with Wikipedia applied a label to each topic, based on the behaviors most heavily associated with each role. Detailed discussion of these roles identified via the LDA method is presented as below.

Social Networker. These editors make frequent edits in

[10] https://en.wikipedia.org/wiki/Wikipedia:Namespace

[11] https://pythonhosted.org/mediawiki-utilities/lib/reverts.html#mw.lib.reverts.api.check

[12] http://ores.wmflabs.org/scores/enwiki/?models=reverted&revids=revision_id

Wikipedia's communication spaces and their profile page but rarely edit articles. As demonstrated in Table 3, social networkers utilized 'Main Talk' and 'User' namespaces extensively. Instead of contributing to articles, social networkers tend to discuss article content and build profiles that show their interests and group membership.

Fact Checker. The most defining characteristics of these editors are the removal of content. Fact Checkers have extensive activities related to information, markup, and wikilink deletion, etc. While this may seem counter-productive on the surface, removing unnecessary content is part of Wikipedia's fact-checking process.

Substantive Expert. Substantive expert contributes by adding substantive content to articles, including providing references to increase the reliability of articles and inserting new knowledge to articles etc. They are the main content producers, engaging in many types of creations, and perform actions more frequently than average contributors.

Copy Editor. Editors who make contribution to improve the format, punctuation, style and accuracy of articles are referred as copy editors. They copy edit Wikipedia articles to make them clear, correct and concise mainly through checking grammar, paraphrasing and adjusting sentences to proper positions.

Wiki Gnome. "Wiki Gnome" is a term used by Wikipedians to refer to uncontroversial, but productive work. These editors make smaller contributions that tend to be focused towards making the content in Wikipedia cleaner by fixing issues with markup and easier to find by fixing and disambiguating links. These editors mainly work on Template insertion, Wikilink modification and Markup modification.

Vandal Fighter. These editors are the gatekeepers of Wikipedia. They monitor the feed of edits as they saved, identify vandalism and revert it (Reverting) and also post warnings to editors who vandalize Wikipedia (User Talk namespace).

Fact Updater. This group of editors contributes mainly to the template content of articles (e.g. Infoboxes – Boxes containing statistics and quick facts that appear on the right-hand side of most Wikipedia articles). Since Wikipedia covers topics that change over time, a lot of work needs to be done to keep these articles up to date. For example, when a company's CEO changes or when a popular band releases a new album.

Wikipedian. Editors in this group contribute to a diverse namespaces such as file, template, draft, etc., some of who might belong to the administrators. These editors work in spaces that are seldom seen by readers to keep the hidden order ordered (Viégas, Wattenberg, and McKeon 2007). Although Wikipedian have limited activities in editing articles, they invest a lot of time to help organize and standardize Wikipedia.

Our findings of Substantive Expert, Vandal Fighter and Social Networker roles are consistent with the roles discovered by (Welser et al. 2011) and Wikipedian role is similar to the Quality Assurance role defined in (Arazy et al. 2015). However, the difference is that our Copy Editor, Wiki Gnome, Fact Checker, Fact Updater roles are obtained through a fine-grained analysis of editors' edits types, which are not directly reflected by simple edit counts in different namespaces.

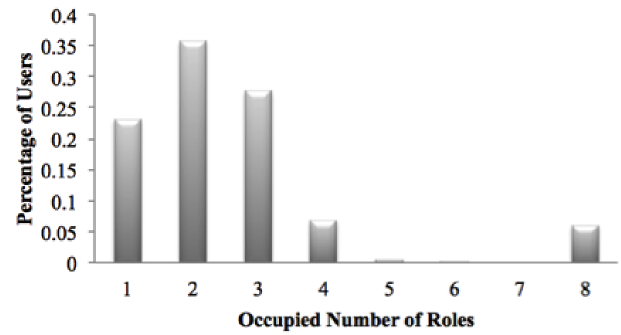

Figure 2: Distribution of Occupied Number of Roles.

We also represented how mixed editors are by computing a Gini coefficient based on how many roles an editor has occupied. A user is considered as occupying a role if he/she has a probability higher than $\frac{1}{8}$ (0.125). The Gini coefficient is 0.3, indicating that editors do occupy different number of roles. This is consistent when we visualized how much percentage of editors occupies a certain number of roles, as shown in Figure 2.

To evaluate the validity of our identified editor roles our methods identified, we estimated the percentage of variance across editors in the number edits of each edit type the roles accounted for. This metric is analogous to communalities in a factor analysis or principal components analysis. In this regression model, the input is an eight dimensional vector indicating how likely the editor belongs to each role and the output indicates how many edits an editor contributes to a specific edit category. We built 24 regression models to predict edit counts in each individual edit category from editors' role distribution. The average R-squared score for these models weighted by the frequency of the predicted behavior is 0.562, indicating that editor roles can explain over 56% of the variability in the numbers of edits of a certain type that an editor makes. The editor roles were especially successful in predicting grammar edits (81% of variance explained), modifying templates (76%), insertion of Wiki links (73%), and additions (62%) and deletions (52%) of information. Roles were poor in explaining insertion, modification and deletion of files, external link deletion and paraphrasing (all with less than 9% of variance explained).

Improving Article Quality

The quality of Wikipedia articles varies widely. Although there are over 4.5 million articles in the English Wikipedia, as of September, 2014 Wikipedians have evaluated fewer than 0.1% of them as good articles or better and over 88% of them as start or stub class articles (the two lowest quality categories). Collaboration among editors with different skills is essential to developing high quality articles (Kittur and Kraut 2008). This section of the paper attempts to determine how contribution by editors occupying different roles at distinct times in an article's history influence changes in its quality. Doing so will allow us to better understand the causes of quality variance in Wikipedia (De la Calzada and Dekhtyar 2010) and will demonstrate the utility of our role .

To do so, we first measured the how much contribution made by a specific role to an article page during a certain time period. Then we explored the correlations between the coordination of editor roles and article quality, controlling for the number of editors, the total number of edits, etc. This analysis is conducted on Article Quality Prediction Dataset. Identifying roles entailed first applying our multi-label classification model of edit categories to categorize the work done during this work and then using LDA techniques to derive the roles from the edit categories performed by each of the editors.

Model Design

We modeled editor roles during the month of Dec 2014 and change in article quality in the first half of 2015 so that the data for modeling roles did not overlap with the data for computing changes in article quality. We measured the contribution of each role in the following six months by summing up all the work of editors who take up that role. Since each editor is a mixture of roles, we attributed the contribution of different roles to an edit in proportion to the probability that that the editor belonged to a specific role. For example, consider editor A who belongs with 80% probability to the Copy Editor role, with 10% to Social Networker and with 10% probability to Vandal Fighter. In this case, we consider one of A's edit consists of 0.8 edits contributions by the copy editor role, 0.1 edits by the social networker role and 0.1 by the vandal fighter role.

Dependent Variable

Article Quality Changes: We validated how our extracted roles and their collaborative interaction contribute to article qualities by framing it as an article quality prediction task. Past work exploring the dynamics of article quality in Wikipedia used assessments applied by Wikipedia editors to articles (Hu et al. 2007; Lipka and Stein 2010). However, these assessments are rarely updated and therefore are often out of sync with the quality level of the article at any given time. To get around this problem, we opted for a different strategy. Researchers have developed robust machine learning strategies for predicting the quality level of an article that do not suffer from such staleness. There are many models to choose from in the literature (e.g. Anderka et al. (2012) quality flaw model and Likpa et al. (2010), which used writing styles to identify featured articles). However, we chose to use the model developed by Warncke-Wang et al. (2013) because it focuses exclusively on current features of the article itself as opposed to the history of activity on the article. This model is currently used by Wikipedia editors and updated by members of the Wikimedia Foundation Staff to measure article quality and identify articles with stale assessment tags[13]. This model classifies articles into the Wikipedia's article assessment scale based on article length, number of headings, number of references, completeness (Warncke-Wang, Cosley, and Riedl 2013), etc. This classifier is highly accurate, with a mean agreement with classification made by Wikipedia editors of 0.609.

Consistent with past work (Kittur and Kraut 2008), we measured article quality using this classifier at two time points six months apart, Jan 1, 2015 and July 1, 2015 (denoted as previous quality score and end quality score respectively). In order to measure sub-class changes in quality we applied a simple weighted sum-based aggregation to the article quality scores such that Stub (the lowest class) was assigned a score of zero and Feature Article (the highest class) was assigned a score of 5 and multiplied the probabilities returned by the classifier by each score and summed the result. With this strategy, if 100% of the probability were centered on Stub, we would arrive at a score of zero. If 100% of the probability were centered on Featured Article, we arrived at a score of five. We calculated change in article quality by subtracting the previous quality score from the end quality score. Spot-checking by comparing changes scores with an examination of the two versions of the article revealed that even small increases in the change score represented clear improvements in the coverage and quality of the article, while decreases represented vandalism and other types of damage.

Control Variables

Previous Quality Score: This is the article quality score in the beginning of Jan 2015. We controlled this variable to validate how role coordination affects the article quality in different stages of an article.

Article Registered Edits: the total number of edits contributed by registered editors (not IP users) to an article page during the six-month time period.

Article Registered Editors: the number of unique registered editors involved in the past six months. Wikipedia is easy to edit does not mean that editors carrying different roles contribute with the same intensity or are needed in the same way.

Talk Registered Edits: This is the total number of edits contributed by registered editors to the article talk pages.

Article Bytes Changed: This variable summed the added (removed) bytes to an article page that increase (decrease) its length. Then we calculated the length increment by subtracting the removed bytes from the added bytes.

Independent Variables

Contribution of Social Networker (Social Networker): We summed all the edits contributed by editors who take up the social network role in the past six month, dividing by the total number of edits in this article.

Similarly, we obtained other seven dependent variables, including Contribution of Fact Checker, Copy Editor, Substantive Expert, Vandal Fighter, Fact Updater and Contribution of Wikipedian.

[13] `https://meta.wikimedia.org/wiki/Research:Screening_WikiProject_Medicine_articles_for_quality`

Variables	**Model 1**		**Model 2**		**Model 3**	
	Coef.	SE	Coef.	SE	Coef.	SE
Previous Quality Score	-.183***	.001	-.188***	.001	-.140***	.003
Article Registered Edits	.129***	.000	.125***	.000	.128***	.000
Article Registered Editors	-.046***	.000	-.045***	.000	-.045***	.000
Talk Registered Edits	-.031***	.000	-.030***	.000	-.030***	.000
Article Bytes Changed	.409***	.000	.407***	.000	.407***	.000
Social Networker			.015***	.006	.023***	.014
Fact Checker			-.009***	.005	-.026***	.013
Substantive Expert			.058***	.005	.017***	.013
Copy Editor			.013***	.003	.029***	.009
Wiki Gnomes			-.033***	.005	-.073***	.012
Vandal Fighter			.008***	.006	.009	.014
Fact Updater			.006*	.005	.012*	.012
Wikipedian			.013***	.005	.047***	.012
Previous Quality Score × Social Networker					-.008	.005
Previous Quality Score × Fact Checker					.021**	.005
Previous Quality Score × Substantive Expert					-.139***	.005
Previous Quality Score × Copy Editor					-.017*	.005
Previous Quality Score × Wiki Gnomes					.049***	.005
Previous Quality Score × Vandal Fighter					.001	.005
Previous Quality Score × Fact Updater					-.008	.005
Previous Quality Score × Wikipedian					-.039***	.005
R-Squared	0.219		0.224		0.228	

Table 4: Article Quality Prediction Performances. P-value: $< .001$:***, $< .01$:**, $< .05$:*

Result Discussion

Results of four regression models are shown in Table 4. Regression Coefficient (Coef.) is reported, which represents the main change in the dependent variable for one standard deviation of change in the predictor variable while holding other predictors constant in the model. Model 1 reports the effects of the control variables.

The strongest predictors were the previous score (-.183) and the article bytes changed (.409). The negative correlation of pretest score with change score reflects both regression towards the mean and the substantive phenomenon that as articles rise to higher quality levels, it is more difficult to increase their quality further. The positive coefficient for edits by registered may simply reflect that more edits generally leads to higher quality or may reflect the distinctive importance of registered as opposed to anonymous editors. The number of editors working on the article (-.046) and the amount of activity on the talk page (-.031) were negatively correlated with quality which may confirm prior work's conclusions (Kittur and Kraut 2008) about cost of coordination in influencing article quality.

Model 2 adds roles' activity to the model and achieves a boost of .005 to the R-Squared. Examining this result in more detail suggests that more activity by substantive expert (.058) and less activity by Wiki gnomes (-.033) predicts of quality improvements. The value of substantive experts is that they add substantive information to an article. In contrast, Wiki gnomes contribute Wikipedia specific cleanup edits. This type of work may be unimportant to article quality or even detrimental, at least based on our automated measures. Alternatively, Wiki gnomes might be drawn to articles whose quality is declining because of the work of other editors.

To determine if the effect on quality of contribution by different roles depends upon the initially quality of the article, Model 3 adds the eight interaction terms between the previous quality score and the contribution of different roles (e.g., Social Networker × Previous Score.). Again we see an improvement to the R-Squared, suggesting that the activities of different types of editors are needed at different stages of article development. The negative coefficient for Substantive expert × Previous score (-.139) suggests that, as articles increase in quality, the substantive content provided by substantive experts is needed less. In contrast, the positive coefficient for Wiki gnomes × Previous score (.021) suggests that, as articles increase in quality, activity the cleanup activities by Wiki gnomes become more important. Although one might have expected the cleanup work done by copy editors, who in conventional publishing are most heavily involved in the final stages of manuscript production, would also become more important for higher quality, more complete articles, the negative coefficients disconfirm this conjecture.

Discussion and Conclusion

This paper focused on identifying editors' roles in Wikipedia and how the work contributed by editors in different roles affect the article quality. To achieve these goals we introduced a fine-grained taxonomy of edit types to characterize users' edits and built machine learning models to automatically identify the edit categories in each edit. We appropriated LDA-like graphical models to extracted latent roles from editors' history of edit activities. Finally, we examined the relationship between contributions of different types of editor to the improvement of article quality.

This research is an initial step in understanding the nature and value of social roles in online production and leaves much room for improvement, which we hope to address in future research. First, our role labeling is based on first identifying semantically meaningful edit types (e.g., adding information or paragraphing). The entire role modeling pipeline depends on creating an appropriate taxonomy of edit types, of accurately classifying each type of edit, of developing models that can account for each edit type. Each of these steps could be improved. Second, our role models take into account only the types of edits editors make and the namespaces where they work. Differentiating types of edits in other namespaces could be valuable (e.g., differentiating supportive versus critical comments in user and article talks pages (Zhu et al. 2011)). In addition, other features used by prior researchers should be included as input to the editor roles models, including user attributes, their social network signatures, users who edit multiple language editions (**?**), and the length of time spent editing (**?**). Future work can extend ours by including a more comprehensive set of relevant features as input to latent role representation. Third, although our findings suggest eight informal editor roles, whether a role accurately represents an editor is not clear. A natural next step is to conduct surveys or interviews, which ask Wikipedians whether our descriptions of them are reasonable. Fourth, our measurement of article quality comes from Wikipedia's Article Quality Predictor. This predictor may be accurate enough in matching human judgments, and because the judgments it is attempting to match are those of committed Wikipedia editors, it may not reflect the characteristics of articles that ordinary readers consider important to quality, such as the recency of the information cited or its accuracy.

We embarked on this research with the hope that automated identification of editors' roles would be useful in building recommender systems to better match editors to work. Although we have demonstrated the promise of social role modeling in Wikipedia, we believe that this approach could be applied to other online production communities, if they require a variety of skills from different contributors to be successful.

Acknowledgement

This research was supported in part by NSF grant IIS-1344768 and a grant from Google.

References

Anderka, M.; Stein, B.; and Lipka, N. 2012. Predicting quality flaws in user-generated content: The case of wikipedia. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, 981–990. New York, NY, USA: ACM.

Arazy, O.; Ortega, F.; Nov, O.; Yeo, L.; and Balila, A. 2015. Functional roles and career paths in wikipedia. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, 1092–1105. New York, NY, USA: ACM.

Bamman, D.; O'Connor, B.; and Smith, N. A. 2013. Learning latent personas of film characters. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 352–361. Sofia, Bulgaria: Association for Computational Linguistics.

Biddle, B. J. 2013. *Role theory: Expectations, identities, and behaviors*. Academic Press.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3:993–1022.

Cosley, D.; Frankowski, D.; Terveen, L.; and Riedl, J. 2007. Suggestbot: Using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces*, IUI '07, 32–41. New York, NY, USA: ACM.

Daxenberger, J., and Gurevych, I. 2012. A corpus-based study of edit categories in featured and non-featured Wikipedia articles. In *Proceedings of COLING 2012*, 711–726. Mumbai, India: The COLING 2012 Organizing Committee.

Daxenberger, J., and Gurevych, I. 2013. Automatically classifying edit categories in Wikipedia revisions. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 578–589. Seattle, Washington, USA: Association for Computational Linguistics.

De la Calzada, G., and Dekhtyar, A. 2010. On measuring the quality of wikipedia articles. In *Proceedings of the 4th Workshop on Information Credibility*, WICOW '10, 11–18. New York, NY, USA: ACM.

Faigley, L., and Witte, S. 1981. Analyzing revision. *College composition and communication* 400–414.

Fisher, D.; Smith, M.; and Welser, H. T. 2006. You are who you talk to: Detecting roles in usenet newsgroups. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences - Volume 03*, HICSS '06, 59.2–. Washington, DC, USA: IEEE Computer Society.

Hu, M.; Lim, E.-P.; Sun, A.; Lauw, H. W.; and Vuong, B.-Q. 2007. Measuring article quality in wikipedia: Models and evaluation. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, CIKM '07, 243–252. New York, NY, USA: ACM.

Kittur, A., and Kraut, R. E. 2008. Harnessing the wisdom of crowds in wikipedia: Quality through coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, 37–46. New York, NY, USA: ACM.

Kittur, A., and Kraut, R. E. 2010. Beyond wikipedia: Coordination and conflict in online production groups. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, 215–224. New York, NY, USA: ACM.

Kittur, A.; Suh, B.; Pendleton, B. A.; and Chi, E. H. 2007. He says, she says: Conflict and coordination in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, 453–462. New York, NY, USA: ACM.

Kittur, A.; Pendleton, B.; and Kraut, R. E. 2009. Herding the cats: The influence of groups in coordinating peer production. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, WikiSym '09, 7:1–7:9. New York, NY, USA: ACM.

Kriplean, T.; Beschastnikh, I.; and McDonald, D. W. 2008. Articulations of wikiwork: Uncovering valued work in wikipedia through barnstars. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, 47–56. New York, NY, USA: ACM.

Landis, J. R., and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *biometrics* 159–174.

Lipka, N., and Stein, B. 2010. Identifying featured articles in wikipedia: writing style matters. In *Proceedings of the 19th international conference on World wide web*, 1147–1148. ACM.

Liu, J., and Ram, S. 2009. Who does what: Collaboration patterns in the wikipedia and their impact on data quality. In *19th Workshop on Information Technologies and Systems*, 175–180.

Liu, J., and Ram, S. 2011. Who does what: Collaboration patterns in the wikipedia and their impact on article quality. *ACM Trans. Manage. Inf. Syst.* 2(2):11:1–11:23.

Orlikowski, W. J. 2000. Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science* 11(4):404–428.

Pfeil, U.; Zaphiris, P.; and Ang, C. S. 2006. Cultural differences in collaborative authoring of wikipedia. *Journal of Computer-Mediated Communication* 12(1):88–113.

Qin, X.; Greene, D.; and Cunningham, P. 2014. A latent space analysis of editor lifecycles in wikipedia. *arXiv preprint arXiv:1407.7736*.

Tsoumakas, G., and Vlahavas, I. 2007. Random k-labelsets: An ensemble method for multilabel classification. In *Machine learning: ECML 2007*, 406–417. Springer.

Tsoumakas, G.; Katakis, I.; and Vlahavas, I. 2010. Mining multi-label data. In *In Data Mining and Knowledge Discovery Handbook*, 667–685.

Viégas, F. B.; Wattenberg, M.; and McKeon, M. M. 2007. The hidden order of wikipedia. In *Proceedings of the 2Nd International Conference on Online Communities and Social Computing*, OCSC'07, 445–454. Berlin, Heidelberg: Springer-Verlag.

Warncke-Wang, M.; Cosley, D.; and Riedl, J. 2013. Tell me more: An actionable quality model for wikipedia. In *Proceedings of the 9th International Symposium on Open Collaboration*, WikiSym '13, 8:1–8:10. New York, NY, USA: ACM.

Welser, H. T.; Gleave, E.; Fisher, D.; and Smith, M. 2007. Visualizing the Signatures of Social Roles in Online Discussion Groups. *The Journal of Social Structure* 8(2).

Welser, H. T.; Cosley, D.; Kossinets, G.; Lin, A.; Dokshin, F.; Gay, G.; and Smith, M. 2011. Finding social roles in wikipedia. In *Proceedings of the 2011 iConference*, iConference '11, 122–129. New York, NY, USA: ACM.

Yang, D.; Wen, M.; and Rose, C. 2015. Weakly supervised role identification in teamwork interactions. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1671–1680. Beijing, China: Association for Computational Linguistics.

Zhu, H.; Kraut, R. E.; Wang, Y.-C.; and Kittur, A. 2011. Identifying shared leadership in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, 3431–3434. New York, NY, USA: ACM.